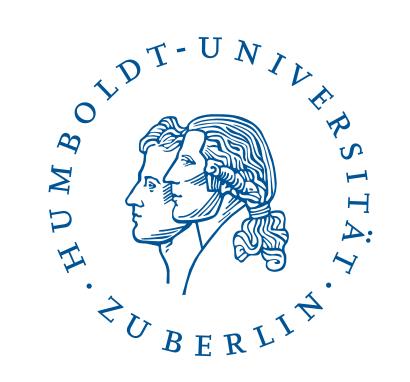


Zufallszahlen: Statistik

Fabian, Marius, Nils, Tim

Institut für Physik, Humboldt-Universität zu Berlin





Einführung

Wie ist es möglich zu entscheiden, ob eine Folge von Zahlen (aus Einsen und Nullen) wirklich zufällig ist? Sicher ist eine Folge wie

sehr unwahrscheinlich. Aber untersucht man wirklich zufällige Folgen nur lange genug, so erwartet man irgendwann auch ein solches Ergebnis. Die Frage lautet also eher: Wie wahrscheinlich ist die Zahlenfolge, wenn die Zahlen wirklich zufällig gezogen worden sind?

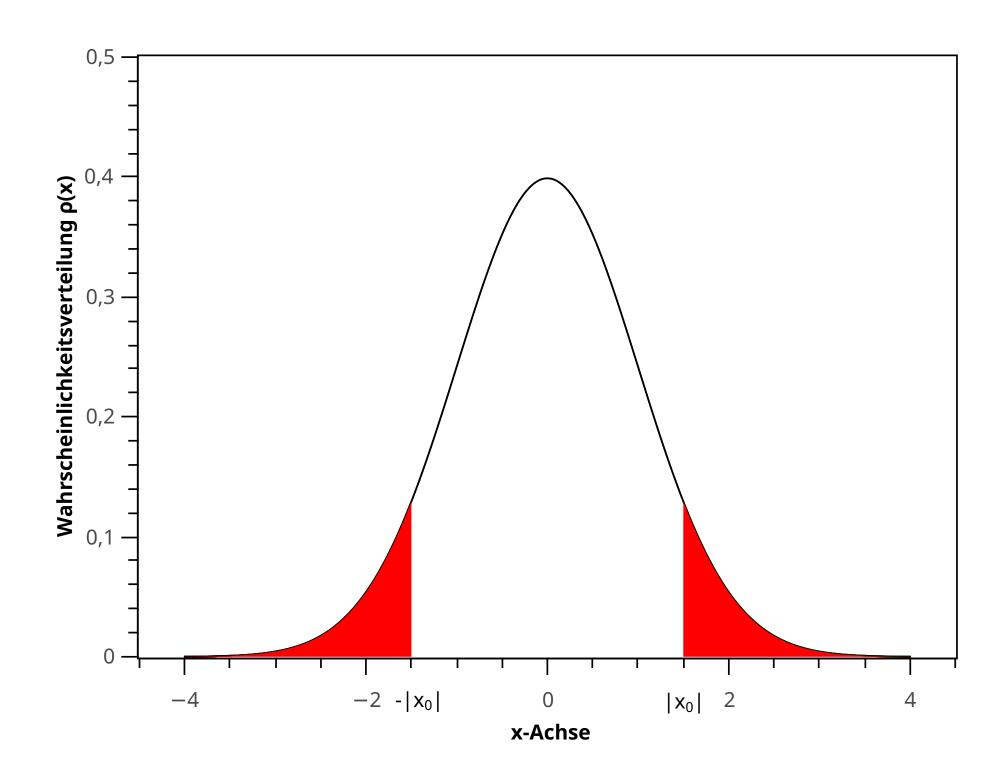


Abbildung 1:Normalverteilung

Oder etwas präziser formuliert: Wie wahrscheinlich ist es für die Zahlenfolge, einen Wert x in Hinblick auf eine bestimmte Eigenschaft aufzuweisen? Ist diese als p-Wert bezeichnete Wahrscheinlichkeit zu klein, so sagen wir, die Folge ist nicht zufällig. Die Grenze für "zu klein" ist willkürlich, aber vor dem Test festgelegt und wird Signifikanzniveau genannt. Für unsere Tests liegt das Signifikanzniveau bei 5%. Abbildung (1) zeigt eine beispielhafte Wahrscheinlichkeitsverteilung für x bei einer sogenannten Normalverteilung. Die rote Fläche gibt den p-Wert wieder, wenn für x der Wert x_0 erhalten wurde. Es gibt unendlich viele Eigenschaften, auf die eine Folge geprüft werden kann. In der Regel wird aber nur eine handvoll Eigenschaften getestet. Eine beispielhafte Auswahl von Tests sind hier aufgelistet.

Verhältnistest

Der wohl einfachste Test ist das Bestimmen des Verhältnisses von 1 zu 0. Wird hundert mal eine Münze geworfen, so erwartet man 50 mal Kopf. Taucht nur einmal Kopf auf, so erscheint das doch sehr unwahrscheinlich. Aber wie sieht es aus mit 45 mal Kopf, oder nur 39 mal, wie wahrscheinlich wäre das?

Die Bernoulli-Formel berechnet diese Wahrscheinlichkeit und für sehr viele Münzwürfe sieht die Wahrscheinlichkeitsverteilung ähnlich aus wie in Abbildung (1).

Für das Beispiel bedeutet das: Bei 100 Münzwürfen sind alle Ergebnisse mit weniger als 41 oder mehr als 59 mal Kopf zu unwahrscheinlich, um für uns als zufällig zu gelten.

Dieser Test liefert zwar einen ersten Anhaltspunkt, versagt aber auch bei einer Folge wie

101010101010101010101010101.

Der Verhältnis-Test findet diese Folge perfekt, auch wenn die Wahrscheinlichkeit für eine exakt alternierende Folge 1 zu 2²⁷ ist, also so wahrscheinlich, wie im Lotto zu gewinnen!

Runs-Test

Der Wald-Wolfowitz Runs-Test zählt die Anzahl an Runs. Was ist ein Run? Ein Run ist ein Durchgehender Block von 1 beziehungsweise 0. So hat beispielsweise die Folge 111010001100100 acht Runs, nämlich

111010001100100

Der Erfahrung nach sind Menschen nicht besonders gut darin, die Anzahl der Runs richtig einzuschätzen. Sind bei Münzwürfen fünfmal hintereinander Kopf unwahrscheinlich? Im Durchschnitt findet sich ein solches Ereignis alle 64 Zahlen in einer zufälligen Zahlenfolge, was leicht unterschätzt werden kann. Mit einfacher Kombinatorik kann ausgerechnet werden, wie wahrscheinlich eine bestimmte Anzahl an Runs in einer Zahlenfolge ist. Für sehr lange Folgen sieht die Wahrscheinlichkeitsverteilung ähnlich aus wie in Abbildung (1).

Autokorrelation

Autokorrelation ist eine Eigenschaft einer Datenreihe, die eine Tendenz angibt, mit der man aus einem Datenwert einen Folgenden voraussagen kann. Hierfür ein Beispiel:

Auf den ersten Blick sieht die Datenreihe recht zufällig aus. Einsen und Nullen sind in etwa gleich viele und auch recht gleich verteilt. Bei genauem Hinsehen wird jedoch deutlich, dass auf jede 1 eine 0 und umgekehrt folgt. Weiterhin folgt zwei Ziffern entfernt auf jede 1 wieder eine 1 und auf jede 0 wieder eine 0. Dies ist nicht sehr zufällig und wird durch die Autokorrelation gemessen.

Die Autokorrelation ρ_n zu einer Verschiebung n ist nun ein Maß für das, was man in diesem Fall durch hinsehen erkennen kann. Autokorrelation bewegt sich stets zwischen den Werten -1 und 1, wobei -1 bedeutet, dass auf große Werte stets kleine folgen und umgekehrt, wohingegen 1 bedeutet, dass auf große (kleine) Werte folgen.

Autokorrelation von Zufallsreihen ist normal verteilt. Demnach kann man eine Wahrscheinlichkeit angeben, mit der die ermittelte Autokorrelation zufällig auftreten würde, also eine Wahrscheinlichkeit, mit der die Datenreihe zufällig bezüglich Autokorrelation wäre. Für die Reihe A beträgt diese Wahrscheinlichkeit weniger als 10⁻¹³%. Es kann also ziemlich sicher gesagt werden, dass diese Datenreihe nicht zufällig entstanden ist.