

Empirische Gesetzmässigkeit für nacheinander m-fach identisch gezogene Zahlen aus unendlich grosser Urne mit „n“ verschiedenen, gleich – oder ungleich verteilten Zahlen zwischen 0 und 9. (Gesetz zur Gruppenbildung)

Der fragliche Beitrag vom 13.02.2025 (mit Präzisierungen) wurde generell zurückgezogen. Grund dafür war ein in den Tabellen 1 – 4 falscher Rechnungsgang bei der Sortierung der aus der Urne gezogenen Zufallszahlen. Die Konzeption als Ganzes, wie auch der Kommentar, die Erwägungen und die Schlussfolgerungen sind aber weiterhin gültig.

Die hier erneuerte Abhandlung übernimmt den früheren Text (mit Anpassungen) und liefert zum Thema überarbeitete Tabellen.

19. 04. 2025

Praktische Frage zu «Kopf oder Zahl» (n = 2)

Wenn unendlich oft «Kopf oder Zahl» geworfen wird, interessiert die Frage, wie häufig (anteilmässig) sich «Kopf» respektive «Zahl» in der Reihenfolge abwechseln, und: wie häufig zweimal nacheinander «Kopf» oder «Zahl» folgt, und: wie oft dreimal sich nacheinander «Kopf» oder «Zahl» folgen, usw. usw.

Praktische Frage zum «Würfel» (n = 6)

Wie oft (anteilmässig) folgt nach einer beliebigen Augenzahl n ($1 \leq n \leq 6$) eine Augenzahl $\neq n$ – und auch: wie oft folgt nacheinander 2-mal, 3-mal, 4-malm-mal) dieselbe Augenzahl.

Die nachstehende Gesetzmässigkeit gilt ebenso bei drei, vier, fünf, sieben, acht, neun oder zehn verschiedenen, in der Urne **gleichverteilten Zahlen** zwischen 0 und 9.

Die Basisformel für gleichverteilte Zahlen lautet: **Promilleanteil = $A \cdot e^{(B \cdot m)}$**

Gruppenbildung «m = 1» bedeutet beispielsweise: Auf eine gezogene Zahl wurde/wird weder unmittelbar vorher noch unmittelbar nachher dieselbe Zahl gezogen. Die Koeffizienten A und B ändern sich je nach Anzahl verschiedener, gleichverteilter Zahlen in der Urne (n = 2 heisst: zwei verschiedene Zahlen, n = 10 heisst: alle Zahlen von 0 bis 9). Der Wert „m“ in der Formel entspricht der Vorgabe des Paketumfangs aus gleicher Zahl.

Je nach «n» unterschiedlichen Zahlen in der Urne sind für die Basisformel die folgenden Kennwerte „A“ und „B“ gegeben. Der Korrelationskoeffizient „r“ für die Funktion liegt immer um den Wert $r \sim 0.999$.

ANZAHL "n" - VERSCHIEDENE ZAHLEN	KONSTANTENTERM "A"	REGRESSIONSKOEFFIZIENT "B"
n = 2	500.745642	-0.693442
n = 3	1110.130301	-1.051134
n = 4	2554.923721	-1.453712
n = 5	2450.381986	-1.502727
n = 6	2803.285179	-1.548605
n = 7	5276.725696	-1.959718
n = 8	7580.027291	-2.173816
n = 9	7462.617685	-2.209558
n = 10	7133.015259	-2.248665

Tabelle 1 Kennwerte zur Basisformel in Abhängigkeit der Anzahl unterschiedlicher, aber stets gleichverteilter (!) Zahlen „n“.

	Promille für n = 2	Promille für n = 3	Promille für n = 4	Promille für n = 5	
m = 1	262	440	0.8	637	
m = 2	130.1	149.7	33.3	125.8	
m = 3	61.9	50.2	8.7	25.7	
m = 4	29.7	15.4	1.6	4.9	
m = 5	13.7	5.3	0.4	1.4	
m = 6	8.3	1.7	0.1	0.2	
m = 7	4.1	0.5			
m = 8	1.9	0.3			
m = 9	1				
m = 10	0.7				
	Promille für n = 6	Promille für n = 7	Promille für n = 8	Promille für n = 9	Promille für n = 10
m = 1	658	740	767	795	809
m = 2	115.5	106.6	93.8	87.4	80.2
m = 3	20.8	14.8	12	9.6	7.1
m = 4	3.3	2	2	1.4	0.9
m = 5	0.9	0.3	0.1	0.2	0.1
m = 6					
m = 7					
m = 8					
m = 9					
m = 10					
	"MÜNZE"				
	"WÜRFEL"				

Tabelle 2 Resultierende Gruppenbildungsanteile in Promille je nach «n» und «m».

Herleitung und Genauigkeit

Für jede «Variabilität n» (n = 2 bis n = 10) wurden jeweils 1000 gleichverteilte einziffrige Zahlen erzeugt sowie die sich daraus gebildeten Gruppierungen ausgezählt und als entsprechende Promillewerte ab m = 1 (steter Wechsel), m = 2,, bis m = 10 festgehalten. Für dieselbe «Variabilität n» wurde das Prozedere 20-mal durchgeführt. Werden 20 mal 1000 Zufallswerte generiert, folgen daraus 20 leicht abweichende Promillewerte, welche um den gewonnenen (und tabellierten) Mittelwert einer Promilleangabe oszillieren. Die Bandbreiten dieser „Fehlerabweichungen“ vom unbekannten wahren Promillewert sind (als Ergänzung zu Tabelle 2) in einer separaten **Tabelle 3** unten zusammengestellt.

	Promille für n = 2	Promille für n = 3	Promille für n = 4	Promille für n = 5	
m = 1	26.92	16.98	21.95	21.78	
m = 2	11.09	6.61	8.53	7.36	
m = 3	7.36	5.04	5.85	5.22	
m = 4	5.44	3.62	2.99	1.81	
m = 5	4.26	2.56	1.49	1.17	
m = 6	2.88	0.85	0.64	0.43	
m = 7	1.92	0.75	0.21	0.21	
m = 8	1.39	0.53			
m = 9	1.19	0.21			
m = 10	0.96				
	Promille für n = 6	Promille für n = 7	Promille für n = 8	Promille für n = 9	Promille für n = 10
m = 1	19.88	23.11	20.65	15.91	19.66
m = 2	6.93	6.4	11.41	9.91	8.1
m = 3	4.58	5.01	3.84	3.84	2.99
m = 4	1.39	1.6	1.51	1.07	0.96
m = 5	0.85	0.53	0.32	0.28	0.21
m = 6					
m = 7					
m = 8					
m = 9					
m = 10					
	"MÜNZE"				
	"WÜRFEL"				

Tabelle 3 (±) Fehler – Bandbreiten als Absolutwerte zu den Angaben laut Tabelle 2

Ungleiche Gewichtung der Zahlen in der Urne

*Vorbemerkung: Der für eine «Gruppenbildung m» tabellierte Promillewert für gleichgewichtete Zahlen in der Urne besagt, dass dabei alle Zahlen von gleicher (eben dieser) «Gruppengrösse m» im Promillewert eingerechnet sind. **Beispiel:** Bei drei vorkommenden Zahlen 1, 4 und 7 in der Urne ist der angegebene Promillewert **50.2** die Folge der Addition aus **allen** (1, 1, 1) und (4, 4, 4) und (7, 7, 7). Sofern diese drei Zahlen in der Urne also gleichverteilt (gleichgewichtet) sind, kann ohne vertiefte Auszählung davon ausgegangen werden, dass sich für diese 3er – Pakete der tabellierte Promilleanteil **50.2** zu gleichen Teilen auf die drei Pakete (1, 1, 1) resp. (4, 4, 4) resp. (7, 7, 7) aufteilt.*

Bei ungleich – gewichteten Zahlen in der Urne müssen die einzelnen Promillewerte aus einem (z.B.) 3er Paket **hingegen** individuell ermittelt werden. **Beispiel:** Angenommen, wir verfügen über einen „Zauberwürfel“, bei welchem die Augenzahl 6 das sechsfache Gewicht gegenüber der Augenzahl 1 (A1) aufweist, die Augenzahl 5 das fünffache Gewicht der Augenzahl 1, etc. Demzufolge sind die Augenzahlen 1 bis 6 zu folgenden Promillesätzen in der Urne vertreten: $A1 = (1/21) \cdot 1000$; $A2 = (2/21) \cdot 1000$; $A6 = (6/21) \cdot 1000$. Man

unterteilt demnach das Zufallsband (1 bis 1000) für die Ziehung in entsprechende Abschnitte und ordnet jede gezogene Zahl (zwischen 1 und 1000) der Urnenzahl mit dem entsprechenden Bandabschnitt zu. Aus der Abfolge von Ziehungen (ad infinitum) bildet sich so das Gesetz zur Gruppenbildung für das betreffende Urnensample. Dabei sind die jeweiligen „m – Pakete“ unter den einzelnen Zahlen nicht mehr gleichverteilt. Da es eine beliebige Anzahl individuell zuteilbarer Gewichtungen gibt, macht es hier keinen Sinn, das Ergebnis der jeweiligen Auszählung in eine Basisformel zu fassen.

Spezialfall zum Gesetz von Benford

Die individuelle Aufteilung der Erstziffern von Datensätzen, welche dem Gesetz von Benford folgen, ist vorgegeben und beschreibt die anteilige Aufteilung der Erstziffern 1 bis 9. Siehe dazu z. B. die Abhandlung unter Wikipedia, mit **Abbildung 1: Benfordsches Gesetz**. Die neun Erstziffern weisen eine ungleiche Gewichtung auf.

*Dazu eine Randbemerkung: Abweichend von der Form der Herleitung der „Zahlenanteile“ laut Wikipedia, gelangt man zum exakt gleichen Ergebnis, wenn die bekannten „Zahlenanteile“ in eine logistisch – lognormale Funktion gemäss **Abbildung 2** gelegt werden.*

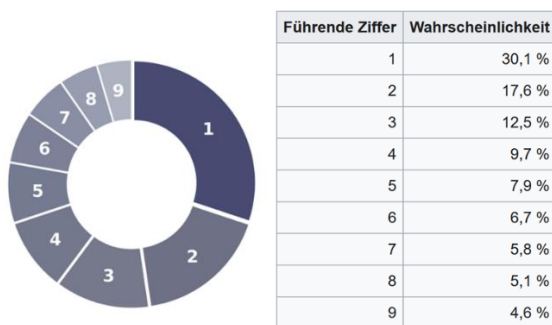


Abbildung 1: «Original» -Benfordsches Gesetz; Grafik aus Wikipedia

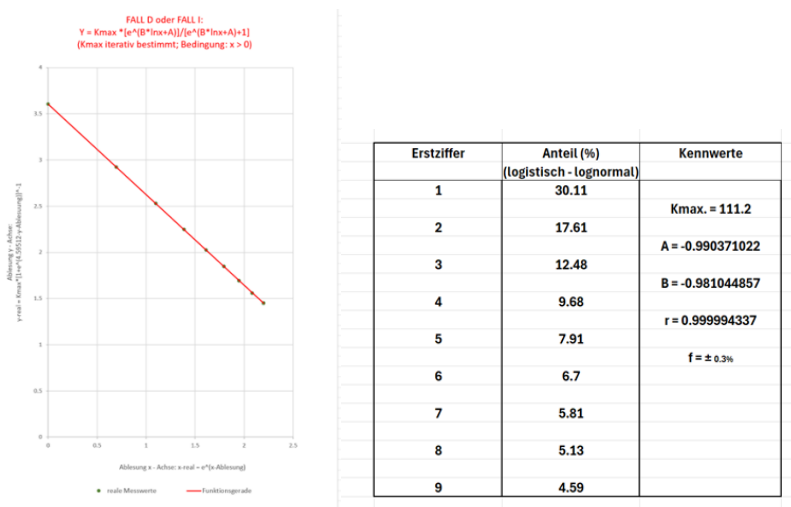


Abbildung 2: Benfordsches Gesetz «logistisch – lognormal»

Die Basisformel für die Erstziffern nach dem Gesetz von Benford lautet:

$$\text{Promilleanteil} = K_{\max} * [e^{(B * \ln z + A)}] / [e^{(\ln z + A)} + 1]$$

Je nach interessierender Gruppenbildung «m» sind in der Basisformel die nachstehenden Kennziffern (K_{max}, A, B) einzusetzen. Der Formelwert z entspricht der gezogenen Erstziffer (1 ≤ z ≤ 9) aus der Urne. K_{max} wurde iterativ ermittelt. Der Korrelationskoeffizient für die Basisfunktion liegt stets um die 0.999.

Gruppenbildung "m"	K _{max}	KONSTANTENTERM "A"	REGRESSIONSKOEFFIZIENT "B"
m = 1	192.4	1.266338386	-1.172238909
m = 2	70	0.534957712	.1.90832476
m = 3	14.7	1.622241034	-3.25693147
m = 4	9.6	-0.397261987	-2.93293715

Tabelle 4 Kennwerte zur Basisformel bei Erstziffern mit benfordscher Verteilung

Gruppenbildung "m"	Promille für Ziffer z = 1	Promille für Ziffer z = 2	Promille für Ziffer z = 3	Promille für Ziffer z = 4	Promille für Ziffer z = 5
m = 1	149	119	96	80	68
m = 2	44.4	20.9	12.5	7.6	5.6
m = 3	12.4	4.7	1.7	0.9	0.4
m = 4	3.9	0.7	0.3	0.3	0.1
m = 5	1.1				
m = 6	0.4				
m = 7	0.3				

Gruppenbildung "m"	Promille für Ziffer z = 6	Promille für Ziffer z = 7	Promille für Ziffer z = 8	Promille für Ziffer z = 9	Promille für ALLE Ziffern z 1 bis 9
m = 1	58	49	46	41	706
m = 2	3.2	2.9	2.7	1.6	99.9
m = 3	0.2	0.2			20.5
m = 4					5.3
m = 5					1.1
m = 6					0.4
m = 7					0.3

Tabelle 5 Promilleanteile für die Erstziffern mit benfordscher Verteilung

Gruppenbildung "m"	Promille für Ziffer z = 1	Promille für Ziffer z = 2	Promille für Ziffer z = 3	Promille für Ziffer z = 4	Promille für Ziffer z = 5
m = 1	12.85	8.87	9.53	8.52	8.41
m = 2	5.44	5.54	3.52	2.99	2.35
m = 3	3.94	2.13	1.49	1.07	0.53
m = 4	1.81	0.96	0.64	0.53	0.21
m = 5	1.17				
m = 6	0.64				
m = 7	0.41				

Gruppenbildung "m"	Promille für Ziffer z = 6	Promille für Ziffer z = 7	Promille für Ziffer z = 8	Promille für Ziffer z = 9	Promille für ALLE Ziffern z 1 bis 9
m = 1	8.65	7.12	7.68	6.98	19.41
m = 2	2.24	2.03	1.6	1.39	8.42
m = 3	0.43	0.41			4.67
m = 4					2.38
m = 5					1.17
m = 6					0.64
m = 7					0.32

Tabelle 6 (±)Fehler - Bandbreiten als Absolutwerte zu den Ergebnissen laut Tabelle 5

Wo das «Gesetz zur Gruppenbildung» nützlich sein kann:

Zuerst einmal darf das Gesetz zur Gruppenbildung zu folgendem **nicht** verleiten:

Zwar beträgt die Wahrscheinlichkeit, dass sich bei der Spielform « $n = 2$ » - z.B. Roulette schwarz oder rot - ein «3er – Paket schwarz» bildet, nur etwa 6.2% (enthaltend 18.6% aller Kugeln in der Urne). Liegt ein solches „Paket schwarz“ vor, könnte dies dazu verleiten, vor der nächsten Ausspielung auf rot zu setzen, zumal die Wahrscheinlichkeit für ein entstehendes „4er Paket schwarz“ nur knapp 3% beträgt. (→ 97% nächste Ziehung = rot ??) Das wäre allerdings Selbstbetrug, denn auch im nächsten Umgang haben schwarz und rot dieselbe Eintreffenschance. Der Grund für den scheinbaren Widerspruch liegt darin, dass die ausgewiesenen Gruppierungsanteile stets unendlich viele Ziehungen umfassen.

Grundsätzlich kann man davon ausgehen, dass sich die Geburten von Knaben und Mädchen «im Schnitt auf die Dauer» in etwa die Waage halten («Kopf oder Zahl») und die Gruppenbildung damit der «Spielform $n = 2$, gleichgewichtet» entspricht. Sollte sich hier nach zeitlicher Abfolge einer Grosszahl von gemeldeten Geburten über eine längere Zeitspanne eine stark abweichende Gruppierung gegenüber der Gesetzmässigkeit zeigen, müsste dies den Schluss zulassen, dass - warum auch immer - z.B. nicht alle Mädchengeburten als solche gemeldet worden sind. Aber auch, wenn in der Summe der Meldungen je 50% als Knaben und Mädchen aufgeführt sind, ist damit noch keineswegs sichergestellt, dass die Liste nicht in diese oder jene Richtung hin frisiert worden ist....?

Ganz generell: Wenn eine in der Abfolge gelistete grosse Datenmenge von Geschehnissen vorliegt, von welchen die Häufigkeit des Eintreffens ihrer Anteile bekannt ist oder „als sicher vermutet“ wird, kann die Plausibilität des gelieferten Datensatzes nach dem «Gesetz zur Gruppenbildung» überprüft werden.

Überprüfte Plausibilität eines Datensatzes nach dem Gesetz zur Gruppenbildung

Entsprechend der bekannten (oder erwarteten) Aufteilung der verschiedenen Zahlen (0 bis 9) innerhalb einer Datenlieferung, kann mittels «Gesetz zur Gruppenbildung» überprüft werden, ob die Lieferung der erwarteten «Sollverteilung» in etwa entspricht. Die Abweichung von der entsprechenden Referenzverteilung kann dabei über ein einfaches, lineares X/Y – Diagramm quantifiziert werden. Es liegt dann im Ermessen des Prüfenden, welche Abweichung (Korrelationskoeffizient « r », $(\pm)$ Fehlerband) der Datenlieferung noch toleriert werden soll.

Für praktische Anwendungen ist Fantasie gefragt!

Nachtrag 24.04.2025

Dem obigen Beispiel „Zauberwürfel“ folgend (Abschnitt: ungleich gewichtete Zahlen in der Urne), sind in den ergänzenden **Tabellen 7 und 8** unten noch die entsprechenden Auswertungen „en detail“ zusammengestellt. Desgleichen wurden die **Tabellen 5 und 6** (Benford – Verteilungen) ergänzend zu den einzelnen Promillewerten je Gruppengrösse (m) und Erstziffer (z) noch um die Spalte „Promille für ALLE Ziffern z = 1 bis 9“ (je nach Gruppengrösse m) ergänzt.

Gruppenbildung "m"	Promille für Augenzahl z = 1	Promille für Augenzahl z = 2	Promille für Augenzahl z = 3	Promille für Augenzahl z = 4
m = 1	43.9	77.3	103.3	124.3
m = 2	2.2	6.4	15.7	23.1
m = 3	0.2	0.5	2.7	4.1
m = 4		0.1	0.2	0.5
m = 5				0.2
m = 6				
m = 7				
Gruppenbildung "m"	Promille für Augenzahl z = 5	Promille für Augenzahl z = 6	Promille für ALLE Augenzahlen z = 1 bis 6	
m = 1	134.9	148.2	631.8	
m = 2	34.8	40.5	122.1	
m = 3	8.5	12.3	28.6	
m = 4	2.2	3.1	8.1	
m = 5	0.3	1.1	1.5	
m = 6	0.1	0.4	0.4	
m = 7			0.1	

Tabelle 7: Promillewerte zur Gruppenbildung des „Zauberwürfels“; Gewichtung der Augenzahlen gemäss Textbeispiel oben.

Gruppenbildung "m"	Promille für Augenzahl z = 1	Promille für Augenzahl z = 2	Promille für Augenzahl z = 3	Promille für Augenzahl z = 4
m = 1	5.38	6.76	8.4	8.04
m = 2	1.28	2.13	6.1	5.65
m = 3	0.43	0.64	1.28	2.45
m = 4		0.32	0.43	0.64
m = 5				0.43
m = 6				
m = 7				
Gruppenbildung "m"	Promille für Augenzahl z = 5	Promille für Augenzahl z = 6	Promille für ALLE Augenzahlen z = 1 bis 6	
m = 1	12.5	10.9	18.51	
m = 2	6.18	7.36	10.45	
m = 3	2.88	3.62	5.01	
m = 4	1.49	2.45	3.62	
m = 5	0.53	1.07	0.96	
m = 6	0.21	0.75	0.74	
m = 7			0.21	

Tabelle 8: (±)Fehler - Bandbreiten als Absolutwerte zu den Ergebnissen laut Tabelle 7