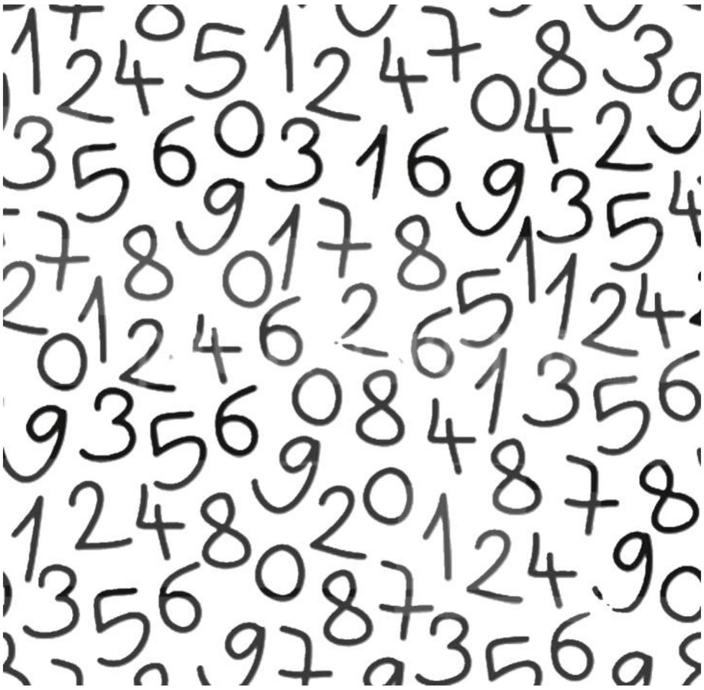
OEKOPRIORITY®

DAS GESETZ ZUR GRUPPENBILDUNG

TEIL 6 – Empirisch ist nicht Punktgenau!

MIT AUSBLICK



siehe auch TEIL 1; TEIL 2; TEIL 3; TEIL 4

TEIL 5A; TEIL 5B; TEIL 5C

April 2026

WORUM ES GEHT

In den hier auf der Titelseite vermerkten LINKS wird die Thematik «Gesetz zur Gruppenbildung» aus verschiedenen Blickwinkeln besprochen und mit praktischen Beispielen verdeutlicht. Das Gesetz kann demnach bei vielen Disziplinen und Lebensbereichen praktisch angewendet werden, wobei letztlich immer dieselbe Frage tangiert wird:

Ist eine gelieferte Datenreihe – aus welchem Themenbereich auch immer – aus Zufall (besser noch: inwieweit aus Zufall) entstanden – oder folgte deren Entstehung einer unbekannten Steuerung?

Es ist genau diese Kernfrage, welche mit Hilfe des «Gesetzes zur Gruppenbildung» nicht nur qualitativ (ja / nein), sondern auch quantitativ (reine Zufälligkeit in Prozent) beantworten werden soll.

Wie der Titel des Beitrags vermuten lässt, gibt es mehrere Arten von Zufälligkeit als Messgrösse für eine gelieferte Zahlenreihe, so zumindest eine *«wahre», naturgegebene – um nicht zu sagen «göttliche» – Zufälligkeit, ferner eine *«pseudo – wahre» Zufälligkeit sowie eine *«logische» Zufälligkeit. Worin besteht der massgebende Unterschied? **Diese Begriffe sind nicht «wissenschaftlich etabliert».*

WAHRE ZUFÄLLIGKEIT

Bleiben wir bei bisher verwendeten Beispielen:

Sowohl die «faire Münze»(n = 2) wie der «faire Würfel» (n = 6) unterliegen der wahren Zufälligkeit. Das braucht nicht näher diskutiert zu werden. Es ist unbestritten, dass nach unendlich vielen Wiederholungen eines Münzenwurfs oder eines Würfels Kopf und Zahl, bzw. die Augenzahlen 1 bis 6 gleich oft gezählt würden. Bei jedem einzelnen Wurf herrscht wahre Zufälligkeit darüber, welches Ergebnis herauskommt. Diese Zufallsergebnisse können punktgenau formuliert werden. Abweichungen davon sind berechenbar.

Das Ergebnis eines Vergleiches «SOLL – IST» kann sowohl empirisch durch eine grosse Anzahl von Versuchen eingegrenzt – **oder formelmässig** berechnet werden.

Angenommen, es stellt sich die Frage:

Wie häufig folgt (in Promille P) beim «fairen Würfel» aus unendlicher Wiederholung ANTEILMÄSSIG **dreimal** hintereinander die Seite mit Augenzahl 4?

Es gilt über die Formel eindeutig:

$P = 1000 * (p^{-1}-1)^2 * p^{(m+2)}$ → ~ 3.215 mal, bezogen auf 1000 Ziehungen

heisst: **3.215 Pakete (P)** aus (p = 1/n), wobei n = Anzahl Variablen → hier sechs Würfelseiten, die sich dreimal hintereinander mit Augenzahl vier folgen

Statt mit geschlossener Formel kann das «wahre» Ergebnis auch **empirisch** (mit Fehlerbandbreite!) **eingegrenzt** werden

P = 3.46 ± 0.76 mal, bezogen auf 1000 Ziehungen

Sowohl der Wert 3.46 wie die Bandbreite ± 0.76 aus 20*1000 Losziehungen sind dabei «leicht variabel», falls x-mal 20*1000 Ziehungen vorgenommen werden. Der «wahre» Wert aus unendlichem Würfeln (hier ~ 3.21 pro Tausend) wird vom empirischen Ergebnis aber immer zuverlässig umschlossen.

Die bisherige Annahme allerdings, dass ungleich häufigen oder gewichteten Variablen – theoretisches Beispiel: Ein «Zauberwürfel» mit Gewicht 1 für die Seite mit Augenzahl 1……Gewicht 6 für die Seite mit Augenzahl 6 – eine punktgenaue Aussage mittels geschlossener Formel nicht möglich ist, **hat sich überholt!**

Tatsächlich wurde zwischenzeitlich eine erweiterte Formel gefunden, welche **auch bei ungleichen Verteilungen** der Zahlen zum **«wahren»** Ergebnis führt!
Siehe dazu auch letzter Abschnitt: AUSBLICK

Die erweiterte Formel lautet:

Aus VORGABE der Anzahl «n» folgt: $P_{(i,m)} = 1000 * (1 - p_{(Zahl\ i)})^2 * (p_{(Zahl\ i)})^m$

$P_{(i,m)}$ heisst: Bei unendlichem Würfeln (hier z.B. des «Zauberwürfels») bildet sich pro 1000 Durchgänge eine eindeutige ANZAHL von Paketen, in denen (nur) die Zahl «i» zufallsbedingt «m - fach» enthalten ist.

$p_{(Zahl\ i)}$ heisst: Die Zahl «i» weist eine relative Häufigkeit oder Gewichtung, bezogen auf alle «n» verschiedenen Zahlen, von «p» (Prozent/100) auf.

*«i» ist die «n-te» verschiedene Zahl (n ≥ 2; nach oben theoretisch offen). Ist das Synonym für **eine der** «n» Variablen (z.B. Würfelseiten → n = 6)*

«m» ist die Anzahl der im jeweiligen Paket enthaltenen $Zahl_i$ (m ≥ 1; nach oben theoretisch offen)

Es gibt viele andere Fälle von «punktgenauer Zufälligkeit», so beispielsweise für die Binomialverteilung:

→ Die Potenz 2^{19} zerlegt sich in 10 verschiedene Binomialkoeffizienten, resultierend aus der Summe aller «19 tief k» (k von 0 bis 19). Aus den 10 verschiedenen Häufigkeiten folgt auch hier die jeweilige Anzahl von Paketen «pro 1000 rein zufälliger Ziehungen».

Eine gelieferte Zahlenreihe kann damit elegant daraufhin überprüft werden, inwieweit sie einer Binomialverteilung entspricht. Weitere Anwendungen aus dem «Reich mathematischer Gesetzmässigkeiten» sind denkbar.

«PSEUDO – WAHRE» ZUFÄLLIGKEIT

Eine solche Form von Zufälligkeit liegt beispielsweise beim «Gesetz von Benford» vor. Hier wurde aus einer **gemachten Beobachtung** über die Häufigkeitsanteile einzelner Ziffern (1 bis 9) in gewissen Datensätzen wie z.B. aus Telefonbüchern eine punktgenaue logarithmische Verteilungsfunktion **«konstruiert»**. Es handelt sich damit jedoch um ein Modell, welches eigentlich zu unrecht mit «wahrer Zufälligkeit» in Verbindung gebracht wird.

LOGISCHE ZUFÄLLIGKEIT

Überall dort, wo sich **in der Praxis** die Frage stellt, «wie zufällig» sind gelieferte Daten entstanden?», kann eine Überprüfung nach dem «Gesetz zur Gruppenbildung» wie folgt vorgenommen werden:

1. Der gelieferte Datensatz enthält *«n» unterschiedliche Werte mit unterschiedlicher Häufigkeit oder Gewichtung. Basierend auf der Anzahl «n» wird mittels der gegebenen Häufigkeitsverteilung unter den «n» - Variablen das früher beschriebene logische Referenzprofil erstellt. Dabei wird vorausgesetzt, dass die (gedanklich!) in eine Urne gelegten Einzelwerte durch «Losziehung» sich rein zufällig folgen. **Es ist zu unterscheiden: Bei verschiedenen «diskreten» Einzelwerten ist die Zuordnung nach Häufigkeit eindeutig. Beim Grossteil aller interessierenden Daten handelt es sich aber – jedenfalls bei «technischen Messungen» – um eine grosse Anzahl von (vielleicht nur minimal) unterschiedlichen Einzelwerten. Hier erfolgt eine Aufteilung der Daten in 10 gleichbreite Zahlenbänder, wodurch sich die zugeteilten Einzelwerte dort mit anteilmässiger Häufigkeit zusammenfinden.*

2. Wenn der gelieferte Datensatz tatsächlich rein zufällig entstanden sein soll – was ja die ursprüngliche Fragestellung ist – so muss dessen Stichprobenprofil (also die nach Einzelwerten aus ihrer Abfolge gebildeten Pakete) mit den berechneten Paketen für das logische Referenzprofil deckungsgleich sein.

3. Das Mass an Übereinstimmung wird über den früher beschriebenen PLAUSIBILITÄTSGRAD ermittelt.

ZUSAMMENFASSUNG

Bei der im Titel gemachten Feststellung: «Empirisch ist nicht Punktgenau!» geht es um den direkten Bezug zur Praxis. Das heisst: Es interessiert konkret die Frage, inwieweit eine erfasste Datenmenge aus Zufall entstanden ist.

Für eine Beantwortung «gemäss Berechnung» kommt man nicht darum herum, den Begriff «Zufall» zu überdenken.

Mit dem «Gesetz zur Gruppenbildung» bietet sich überall dort, wo für die Entstehung einer Datenabfolge **reine Zufälligkeit** der Einzelwerte **angenommen** wird (keine gegenseitige Abhängigkeit, keine Steuerung, etc.) die Möglichkeit, eine nachvollziehbare Überprüfung unter definierten Randbedingungen vorzunehmen.

Ein dazu erstelltes «logisches Referenzprofil» stellt dazu die sogenannte Nullhypothese dar, und der ermittelte PLAUSIBILITÄTSGRAD das Mass an Zufälligkeit, mit welchem der gelieferte Datensatz entstanden ist. Je kleiner der PLAUSIBILITÄTSGRAD, desto stärker war bei der Bildung der Datenreihe die Abweichung von der vorausgesetzten reinen Zufälligkeit.

Man kann auch sagen:

Wenn eine gelieferte Datenmenge in ihrer Reihenfolge (also entsprechend der Abfolge einer Ziehung aus der Urne) nach dem «Gesetz zur Gruppenbildung» mit ihrem «logischen Referenzprofil» vollständig übereinstimmt, **dann** sind die gelieferten Daten aus **reinem Zufall** entstanden.

AUSBLICK

Dank dem Umstand, dass zwischenzeitlich eine exakte Formel für die Paketbildung, gültig für gleichgewichtete wie für ungleichgewichtete Merkmale (Variablen, Zahlen, Gruppen/Zahlenbänder) gefunden wurde, ergeben sich **neue** (weitere) Möglichkeiten der **Überprüfung**:

Mit der **gefundenen Formel** kann für jede Paketbildung ein «wahrer» (sicherer) Erwartungswert (E) ohne Fehlerbandbreite bestimmt werden, welchem das jeweilige Paket aus der realen Datenabfolge (O) gegenübergestellt wird.

Auf diesem Weg lässt sich aus den «n x m» Paketen der massgebende Wert **«Chi$_X^2$»** bestimmen. Zusammen mit dem entsprechenden Freiheitsgrad **«df»** leitet sich daraus mit P* ein Wert ab, welcher besagt, zu wieviel Prozent «P*(%)» der erfasste Datensatz tatsächlich reiner Zufälligkeit entspringt. Das Mass ergibt sich hier aber nicht aus der Verhältniszahl P = (Pakete $_{IST}$ / (Pakete ± Fehlerbandbreite) wie beim empirischen Nachweis. Vielmehr folgt hier der ausgewiesene Prozentwert aus dem statistischen Ansatz P*(%) = $[1 - F_X{}^2 (X^2,df)]*100$.

In Kenntnis ihrer unterschiedlichen Randbedingungen können sowohl das «empirische Verfahren» wie das «statistische Verfahren» als «in sich kohärent» bezeichnet werden.

Sie sollen jedenfalls NICHT in Konkurrenz zueinander gestellt werden.

Anderseits ist eine begriffliche Unterscheidung zweckmässig:

Ergebnis aus empirischer Berechnung: PLAUSIBILITÄTSGRAD = $P_{empirisch}$ (%)

Ergebnis aus statistischer Berechnung: PLAUSIBILITÄTSGRAD = $P_{statistisch}$ (%)

Alle bis und mit März 2026 erstellten Beiträge und Beispiele basieren auf dem «empirische Verfahren».

Künftige Beiträge und Beispiele zum Thema sollen (nicht zuletzt, weil damit auf Zufallsgeneratoren verzichtet werden kann) das «statistische Verfahren» favorisieren.

April 2026