ST. PÖLTEN UNIVERSITY
OF APPLIED SCIENCES

Informatik
& Security

fh
st. pölten

Zufallsvariablen

Dipl.-Ing. Hubert Schölnast, BSc
Stand: 03. Juli 2023

Fachhochschule St. Pölten GmbH, Campus-Platz 1, 3100 St. Pölten, T: +43 (2742) 313 228, F: +43 (2742) 313 228-339, E: csc@fhstp.ac.at, I: www.fhstp.ac.at

Inhaltsverzeichnis

1 Einführung

In der Wahrscheinlichkeitsrechnung konnte das Ergebnis eins Zufallsergebnisses alles Mögliche sein. Beim Roulette waren die Ergebnisse »rot« und »schwarz« möglich, beim Münzwurf »Kopf« und »Zahl«, und eine gezogene Spielkarte konnte die Farbe ♥, ♦, ♠ oder ♣ haben.

Andere Ergebnisse waren Zahlen, oder sie waren Zahlen zumindest so ähnlich, dass leicht eine eindeutige Zuordnung möglich war. Ein typisches Beispiel sind die Ergebnisse beim Wurf eines Spielwürfels. Hier drängt sich die folgende Zuordnung geradezu auf:

⚀ $\widehat{=}$ 1

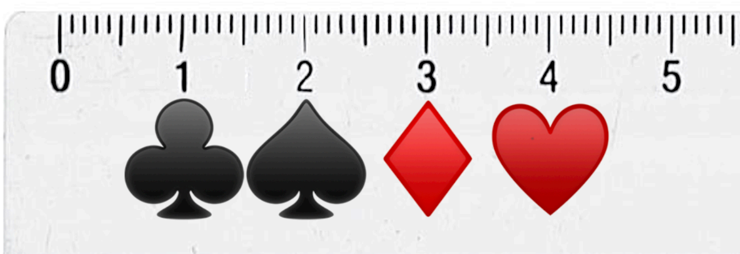
⚁ $\widehat{=}$ 2

⚂ $\widehat{=}$ 3

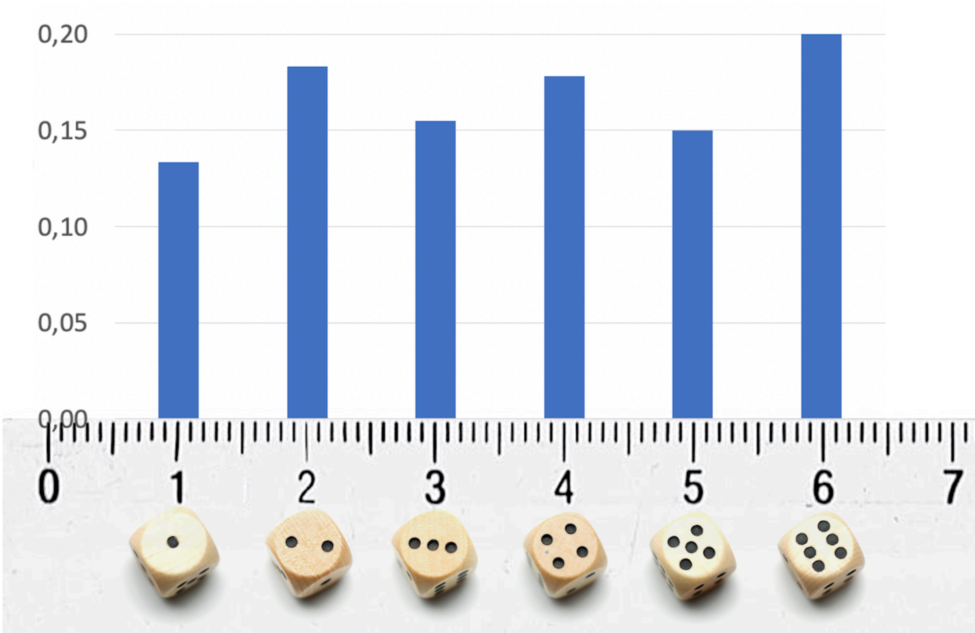
usw.

Prinzipiell ist aber eigentlich immer eine eindeutige Zuordnung zwischen einem Elementarereignis und einer reellen Zahl möglich, oft sind es sogar ganze und sehr häufig sogar natürliche Zahlen. Wenn man so eine Zuordnung definiert hat (oder wenn sie sich ohnehin fast von selbst aufdrängt), bekommen die Elementarereignisse, die bislang ja nur Elemente einer Menge waren, auch eine bestimmte Reihenfolge, und man kann sie in eben dieser Reihenfolge auf der reellen Zahlengeraden anordnen.

In den folgenden Beispielen werden die Augenzahlen eines Würfels den üblichen Zahlenwerten auf dem Zahlenstrahl zugeordnet. Bei den Münzwürfen wird Kopf=0 und Zahl=1 gesetzt, und die Spielkartenfarben werden in jener Reihenfolge den Werten 1 bis 4 zugeordnet, die ihrer Wertigkeit im Spiel Préférence entspricht.



Wenn man das gemacht hat, kann man über jedem Elementarereignis dessen Wahrscheinlichkeit auftragen.



Wahrscheinlichkeiten der Elementarereignisse eins nicht fairen Würfels.

Auf diese Weise kann die Wahrscheinlichkeit jedes Elementarereignisses als eine Funktion des Ereignisses angesehen werden.

Man kann nun einem Zufallsexperiment eine Zufallsvariable X zuordnen. Das Ergebnis des Zufallsexperiments ist der Wert dieser Zufallsvariablen. In diesem Kontext bezeichnet man ein Ergebnis auch als eine Realisierung des Zufallsexperiments. Die Wahrscheinlichkeit einer solchen Realisierung ist dann eine Funktion dieser Zufallsvariablen.

2 Diskrete Zufallsvariablen

Wenn es so ist, dass jede Realisierung einer Zufallsvariablen einen eindeutigen Vorgänger und einen eindeutigen Nachfolger hat (von den beiden Randwerten abgesehen), und wenn zwischen zwei solchen benachbarten Realisierungen keine weitere Realisierung liegt, dann nennt man diese Realisierung diskret. Alle bisher in diesem Skriptum gezeigten Beispiele (Würfel, Münzen, Karten) fallen in diese Kategorie.

Man spricht auch dann von diskreten Zufallsvariablen, wenn es nur einen oder gar keinen Randwert gibt, wenn dabei aber trotzdem die Eigenschaft der fehlenden Zwischenwerte aufrecht bleibt. In diesem Fall gibt es dann unendlich viele Realisierungen einer diskreten Zufallsvariablen. Dieser Fall tritt in der Praxis aber so gut wie nie auf, weil man bei diskreten Zufallsexperimenten, die »im wirklichen Leben« stattfinden, immer eine untere und eine obere Schranke für die Realisierungen finden kann. Daher wird dieser Fall hier nicht weiter behandelt. Wir gehen hier bei diskreten Zufallsvariablen immer von einer endlichen Anzahl aus.

Wenn es sowohl eine kleinste als auch eine größte Realisierung gibt (oder wenn es zumindest eine untere und eine obere Schranke gibt) ist die Menge der möglichen Realisierungen endlich.

2.1.1 **Notation**

Wenn eine Zufallsvariable X den Wert (die Realisierung) x_i annimmt, und die Wahrscheinlichkeit dieser Realisierung gleich p_i ist, schreibt man das so:

$$P(X = x_i) \ = \ p_i$$

»Die Wahrscheinlichkeit dafür, dass X den Wert x_i annimmt, ist gleich p_i.«

Es wäre natürlich auch denkbar, denselben Zusammenhang so zu schreiben: $P(x_i) = p_i$. Es gibt aber viele Fälle, in denen man Angaben braucht, die so ähnlich wie diese aussehen $P(x_1 < X < x_2) = p_i$ und weil man dann auf das X nicht mehr verzichten kann, schreibt man es immer in die runde Klammer.

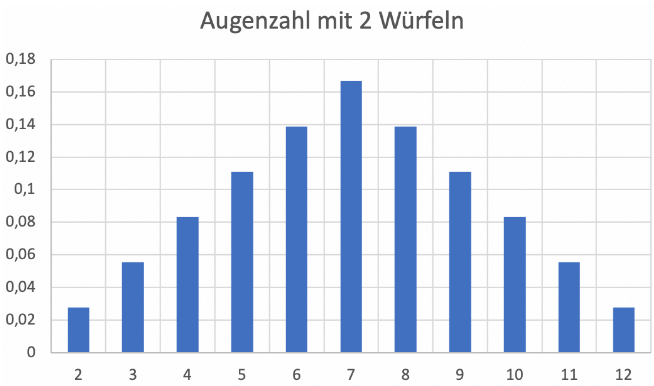
2.1.2 **Beispiel**

Das Zufallsexperiment sei der gleichzeitige Wurf zweier gewöhnlicher sechsseitiger Spielwürfel. Als Ergebnis wird die Augensumme der beiden Würfel gewertet. Dann sind die Wahrscheinlichkeiten einiger der 11 möglichen Realisierungen wie folgt zu notieren:

- $P(X = 2) = \frac{1}{36}$
- $P(X = 3) = \frac{2}{36}$
- $\dots$
- $P(X = 7) = \frac{6}{36}$
- $P(X = 8) = \frac{5}{36}$
- $\dots$
- $P(X = 12) = \frac{1}{36}$

2.2 Wahrscheinlichkeitsfunktion

Alle 11 Funktionswerte kann man schön in einem Diagramm (hier ein Säulendiagramm) darstellen:



Wahrscheinlichkeitsfunktion

Diese Wahrscheinlichkeiten der einzelnen Realisierungen entsprechen genau den relativen Häufigkeiten in der beschreibenden Statistik. Die Kolmogorow-Axiome 1 und 2 wurden auch mit Absicht genau so gewählt, dass sich genau dieser Zusammenhang ergibt.

2.3 Diskrete Verteilungsfunktion

Bei Zufallsvariablen spricht man aber nicht von einer Häufigkeitsverteilung, sondern von einer Wahrscheinlichkeitsverteilung, und dazu gehört eine Verteilungsfunktion. Diese Funktion gibt an, wie groß die Wahrscheinlichkeit dafür ist, bei einem Zufallsexperiment einen Wert zu erhalten, der kleiner oder gleich einer bestimmten Realisierung ist:

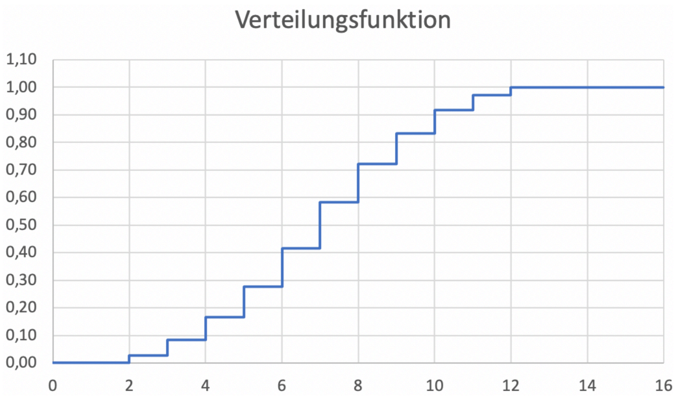
$$F(x) := P(X \leq x)$$

Wie man sich leicht überlegen kann, ist diese Definition äquivalent zu dieser Formulierung desselben Sachverhaltes:

$$F(x) := \sum_{i:\, x_i \leq x} p_i$$

Der Ausdruck »$i: x_i \leq x$«, der unter dem Summen-Operator steht, ist wie folgt zu lesen: »Summe über alle i für die der Wert von x_i kleiner oder gleich x ist.«

Beachte, dass der Wertebereich von x nicht auf die möglichen Realisierungen der Zufallsvariable beschränkt ist. x kann jede beliebige reelle Zahl sein.



Eine diskrete Verteilungsfunktion

2.3.1 Eigenschaften einer diskreten Verteilungsfunktion:

- Sie wächst monoton von 0 bis 1
- Sie hat Sprungstellen überall dort, wo es Realisierungen gibt, deren Wahrscheinlichkeit größer als 0 ist. Der Wert genau an der Sprungstelle ist immer der obere Wert (wegen des monotonen Wachstums ist das auch der Wert des rechts anschließenden horizontalen Teilstücks der Funktion).
- Zwischen 2 Sprungstellen ist der Funktionswert konstant
- Die Differenz zwischen unterem und oberem Grenzwert an jeder Sprungstelle ist genau die Wahrscheinlichkeit der jeweiligen Realisierung der Zufallsvariablen.

3 Stetige Zufallsvariablen

Wenn eine Zufallsvariable X, die die beiden Realisierungen x und $x+\varepsilon$ annehmen kann, in jedem Fall auch immer einen Wert irgendwo zwischen x und $x+\varepsilon$ haben kann, egal wie klein ε ist, dann ist diese Variable nicht diskret, sondern stetig. Das ist gleichbedeutend mit der Tatsache, dass die Werte, die X annehmen kann, alle reelle Zahlen innerhalb eines bestimmten Intervalls sind.

Das bedeutet aber auch, dass es unendlich viele Werte gibt, die X annehmen kann. Tatsächlich ist es sogar so, dass die Anzahl der möglichen Werte größer ist als die Anzahl der natürlichen Zahlen. Der Wertebereich ist also nicht nur unendlich, sondern sogar überabzählbar groß (siehe Skriptum Mengenlehre, Beweis $|\mathbb{R}| > |\mathbb{N}|$). Und damit ist das dritte Kolmogorow-Axiom (Skriptum Wahrscheinlichkeit, Kolmogorow-Axiome) nicht mehr anwendbar.

Was genau bedeutet das?

3.1 Beispiel Gleichverteilung

Sehen wir uns eine sehr einfache stetige Verteilung an, nämlich die Gleichverteilung innerhalb eines festen Intervalls auf der reellen Zahlengeraden. Der Einfachheit halber soll das Intervall von 0 bis 1 gehen. Das ist die Standard-Verteilung eines herkömmlichen Zufallszahlen-generators. Wir gehen in diesem Beispiel von der idealisierten Annahme aus, dass dieser Generator im mathematischen Sinne ideal funktioniert, also reelle Zahlen exakt ausgibt (d.h. mit unendlich vielen Nachkommastellen).

Überlegen wir uns, wie groß die Wahrscheinlichkeit dafür ist, dass dieser Generator eine ganz bestimmte Zahl im erlaubten Intervall ausspuckt, z.B. genau $\frac{\sqrt{2}}{2}$.

Nach wie vor ist die Wahrscheinlichkeit für das sichere Ereignis gleich 1, das heißt wir wissen mit Sicherheit, dass beim nächsten Aufruf des Generators eine Zahl zwischen 0 und 1 herauskommen wird. Es gibt aber überabzählbar viele mögliche Elementarereignisse. Deren Anzahl ist also größer als $\aleph_0$. ($\aleph_0$ist die Mächtigkeit der natürlichen Zahlen.) Wenn das dritte Axiom auch in diesem Fall gelten würde, wäre bei unserem Zufallszahlengenerator die Wahrscheinlichkeit für das Auftreten einer bestimmten Zahl gleich einer reellen Zahl α, für die folgendes gelten muss:[1]

$$\alpha < \frac{1}{\aleph_0} \quad \text{und daher auch} \quad \aleph_0 < \frac{1}{\alpha}$$

Der Kehrwert von α müsste also größer als unendlich sein, bzw. müsste α kleiner als jede positive reelle Zahl sein. Trotzdem müsste α aber auch größer als 0 sein. So eine Zahl gibt es in der Menge der reellen Zahlen aber nicht.[2]

Daher setzt man für stetige Zufallsvariable das dritte Kolmogorow-Axiom außer Kraft und definiert stattdessen, dass die Wahrscheinlichkeit dafür, dass eine stetige Zufallsvariable einen bestimmten Wert (z.B. n) annimmt, gleich 0 ist.

$$P(X = n) \; = \; 0$$

Das ist aber trotzdem kein unmögliches Ereignis! Es ist im Fall der stetigen Verteilungen aber trotzdem so, dass das unmögliche Ereignis gleich wahrscheinlich ist wie jedes einzelne mögliche Elementarereignis.

---

[1] Es mag naheliegend erscheinen, den Bruch $\frac{1}{\aleph_0}$ als $\frac{1}{\infty}$ zu schreiben, weil wir ja von unendlich vielen Werten sprechen, aber ∞ ist keine Zahl, sondern das Symbol dafür, dass eine Folge über alle Grenzen wächst. $\aleph_0$ ist hingegen eine ganz konkrete unendlich große Zahl. (Nur um Missverständnisse zu vermeiden: $\aleph_0$ ist zwar eine Zahl, aber keine reelle Zahl. $\aleph_0$ ist die kleinste nicht-endliche Kardinalzahl. $\aleph_0$ ist so etwas wie das rechte Ende der reellen Zahlengeraden.

[2] In der Klasse der surrealen Zahlen gibt es Zahlen mit solchen Eigenschaften. Das sind aber keine reellen Zahlen. Surrealen Zahlen bilden auch keine Menge, sondern eine Klasse, und jede von ihnen hat einen Geburtstag, der selbst auch wieder eine surreale Zahl ist und einen Geburtstag hat. Auch sonst sind surrealen Zahlen etwas merkwürdig.

Dieser Sachverhalt entspricht der Tatsache, dass man auch bei einem extrem schnell laufenden Zufallszahlengenerator unendlich lange warten muss, bis er irgendwann genau jene Zahl ausgibt, auf die man hofft. Aber jede Zahl, die er in dieser unendlich langen Zeit stattdessen ausgibt, während Sie auf die »richtige« Zahl warten, hatte genau dieselbe Wahrscheinlichkeit wie »Ihre« Zahl.

Daher macht es bei stetigen Verteilungen auch keinen Sinn, diese Elementarereignisse zu untersuchen. Sie treten zwar ein, aber mit einer Wahrscheinlichkeit von genau 0, und damit kann man nicht sinnvoll rechnen.

3.2 **Wahrscheinlichkeitsdichte statt Wahrscheinlichkeit**

Man kann bei stetigen Wahrscheinlichkeiten zwar nicht sinnvoll über Elementarereignisse reden, aber es gibt trotzdem Ereignisse mit einer Wahrscheinlichkeit, die größer als 0 ist. Das sichere Ereignis ist ein Beispiel für so ein Ereignis. Es tritt mit Sicherheit immer ein und hat die Wahrscheinlichkeit 1. Das sichere Ereignis ist aber darüber definiert, dass der Wert der Zufallsvariablen X irgendwo in einem Intervall liegt, das den ganzen Wertebereich der Verteilung umfasst.

Der wesentliche Punkt dabei ist das Wort »Intervall«. Sobald wir davon reden, dass die Zufallsvariable einen beliebigen Wert zwischen zwei Grenzwerten hat, können wir diesem Ereignis auch eine Wahrscheinlichkeit zuordnen, die größer als 0 sein kann:

$$P(X > x_1 \wedge X < x_2) \geq 0$$

Alternative und inhaltlich gleichwertige Schreibweise:

$$P(x_1 < X < x_2) \geq 0$$

(Überlegen Sie sich, warum es egal ist, ob man innerhalb der runden Klammer die Zeichen $>$ und $<$ oder $\geq$ und $\leq$ verwendet.)

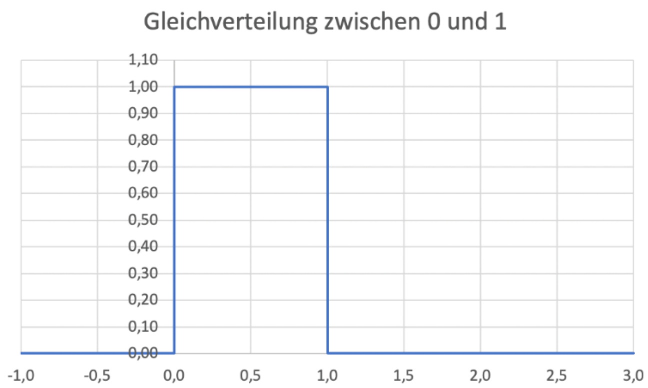
Wir können diese Intervalle (also die Differenz $x_2 - x_1$) beliebig klein machen, und können anschließend unterschiedliche, aber gleich breite Intervalle vergleichen. Wir werden dann bei stetigen Verteilungen (fast) immer feststellen, dass die Ereignisse, die diesen Intervallen entsprechen, unterschiedlich große Wahrscheinlichkeiten haben.

3.3 Beispiel Gleichverteilung, Wahrscheinlichkeitsdichtefunktion

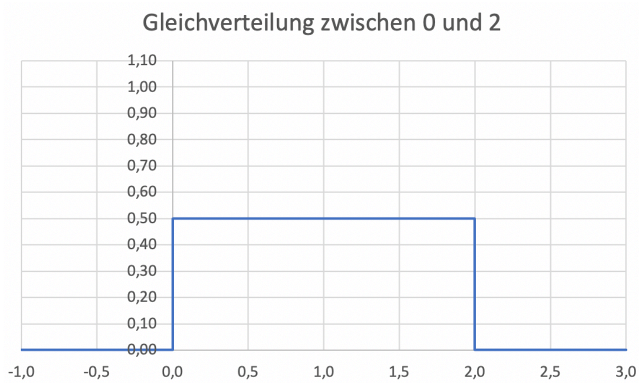
Lediglich die bereits weiter oben als Beispiel verwendete Gleichverteilung ist hier eine Ausnahme. Innerhalb des Intervalls von 0 bis 1 haben gleich breite Teil-Intervalle immer dieselbe Wahrscheinlichkeit.

Wenn die Breite des Intervalls 1 ist, ist die Wahrscheinlichkeit genau 1. Wenn die Breite $\frac{1}{2}$ ist, ist auch die Wahrscheinlichkeit, einen Wert in diesem Intervall zu erhalten, gleich $\frac{1}{2}$ usw.

Ganz allgemein stellt man bei der Gleichverteilung fest, dass, solange man im Intervall von 0 bis 1 bleibt, der Quotient aus Wahrscheinlichkeit und Intervallbreite immer den Wert 1 ergibt. Außerhalb des Intervalls $0 \cdots 1$ liegt immer das unmögliche Ereignis vor, und beliebig breite Intervalle außerhalb dieses Wertebereichs haben immer die Wahrscheinlichkeit 0. Das kann man schön in einem Diagramm darstellen:



Wenn aber der Wertebereich des Zufallszahlengenerators von 0 bis 2 reichen würde, müsste das Intervall von 0 bis 2, also ein Intervall mit der Breite 2 die Wahrscheinlichkeit 1 haben. Ein Intervall der Breite 1 hätte eine Wahrscheinlichkeit von $\frac{1}{2}$ usw. Der Quotient aus Wahrscheinlichkeit geteilt durch Breite hat in diesem Fall zwischen 0 und 2 den konstanten Wert $\frac{1}{2}$ und außerhalb dieses Intervalls den Wert 0:



Im ersten Fall (Intervall $0 \cdots 1$) waren die Wahrscheinlichkeiten also auf einen schmaleren Bereich komprimiert, in gewisser Weise also *dichter* als im zweiten Fall (Intervall $0 \cdots 2$). Daher nennt man den Wert, der hier in Richtung der Y-Achse aufgetragen wird, die Wahrscheinlichkeits**dichte**, und die Funktion, deren Werte in den beiden Diagrammen dargestellt sind, heißt Wahrscheinlichkeitsdichtefunktion.

So einfach, wie bei der Gleichverteilung, lässt sich die Wahrscheinlichkeitsdichte aber nicht immer bestimmen. Im allgemeinen Fall ist die Wahrscheinlichkeitsdichte durch einen Grenzwert bestimmt:

$$f(x) = \lim_{\varepsilon \to 0} \frac{P(x < X < x + \varepsilon)}{\varepsilon}$$

3.3.1 Eigenschaften der Wahrscheinlichkeitsdichtefunktion:

Die folgenden Eigenschaften gelten nicht nur für die Gleichverteilung, sondern für die Verteilungen aller stetigen Zufallsvariablen:

- Die Wahrscheinlichkeitsdichtefunktion ist auf ganz $\mathbb{R}$ definiert.
- Diese Funktion nimmt nirgendwo negative Werte an. Ihr Graph liegt daher immer auf oder oberhalb der X-Achse.
- Wenn der Wert der Realisierung gegen plus oder minus unendlich strebt, strebt die Wahrscheinlichkeitsdichte gegen 0.
$$\lim_{x \to \pm\infty} f(x) = 0$$
- Der Inhalt der Fläche, die zwischen der X-Achse und der Wahrscheinlichkeitsdichtefunktion eingeschlossen ist, ist genau 1.
$$\int_{-\infty}^{+\infty} f(x)dx = 1$$
- Die Funktion macht an keiner Stelle einen Sprung. Das ist gleichbedeutend mit der Tatsache, dass die Wahrscheinlichkeit einen ganz bestimmten einzelnen Wert anzunehmen, bei allen Werten gleich 0 ist.

3.4 Stetige Verteilungsfunktion

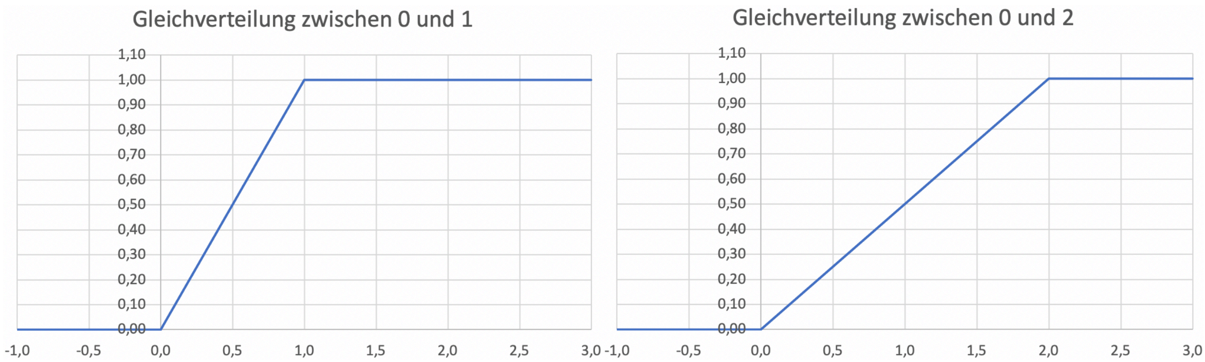
Die stetigen Verteilungsfunktionen sind genau gleich definiert wie die diskreten Verteilungs-funktionen:

$$F(x) := P(X \leq x)$$

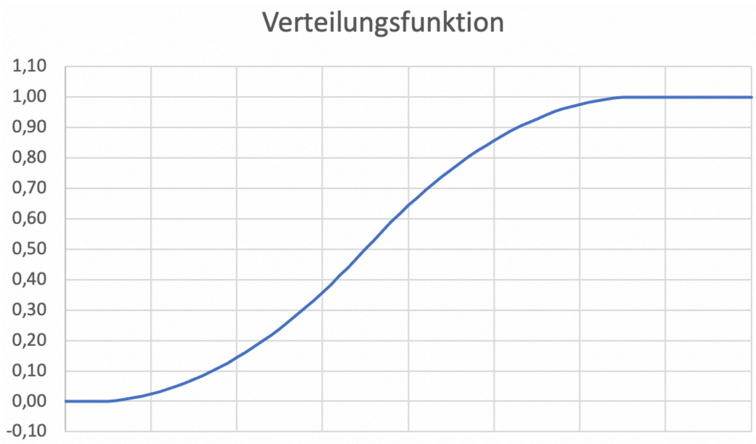
Die alternative Darstellung ist nun aber keine Summe mehr, sondern ein Integral:

$$F(x) := \int_{-\infty}^{x} f(u)du = \mathrm{p}$$

Dabei ist $f(u)$ genau die Wahrscheinlichkeitsdichtefunktion $f(x)$ (nur mit anders benanntem Argument).



Verteilungsfunktionen der beiden zuvor behandelten Gleichverteilungen



Eine typische stetige Verteilungsfunktionen hat immer ungefähr die hier abgebildete Form

3.4.1 **Eigenschaften einer stetigen Verteilungsfunktion:**

- Sie wächst monoton von 0 bis 1
- Sie hat nirgendwo Sprungstellen
- Nach links nähert sie sich immer weiter dem Funktionswert 0 oder nimmt links von einer bestimmten Stelle sogar überall den Wert 0 an
- Nach rechts nähert sie sich immer weiter dem Funktionswert 1 oder nimmt rechts von einer bestimmten Stelle sogar überall den Wert 1 an

4 Erwartungswert einer Wahrscheinlichkeitsverteilung

So, wie es in der beschreibenden Statistik bestimmte Parameter gab, welche die Lage oder die Breite (Streuung) einer Häufigkeitsverteilung beschrieben haben, gibt es auch Parameter, welche die Lage und Breite einer Wahrscheinlichkeitsverteilung beschreiben. Diese Parameter sind der Erwartungswert (für die Lage) und die Varianz bzw. Standardabweichung (für die Breite).

Der Erwartungswert einer Zufallsvariablen X gibt an, welchen Durchschnittswert man bei dem Zufallsexperiment erwarten kann, zu dem die Zufallsvariable X gehört. Der Erwartungswert ist das gewichtete arithmetische Mittel aller möglichen Realisierungen von X. Als Gewichte dienen dabei die Wahrscheinlichkeiten dieser Realisierungen.

Erinnern wir uns an die allgemeine Formel für das gewichtete arithmetische Mittel:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} x_i \cdot w_i}{\sum_{i=1}^{n} w_i}$$

Dabei sind die x_i jene Werte, von denen wir den Mittelwert haben wollen, und die w_i sind die dazugehörenden Gewichte.

Die x_i sind nun die möglichen Realisierungen von X und sie heißen auch weiterhin x_i.

Die w_i sind aber die Wahrscheinlichkeiten $p_i = P(X = x_i)$. Und die Summe all dieser Wahrscheinlichkeiten ist die Wahrscheinlichkeit des sicheren Ereignisses, also 1.

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{n} p_i = 1$$

Das heißt, dass in der Formel für das gewichtete arithmetische Mittel unter dem Bruchstrich die Zahl 1 steht, und wir daher gar keinen Bruch mehr schreiben müssen:

Erwartungswert einer diskreten Zufallsvariablen:

$$\mu = E(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

Im Fall einer stetigen Zufallsvariablen muss die Summe durch ein Integral ersetzt werden, und an die Stelle der Wahrscheinlichkeiten der Elementarereignisse tritt die Wahrscheinlichkeitsdichtefunktion.

Erwartungswert einer stetigen Zufallsvariablen:

$$\mu = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$$

Notation

Wie in den hervorgehobenen Formeln schon gezeigt, werden für den Erwartungswert die Symbole μ und $E(X)$ verwendet.

4.1.1 Beispiel diskret

An einem Spielwürfel wurde an einer Seite ein kleines Bleigewicht eingesetzt, daher erscheinen die sechs Augenzahlen mit unterschiedlichen Wahrscheinlichkeiten.[3]

- $p_1 = P(X = 1) = \frac{1}{10}$
- $p_2 = P(X = 2) = \frac{2}{15}$
- $p_3 = P(X = 3) = \frac{3}{20}$
- $p_4 = P(X = 4) = \frac{1}{6}$
- $p_5 = P(X = 5) = \frac{1}{5}$
- $p_6 = P(X = 6) = \frac{1}{4}$

$$\mu = \sum_{i=1}^{n} x_i \cdot p_i$$

$$\mu = 1 \cdot \frac{1}{10} + 2 \cdot \frac{2}{15} + 3 \cdot \frac{3}{20} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{5} + 6 \cdot \frac{1}{4}$$

$$\mu = \frac{1}{10} + \frac{4}{15} + \frac{9}{20} + \frac{4}{6} + 1 + \frac{6}{4}$$

$$\mu = \frac{6 + 16 + 27 + 40 + 60 + 90}{60}$$

$$\mu = \frac{239}{60} \cong 3{,}9833$$

Bei diesem Würfel beträgt der Durchschnittswert der zu erwartenden Augenzahlen knapp 4. Bei einem fairen Würfel liegt dieser Wert bei genau 3,5, was man anhand einer ähnlichen Rechnung leicht selbst nachrechnen kann.

---

[3] Einen Würfel, der so manipuliert wurde, dass nicht mehr alle 6 Augenzahlen gleich wahrscheinlich gewürfelt werden, bezeichnet man als »gezinkt«. Das Verb »zinken« bezog sich ursprünglich auf Spielkarten, die so bearbeitet worden sind, dass an den Ecken kleine Zacken (»Zinken«) hervorstanden. Falschspieler konnten dann durch Betasten der verdeckt liegenden Karten deren Wert erkennen. In weiterer Folge wurde jede Manipulation eines Glückspielgeräts als »zinken« bezeichnet.

4.1.2 Beispiel stetig

Sie kommen zu einer Schnellbahnstation und wissen nicht genau, wann die Züge fahren. Sie wissen aber, dass Schnellbahnzüge im gesamten Schnellbahnnetz in Intervallen von genau 15 Minuten fahren. Wie lange müssen Sie bis zur Abfahrt des nächsten Zuges warten, wenn Sie annehmen, dass die Züge pünktlich verkehren?

Die Dauer ihrer Wartezeit ist eine gleichverteilte stetige Zufallsvariable. Die kleinstmögliche Wartezeit beträgt 0 Minuten, jedoch müssen Sie niemals länger als 15 Minuten warten (weil die Züge laut Annahme immer pünktlich abfahren).

Sie wissen also:

- Die Wartezeit kann nicht negativ sein: $f(x) = 0$ für $x < 0$
- Die Wartezeit kann nicht größer als 15 sein: $f(x) = 0$ für $x > 15$
- Der Zug kommt ganz sicher innerhalb der nächsten 15 min: $\int_0^{15} f(x)\,dx = 1$
- Die Wahrscheinlichkeit dazwischen ist gleichverteilt, also $f(x) = k$ für $0 < x < 15$

$$\int_0^{15} k\,dx = 1$$

$$(kx + c)\,|_0^{15} = 1$$

$$(15k + c) - (0k + c) = 1$$

$$15k = 1$$

$$k = \frac{1}{15} = f(x)$$

Formel für den Erwartungswert einer stetigen Zufallsvariablen:

$$\mu = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$$

Man muss in diesem Beispiel aber gar nicht von $-\infty$ bis $+\infty$ integrieren, sondern nur von 0 bis 15:

$$\mu = \int_0^{15} x \cdot f(x)\,dx$$

Einsetzen: $f(x) = \frac{1}{15}$

$$\mu = \int_0^{15} x \cdot \frac{1}{15}\, dx$$

$$\mu = \left(\frac{1}{15} \cdot \frac{x^2}{2} + c\right)\Bigg|_0^{15}$$

$$\mu = \left(\frac{15^2}{15 \cdot 2} + c\right) - \left(\frac{0^2}{15 \cdot 2} + c\right)$$

$$\mu = \frac{15}{2} = 7{,}5$$

Der Erwartungswert beträgt 7,5 Minuten. Sie müssen im Schnitt jedes Mal 7 Minuten und 30 Sekunden lang warten.

4.2 Der Erwartungswert ist linear

Wenn X und Y zwei Zufallsvariablen sind, ist der Erwartungswert der Summe der Zufallsvariablen gleich der Summe der Erwartungswerte:

$$E(X + Y) = E(X) + E(Y)$$

Das gilt unabhängig davon, ob X und Y voneinander abhängig sind oder nicht.

»Die Bildung des Erwartungswertes ist invariant gegenüber der Addition.«

Der Erwartungswert einer Zufallsvariablen, die mit einer reellen Zahl multipliziert wurde, ist das Produkt aus dem ursprünglichen Erwartungswert und dieser reellen Zahl.

$$a \in \mathbb{R} \quad \Longrightarrow \quad E(a \cdot X) = a \cdot E(X)$$

»Die Bildung des Erwartungswertes ist invariant gegenüber der Multiplikation mit einem Skalar.«

Wenn sich ein Operator bei Addition und Multiplikation mit einem konstanten Wert so verhält wie eben beschrieben, wird dieser Operator als linear bezeichnet.

Die Linearität gilt sowohl für diskrete als auch für stetige Zufallsvariablen.

4.2.1 Beweis der Linearität

Stetige Zufallsvariablen können durch diskrete beliebig genau approximiert werden, daher wird hier auf den expliziten Beweis für stetige Zufallsvariablen verzichtet, könnte bei Bedarf aber analog zu den hier gezeigten Beweisen erfolgen.

Beginnen wir mit der Multiplikation mit einem Skalar, das geht in einer einzigen Zeile:

$$E(a \cdot X) \ = \ \sum_{i=1}^{n}(a \cdot x_i) \cdot p_i \ = \ \sum_{i=1}^{n} a \cdot (x_i \cdot p_i) \ = \ a \cdot \sum_{i=1}^{n}(x_i \cdot p_i) \ = \ a \cdot E(X)$$

□

Nun zur Summe. Was bedeutet es überhaupt, zwei Zufallsvariablen zu addieren? Wir haben

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

und

$$E(Y) = \sum_{j=1}^{m} y_j \cdot p_j$$

Erinnern wir uns daran, dass p_i nur eine Kurzschreibweise für $P(X = x_i)$ ist, und analog $p_j = P(Y = y_j)$.

Im Fall von $E(X + Y)$ wird hinter dem Summenzeichen sicherlich $(x_i + y_j)$ stehen, und das muss mit dieser Wahrscheinlichkeit multipliziert werden

$$P(X = x_i \wedge Y = y_j)$$

Dafür verwenden wir das Kürzel p_{ij}. Achtung! p_{ij} ist nicht dasselbe wie $p_i \cdot p_j$. Das wäre nur dann der Fall, wenn X und Y unabhängig voneinander wären. Wir wollen für den Beweis aber keinerlei Annahmen über die Abhängigkeit machen. Und weil X und Y ja ihre eigenen Laufvariablen und Wertebereiche für diese Laufvariablen haben, haben wir es mit der Summe einer Summe zu tun:

$$E(X + Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_j) \cdot p_{ij}$$

Ausmultiplizieren

$$E(X + Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot p_{ij} + y_j \cdot p_{ij}$$

In 2 Summen aufspalten

$$E(X + Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot p_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} y_j \cdot p_{ij}$$

Die folgenden Schritte werden hier nur für die erste Doppelsumme gezeigt, für die zweite sind exakt dieselben Schritte auszuführen

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot p_{ij} = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} x_i \cdot p_{ij}\right)$$

Innerhalb der Klammer den konstanten Faktor vor das Summenzeichen ziehen

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{m} x_i \cdot p_{ij}\right) = \sum_{i=1}^{n}\left(x_i \cdot \sum_{j=1}^{m} p_{ij}\right)$$

Erinnern wir uns, wofür p_{ij} eigentlich steht

$$\sum_{j=1}^{m} p_{ij} = \sum_{j=1}^{m} P\left(X = x_i \wedge Y = y_j\right)$$

Wenn man hier alle möglichen Werte von y_j durchläuft, erhält man für Y in Summe das sichere Ereignis. Es gilt also:

$$\sum_{j=1}^{m} p_{ij} = \sum_{j=1}^{m} P\left(X = x_i \wedge Y = y_j\right) = P(X = x_i) = p_i$$

Daher auch:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot p_{ij} = \sum_{i=1}^{n}\left(x_i \cdot \sum_{j=1}^{m} p_{ij}\right) = \sum_{i=1}^{n} x_i \cdot p_i = E(X)$$

Wie gesagt gilt das alles ganz analog auch für den zweiten Summanden in

$$E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i \cdot p_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} y_j \cdot p_{ij}$$

Und daher haben wir:

$$E(X+Y) = E(X) + E(Y)$$

was zu beweisen war.

□

4.2.2 Beispiel

Sie haben einen gewöhnlichen sechsseitigen Spielwürfel und einen weiteren Wurfstein in der Form eines regelmäßigen Oktaeders. Der Würfel kann die Augenzahlen 1 bis 6 zeigen, der Oktaeder die Zahlen von 1 bis 8. Beide Wurfsteine sind fair (die Augenzahlen eines Steins sind jeweils gleich wahrscheinlich) und werden zugleich geworfen. Wie groß ist der Erwartungswert dieses Spiels?

Berechnen wir zunächst die Erwartungswerte der beiden Steine, wenn sie alleine geworfen werden.

Formel:

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

Würfel:

$$E(W) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{i=1}^{6} i = \frac{1}{6} \cdot \frac{(6+1) \cdot 6}{2} = \frac{7}{2} = 3{,}5$$

Oktaeder:

$$E(O) = \sum_{i=1}^{8} i \cdot \frac{1}{8} = \frac{1}{8} \cdot \sum_{i=1}^{8} i = \frac{1}{8} \cdot \frac{(8+1) \cdot 8}{2} = \frac{9}{2} = 4{,}5$$

Beide zugleich:

$$E(W+O) = E(W) + E(O) = 3{,}5 + 4{,}5 = 8$$

Überprüfen wir das mit einer anderen Methode. Wir ermitteln alle Elementarereignisse und deren Wahrscheinlichkeiten und berechnen damit den Erwartungswert.

Die folgende Tabelle enthält alle möglichen Augenzahl-Kombinationen.

O / *W*	1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8	9
2	3	4	5	6	7	8	9	10
3	4	5	6	7	8	9	10	11
4	5	6	7	8	9	10	11	12
5	6	7	8	9	10	11	12	13
6	7	8	9	10	11	12	13	14

Es gibt 48 mögliche Kombinationen, die alle gleich wahrscheinlich sind. (Es sind Laplace-Experimente.) Gleiche Augensummen kann man zusammenfassen.

$$E(W+O) = 2\cdot\frac{1}{48} + 3\cdot\frac{2}{48} + 4\cdot\frac{3}{48} + 5\cdot\frac{4}{48} + 6\cdot\frac{5}{48} + 7\cdot\frac{6}{48} + 8\cdot\frac{6}{48} + 9\cdot\frac{6}{48} + 10\cdot\frac{5}{48} + 11\cdot\frac{4}{48} + 12\cdot\frac{3}{48} + 13\cdot\frac{2}{48} + 14\cdot\frac{1}{48}$$

$$E(W+O) = \frac{2+6+12+20+30+42+48+54+50+44+36+26+14}{48}$$

$$E(W+O) = \frac{384}{48} = 8$$

Die Probe geht auf, die Rechnung brachte also erwartungsgemäß das richtige Resultat.

4.3 Multiplikations-Invarianz

Wenn X und Y zwei **unabhängige** Zufallsvariablen sind, ist der Erwartungswert des Produkts der Zufallsvariablen gleich dem Produkt der Erwartungswerte:

$$E(X\cdot Y) = E(X)\cdot E(Y)$$

Dieser Zusammenhang gilt **nicht**, wenn X und Y voneinander abhängig sind.

»Die Bildung des Erwartungswertes ist bei **unabhängigen** Zufallsvariablen invariant gegenüber der Multiplikation.«

5 Varianz und Standardabweichung

Zur Erinnerung: Bei den Häufigkeitsverteilungen hat man von den Einzelwerten jeweils den Mittelwert der gesamten Verteilung abgezogen, diese Differenz quadriert und dann diese quadrierten Werte aufsummiert. Das Ergebnis ist die Varianz. Wenn man daraus dann noch die Wurzel zieht, hat man die Standardabweichung.

Und genau so macht man das auch bei den diskreten Wahrscheinlichkeitsverteilungen, wobei die Rolle des Mittelwerts vom Erwartungswert eingenommen wird. Bei den stetigen Wahrscheinlichkeitsverteilungen muss man von Summen zu Integralen übergehen, der Rest erfolgt ganz analog.

Bei Wahrscheinlichkeitsverteilungen bezeichnet man die Varianz manchmal auch als den *Erwartungswert für die mittlere quadratische Abweichung vom Erwartungswert.*

Varianz einer diskreten Zufallsvariablen X:

$$\mathrm{VAR}(X) = \sigma^2 = E((X-\mu)^2) = \sum_{i=1}^{n}(x_i-\mu)^2 \cdot p_i$$

Varianz einer stetigen Zufallsvariablen X:

$$\mathrm{VAR}(X) = \sigma^2 = E((X-\mu)^2) = \int_{-\infty}^{+\infty}(x-\mu)^2 \cdot f(x)\,dx$$

Standardabweichung einer Zufallsvariablen X:

$$\sigma = \sqrt{\mathrm{VAR}(X)} = \sqrt{E((X-\mu)^2)}$$

Achtung!

$$\sqrt{E((X-\mu)^2)} \neq E(X-\mu)$$

Bitte selbst nachdenken, warum das nicht dasselbe ist. (Welchen Wert hat denn der Erwartungswert für $X-\mu$ in jeder beliebigen Verteilung? Was ist denn μ?)

5.1.1 Entartete Wahrscheinlichkeitsverteilung

Die Varianz einer Wahrscheinlichkeitsverteilung (und damit auch die Standardabweichung) ist genau dann gleich 0, wenn die Zufallsvariable nur eine einzige Realisierung hat. Das ist gleichbedeutend damit, dass das »Zufalls«-Experiment immer dasselbe Ergebnis bringt, also gar keine Zufälligkeit enthält. Dann ist jedes Ergebnis immer mit dem Erwartungswert identisch, und nur unter dieser Voraussetzung ist die Differenz zwischen dem Ergebnis eines Zufallsexperiments und seinem Erwartungswert immer 0. Eine Wahrscheinlichkeitsverteilung, die so beschaffen ist, nennt man »entartet«.

5.1.2 **Beispiel 1**

Varianz und Standardabweichung eines fairen sechsseitigen Spielwürfels

Formel

$$\mathrm{VAR}(X) = \sigma^2 = E((X-\mu)^2) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$$

Den Wert für μ haben wir schon auf Seite 17 ausgerechnet, dort hieß dieser Wert $E(W)$.

$$\mu = E(W) = 3{,}5 = \frac{7}{2}$$

Die Summe ausschreiben und dabei μ einsetzen:

$$\sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$$

$$= \left(1-\frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2-\frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(3-\frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(4-\frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(5-\frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(6-\frac{7}{2}\right)^2 \cdot \frac{1}{6}$$

$$\mathrm{VAR}(X) = \left(\left(\frac{2-7}{2}\right)^2 + \left(\frac{4-7}{2}\right)^2 + \left(\frac{6-7}{2}\right)^2 + \left(\frac{8-7}{2}\right)^2 + \left(\frac{10-7}{2}\right)^2 + \left(\frac{12-7}{2}\right)^2\right) \cdot \frac{1}{6}$$

$$\mathrm{VAR}(X) = \left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) \cdot \frac{1}{6}$$

$$\mathrm{VAR}(X) = \frac{70}{24} = \frac{35}{12} \cong 2{,}9167$$

Die Standardabweichung ist die Wurzel aus der Varianz

$$\sigma = \sqrt{\mathrm{VAR}(X)} = \sqrt{\frac{35}{12}} = \sqrt{\frac{35 \cdot 12}{12 \cdot 12}} = \frac{\sqrt{35 \cdot 12}}{\sqrt{12 \cdot 12}} = \frac{\sqrt{35 \cdot 3 \cdot 4}}{12} = \frac{\sqrt{35 \cdot 3} \cdot \sqrt{4}}{12} = \frac{\sqrt{105} \cdot 2}{12} = \frac{\sqrt{105}}{6}$$

$$\sigma \cong 1{,}7078$$

5.1.3 Beispiel 2

Schnellbahnhaltestelle

Formel

$$\mathrm{VAR}(X) = \sigma^2 = E((X-\mu)^2) = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\, dx$$

Den Wert für μ haben wir auf Seite 14 berechnet:

$$\mu = \frac{15}{2} = 7{,}5$$

In Formel einsetzen:

$$\mathrm{VAR}(X) = \int_{-\infty}^{+\infty} \left(x - \frac{15}{2}\right)^2 \cdot f(x)\, dx$$

Wir können von Seite 14 auch übernehmen, dass wir nur von 0 bis 15 integrieren müssen, und dass $f(x)$ in diesem Intervall den konstanten Wert $\frac{1}{15}$ hat:

$$\mathrm{VAR}(X) = \int_{0}^{15} \left(x - \frac{15}{2}\right)^2 \cdot \frac{1}{15}\, dx$$

Vereinfachen des Integranden:

$$\mathrm{VAR}(X) = \int_{0}^{15} \frac{(2x-15)^2}{4} \cdot \frac{1}{15}\, dx$$

$$\mathrm{VAR}(X) = \int_{0}^{15} \frac{4x^2 - 2 \cdot 15 \cdot 2x + 15^2}{60}\, dx$$

$$\mathrm{VAR}(X) = \frac{1}{60} \cdot \int_{0}^{15} 4x^2 - 60x + 225\, dx$$

Integrieren

$$\mathrm{VAR}(X) = \frac{1}{60} \cdot \left(\frac{4}{3}x^3 - 30x^2 + 225x + c\right)\Bigg|_0^{15}$$

wieder ausmultiplizieren

$$\mathrm{VAR}(X) = \frac{1}{45}x^3 - \frac{1}{2}x^2 + \frac{15}{4}x + \frac{1}{60}c\Bigg|_0^{15}$$

Grenzen einsetzen

$$VAR(X) = \left(\frac{3375}{45} - \frac{225}{2} + \frac{225}{4} + \frac{1}{60}c\right) - \left(\frac{1}{45} \cdot 0^3 - \frac{1}{2} \cdot 0^2 + \frac{15}{4} \cdot 0 + \frac{1}{60}c\right)$$

$$VAR(X) = 75 - \frac{225}{2} + \frac{225}{4} = \frac{300 - 450 + 225}{4} = \frac{75}{4} = 18{,}75$$

Standardabweichung durch Wurzelziehen

$$\sigma = \sqrt{\mathrm{VAR}(X)} = \sqrt{\frac{75}{4}} = \frac{\sqrt{75}}{\sqrt{4}} = \frac{\sqrt{3 \cdot 25}}{2} = \frac{\sqrt{3} \cdot \sqrt{25}}{2} = \sqrt{3} \cdot \frac{5}{2} \cong 4{,}3301$$

5.2 Momente einer Wahrscheinlichkeitsverteilung

Den Ausdruck $E((X-\mu)^2)$ über den die Varianz definiert ist, nennt man auch *»zweites zentrales Moment«* der Verteilung, und $E(X)$ (das ist gleich μ) nennt man *»erstes Moment«*. Ganz allgemein gilt:

nulltes Mom.	$m_0(X) = E(1) = 1$
1. Moment	$m_1(X) = E(X) = \mu$
2. Moment	$m_2(X) = E(X^2) = \sigma^2 + \mu^2$
3. Moment	$m_3(X) = E(X^3)$
n-tes Mom.	$m_n(X) = E(X^n)$

nulltes zentr. M.	$m_0(X-\mu) = E(1) = 1$
1. zentr. M.	$m_1(X-\mu) = E(X-\mu) = 0$
2. zentr. M.	$m_2(X-\mu) = E((X-\mu)^2) = \sigma^2$
3. zentr. M.	$m_3(X-\mu) = E((X-\mu)^3)$
n-tes zentr. M.	$m_n(X-\mu) = E((X-\mu)^n)$

Und auch bei Wahrscheinlichkeitsverteilung kann man die höheren Momente verwenden, um mehr über die Form der Verteilung in einfache Zahlen fassen zu können:

5.2.1 Schiefe

$$\gamma = \frac{m_3(X-\mu)}{\sigma^3}$$

5.2.2 Kurtosis (Wölbung)

$$\kappa = \frac{m_4(X-\mu)}{\sigma^4}$$

Bei Schiefe und Kurtosis gilt alles, was bereits in den entsprechenden Kapiteln über Häufigkeitsverteilungen gesagt worden ist.

5.3 Verhalten bei linearer Transformation

Wenn man eine Wahrscheinlichkeitsverteilung X mit bekannter Varianz bzw. Standardabweichung nach folgender Formel linear transformiert

$$Y = a \cdot X + b \qquad a, b \in \mathbb{R}$$

Dann gilt für die Varianz

$$VAR(Y) \ = \ a^2 \cdot VAR(X)$$

und für die Standardabweichung

$$\sigma_y \ = \ |a| \cdot \sigma_x$$

Auf den ersten Blick erscheint es vielleicht seltsam, dass die transformierte Varianz überhaupt nicht von dem Wert b abhängt, aber es leuchtet ein, wenn man sich klar macht, dass Varianz und Standardabweichung die Breite einer Verteilung messen, eine Verteilung aber nicht breiter oder schmäler wird, wenn man sie auf der Zahlengeraden einfach nur nach links oder rechts verschiebt.

5.4 Standardisierte Zufallsvariablen

Wir haben nun Formeln kennengelernt, die es uns ermöglichen, beliebige Verschiebungen und Streckungen auf Wahrscheinlichkeitsverteilungen anzuwenden und dann einfach zwischen den Erwartungswerten und Varianzen dieser Wahrscheinlichkeitsverteilungen umrechnen zu können. Das können wir ausnützen, um aus jeder beliebigen Verteilung eine neue Verteilung zu machen, die den Erwartungswert 0 hat. Wenn die Verteilung nicht entartet ist, wird es auch immer gelingen, sie so zu stauchen oder strecken, dass die Standardabweichung den Wert 1 bekommt.

Wenn man eine nicht-entartete Wahrscheinlichkeitsverteilung X hat und davon bereits den Erwartungswert μ und die Standardabweichung σ kennt, kann man mit folgender Formel

$$Z = \frac{X - \mu_x}{\sigma_x}$$

daraus die Wahrscheinlichkeitsverteilung Z machen. Für diese standardisierte Wahrscheinlichkeitsverteilung Z gilt

$$\mu_z = 0 \qquad \sigma_z = 1$$

Umgekehrt kann man aus einer standardisierten Wahrscheinlichkeitsverteilung Z mit folgender Formel immer eine beliebige Wahrscheinlichkeitsverteilung X machen, bei der μ und σ die vorgegebenen Werte μ_x und σ_x annehmen:

$$X = \sigma_x \cdot Z + \mu_x$$

Die Verteilungsfunktionen der beiden Verteilungen kann man wie folgt ineinander umrechnen:

$$F_X(x) = F_Z\left(\frac{x-\mu_x}{\sigma_x}\right)$$

$$F_Z(z) = F_X(\sigma_x \cdot x + \mu_x)$$

Und auch für die Wahrscheinlichkeitsdichtefunktion stetiger Verteilungen gibt es eine Umrechnungsformel:

$$f_X(x) = \frac{1}{\sigma_x} \cdot f_Z\left(\frac{x-\mu_x}{\sigma_x}\right)$$

Warum sollte man das überhaupt machen wollen?

In den folgenden Kapiteln werden mehrere spezielle Wahrscheinlichkeitsverteilungen vorgestellt, und das wird viel einfacher, wenn man einfach immer nur auf den Fall $\mu = 0;\ \sigma = 1$ untersuchen muss. Mit den Formeln auf dieser Seite kann man sich daraus jede beliebige Verteilung basteln.

5.5 Varianz der Summe zweier Wahrscheinlichkeitsverteilungen (Kovarianz)

Wenn man zwei Wahrscheinlichkeitsverteilungen addiert, erhält man den Erwartungswert dieser Summe indem man die Erwartungswert der beiden ursprünglichen Verteilungen addiert. Für die Varianz gilt diese Summenregel nur bei **unabhängigen** Variablen.

Wenn X und Y zwei **unabhängige** Zufallsvariablen sind, ist die Varianz der Summe der Zufallsvariablen gleich der Summe der Varianzen:

$$\mathrm{VAR}(X+Y) \quad = \quad \mathrm{VAR}(X) + \mathrm{VAR}(Y)$$

Wenn X und Y voneinander **abhängig** sind, kommt ein weiterer Summand hinzu, nämlich die Kovarianz:

$$\mathrm{COV}(X,Y) \quad = \quad E\Big((X-\mu_x)\cdot(Y-\mu_y)\Big) \quad = \quad E(X\cdot Y) - \mu_x \cdot \mu_y$$

Dieser Summand muss zweimal hinzuaddiert werden (einmal als $\mathrm{COV}(X,Y)$, das zweite Mal umgekehrt als $\mathrm{COV}(Y,X)$, weil das wegen der Kommutativität der Multiplikation aber immer dasselbe ergibt, erscheinen diese beiden Ausdrücke gemeinsam als $2\cdot\mathrm{COV}(X,Y)$ in der folgenden Formel.

Wenn X und Y zwei Zufallsvariablen sind, die auch voneinander abhängig sein dürfen, ist die Varianz der Summe der Zufallsvariablen gleich der Summe der Varianzen plus der doppelten Covarianz:

$$\mathrm{VAR}(X+Y) \quad = \quad \mathrm{VAR}(X) + \mathrm{VAR}(Y) + 2 \cdot \mathrm{COV}(X,Y)$$

Man kann diesen Sachverhalt auch verwenden, um herauszufinden, ob zwei Wahrscheinlichkeitsverteilungen voneinander linear abhängig sind:

5.5.1 **Korrelation**

Wenn $\mathrm{COV}(X,Y) = 0$, dann sind X und Y linear unkorreliert. Linear unkorreliert zu sein ist eine schwächere Bedingung als unabhängig zu sein. Es gibt Wahrscheinlichkeitsverteilungen, die voneinander abhängig sind, bei denen diese Abhängigkeit aber keine lineare Komponente hat. In diesem Fall ist $\mathrm{COV}(X,Y) = 0$, obwohl die beiden Verteilungen voneinander abhängig sind.

Wenn $\mathrm{COV}(X,Y) > 0$ spricht man von einer positiven Korrelation. Wegen $\mathrm{COV}(X,Y) = E(X \cdot Y) - \mu_x \cdot \mu_y$ gilt dann:

$$E(X \cdot Y) > \mu_x \cdot \mu_y$$

5.5.2 **Maximale Kovarianz**

Die Kovarianz ist dann maximal groß, wenn man eine der beiden Verteilungen durch eine lineare Transformation aus der anderen erhalten kann. (Das schließt den Fall ein, dass eine der beiden Verteilungen entartet ist.) Es gilt immer diese Ungleichung:

$$\mathrm{COV}(X,Y)^2 \leq VAR(X) \cdot VAR(Y)$$

Dabei gilt die Gleichheit nur genau bei linearer Abhängigkeit:

$$\mathrm{COV}(X,Y)^2 = VAR(X) \cdot VAR(Y) \quad \Leftrightarrow \quad Y = a \cdot X + b$$

5.6 **Tschebyschowsche Ungleichung**[4]

Manchmal möchte man, ohne sich groß den Kopf über die genaue Verteilung zu zerbrechen, abschätzen, wie groß die Wahrscheinlichkeit dafür ist, dass die Realisierung einer Zufallsvariablen innerhalb eines bestimmten Intervalls liegt. Es wird vorausgesetzt, dass μ und σ bekannt sind, und dass das Intervall genau um den Erwartungswert zentriert ist. Die Breite dieses Intervalls sei k. Dann gilt diese Ungleichung:

$$P(|X-\mu|<k) \;\;\geq\;\; 1-\frac{\sigma^2}{k^2}$$

andere Schreibweise:

$$P(\mu-k \;<\; X \;<\; \mu+k) \;\;\geq\;\; 1-\frac{\sigma^2}{k^2}$$

In Worten: »Die Wahrscheinlichkeit dafür, dass die Zufallsvariable X eine Realisierung hat, die zwischen $\mu-k$ und $\mu+k$ liegt, ist mindestens so groß wie $1-\frac{\sigma^2}{k^2}$«.

Diese Grenze gilt immer, für jede beliebige Verteilungen, egal welche Form sie hat. Das hat aber bei den meisten Verteilungen zur Folge, dass der Ausdruck auf der linken Seite der Ungleichung erheblich größer als der Wert auf der rechten Seite ist.

5.6.1 **Beispiel 1**

Nehmen wir an, dass Bücher im Schnitt 300 Seiten haben und dass die Standardabweichung für die Seitenanzahl 50 ist. Wie groß ist die Untergrenze für die Wahrscheinlichkeit, dass ein beliebiges Buch zwischen 200 und 400 Seiten hat?

$$\mu=300;\quad \sigma=50;\quad k=100$$

$$P(|X-300|<100) \;\;\geq\;\; 1-\frac{50^2}{100^2}=1-\frac{2500}{10000}=1-\frac{1}{4}=\frac{3}{4}$$

Die Wahrscheinlichkeit dafür ist größer als 75%.

---

[4] Pafnuti Lwowitsch Tschebyschow (Пафнутий Львович Чебышёв) (1821-1894) war ein russischer Mathematiker, der viele Beiträge zu Zahlentheorie, Wahrscheinlichkeitsrechnung, Funktionen, Mechanik und Ballistik publiziert hat. Sein Nachname wird nach aktuellem Standard zu *Čebyšëv* transliteriert und in einem deutschen Kontext zu *Tschebyschow* transkribiert (letzte Silbe betont). Ältere, heute nicht mehr gebräuchliche deutsche Transkriptionen sind Tschebyschef, Tschebyscheff, Tschebyschew und Tschebyschev. In englischen Text findet man oft die Schreibweise Chebyshev. Viele seiner Werke erschienen zuerst aber auf französisch, daher findet man oftmals auch in deutschen Texten die französische Transkription *Tchebychef.*

5.6.2 Beispiel 2

Selbes Beispiel wie zuvor, wir wollen aber wissen, wie groß die Wahrscheinlichkeit dafür ist, bei einer zufälligen Wahl ein Buch zu erwischen, das zwischen 275 und 325 Seiten hat.

$$\mu = 300; \quad \sigma = 50; \quad k = 25$$

$$P(|X - 300| < 25) \;\; \geq \;\; 1 - \frac{50^2}{25^2} = 1 - \frac{2500}{625} = 1 - 4 = -3$$

Die Wahrscheinlichkeit dafür ist größer als -3. Nachdem wir aber das erste Kolmogorow-Axiom kennen, das ohnehin schon aussagt, dass alle Wahrscheinlichkeiten größer als 0 sind, hat uns diese Rechnerei also keine neuen Erkenntnisse gebracht.

Wie man sich leicht klar machen kann, ist der Grenzwert, den man bei der Tschebyschow-schen Ungleichung erhält, nur dann größer als 0, wenn $k > \sigma$. Aber selbst, wenn k deutlich größer als σ ist, liefert die Tschebyschowsche Ungleichung nur eine sehr, sehr vorsichtige Abschätzung. Dafür funktioniert diese Abschätzung auch ohne irgend etwas über die genaue Form der Verteilung zu wissen.

6 Gesetz der großen Zahlen

Bleiben wir beim fairen sechsseitigen Spielwürfel. Wir kennen den Erwartungswert für einen einfachen Wurf und auch die Standardabweichung dieses Zufallsexperiments.

Nun werfen wir denselben Würfel ein zweites Mal, interessieren uns nun aber für den Mittelwert der Augensummen der beiden Würfe. Dieses Experiment ist, bis auf die zeitliche Abfolge, die uns aber nicht interessiert, mit dem gleichzeitigen Wurf zweier Würfel identisch.

Das ist so, weil die beiden zugleich geworfenen Würfel, während sie fallen, nicht vom jeweils anderen Würfel beeinflusst werden, und weil der zweimal hintereinander geworfene Würfel weder ein Gedächtnis hat, noch Informationen über zukünftige Würfe zur Verfügung hat. In beiden Durchführungsarten sind die beiden Würfel also voneinander unabhängig.

Halten wir zwei Kriterien fest, die wichtig sind:

- Die einzelnen Zufallsexperimente sind voneinander unabhängig (engl.: independent)
- Für die einzelnen Zufallsexperimente gilt immer dieselbe Wahrscheinlichkeitsverteilung (engl.: identically distributed)

Weil diese beiden Eigenschaften so wichtig sind, und man in wissenschaftlichen Arbeiten oft auf deren Gültigkeit verweisen muss, hat sich aus der englischen Floskel »independent and identically distributed« die Abkürzung »iid« als Standard-Bezeichnung für diese beiden Kriterien etabliert.

Beim Wurf zweier Würfel (egal ob zugleich oder hintereinander) erhalten wir diese Augenzahlensummen:

W_1 \ W_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Der Mittelwert der Augenzahlen der beiden Würfe ist bei 2 Würfen jeweils die Hälfte:

W_1 \ W_2	1	2	3	4	5	6
1	1	1,5	2	2,5	3	3,5
2	1,5	2	2,5	3	3,5	4
3	2	2,5	3	3,5	4	4,5
4	2,5	3	3,5	4	4,5	5
5	3	3,5	4	4,5	5	5,5
6	3,5	4	4,5	5	5,5	6

Der Erwartungswert für diesen Mittelwert errechnen wir wie folgt:

$$E(\overline{X}) = \sum_{i=1}^{n} \overline{x}_i \cdot p_i$$

$$E(\overline{X}) = 1 \cdot \frac{1}{36} + \frac{3}{2} \cdot \frac{2}{36} + 2 \cdot \frac{3}{36} + \frac{5}{2} \cdot \frac{4}{36} + 3 \cdot \frac{5}{36} + \frac{7}{2} \cdot \frac{6}{36} + 4 \cdot \frac{5}{36} + \frac{9}{2} \cdot \frac{4}{36} + 5 \cdot \frac{3}{36} + \frac{11}{2} \cdot \frac{2}{36} + 6 \cdot \frac{1}{36}$$

$$E(\overline{X}) = \frac{1}{36} + \frac{3}{36} + \frac{6}{36} + \frac{10}{36} + \frac{15}{36} + \frac{21}{36} + \frac{20}{36} + \frac{18}{36} + \frac{15}{36} + \frac{11}{36} + \frac{6}{36}$$

$$E(\overline{X}) = \overline{\mu} = \frac{126}{36} = \frac{7}{2} = 3{,}5$$

Der Erwartungswert für den Mittelwert zweier Würfe ist also mit dem Erwartungswert für einen einzelnen Wurf identisch.

Die Varianz berechnen so:

$$\mathrm{VAR}(\overline{X}) = \sum_{i=1}^{n} (\overline{x}_i - \overline{\mu})^2 \cdot p_i$$

$$\mathrm{VAR}(\overline{X}) = \left(1-\frac{7}{2}\right)^2 \cdot \frac{1}{36} + \left(\frac{3}{2}-\frac{7}{2}\right)^2 \cdot \frac{2}{36} + \left(2-\frac{7}{2}\right)^2 \cdot \frac{3}{36} + \left(\frac{5}{2}-\frac{7}{2}\right)^2 \cdot \frac{4}{36} + \left(3-\frac{7}{2}\right)^2 \cdot \frac{5}{36} + \left(\frac{7}{2}-\frac{7}{2}\right)^2 \cdot \frac{6}{36} + \left(4-\frac{7}{2}\right)^2 \cdot \frac{5}{36} + \left(\frac{9}{2}-\frac{7}{2}\right)^2 \cdot \frac{4}{36} + \left(5-\frac{7}{2}\right)^2 \cdot \frac{3}{36} + \left(\frac{11}{2}-\frac{7}{2}\right)^2 \cdot \frac{2}{36} + \left(6-\frac{7}{2}\right)^2 \cdot \frac{1}{36}$$

$$\mathrm{VAR}(\overline{X}) = \left(\frac{-5}{2}\right)^2 \cdot \frac{1}{36} + \left(\frac{-4}{2}\right)^2 \cdot \frac{2}{36} + \left(\frac{-3}{2}\right)^2 \cdot \frac{3}{36} + \left(\frac{-2}{2}\right)^2 \cdot \frac{4}{36} + \left(\frac{-1}{2}\right)^2 \cdot \frac{5}{36} + \left(\frac{0}{2}\right)^2 \cdot \frac{6}{36} + \left(\frac{1}{2}\right)^2 \cdot \frac{5}{36} + \left(\frac{2}{2}\right)^2 \cdot \frac{4}{36} + \left(\frac{3}{2}\right)^2 \cdot \frac{3}{36} + \left(\frac{4}{2}\right)^2 \cdot \frac{2}{36} + \left(\frac{5}{2}\right)^2 \cdot \frac{1}{36}$$

$$\mathrm{VAR}(\overline{X}) = \frac{25\cdot 1}{144} + \frac{16\cdot 2}{144} + \frac{9\cdot 3}{144} + \frac{4\cdot 4}{144} + \frac{1\cdot 5}{144} + \frac{0\cdot 6}{144} + \frac{1\cdot 5}{144} + \frac{4\cdot 4}{144} + \frac{9\cdot 3}{144} + \frac{16\cdot 2}{144} + \frac{25\cdot 1}{144}$$

$$\mathrm{VAR}(\overline{X}) = \frac{25+32+27+16+5+0+5+16+27+32+25}{144}$$

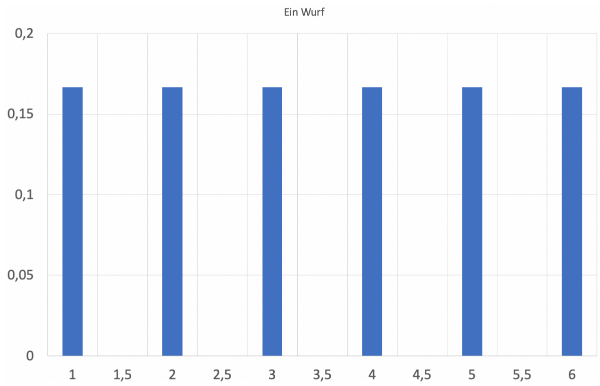
$$\mathrm{VAR}(\overline{X}) = \frac{210}{144} = \frac{35}{24} \cong 1{,}4583$$

Erinnern wir uns an die Varianz beim Wurf eines einzelnen Würfels (Seite 20):

$$\mathrm{VAR}(X) = \frac{35}{12} \cong 2{,}9167$$

Die Varianz für den Mittelwert der Ergebnisse zweier gleicher unabhängiger Zufallsexperimente ist also die Hälfte der Varianz der Einzelexperimente.

Der Vergleich der beiden Wahrscheinlichkeitsfunktionen macht diesen Befund plausibel:



Wahrscheinlichkeitsfunktionen für einen Wurf eines Würfels und für zwei Würfe

Bei zwei Würfeln sind die Werte am Rand viel weniger wahrscheinlich als in der Mitte. Die einzelnen Ergebnisse liegen im Durchschnitt also näher beim Mittelwert als bei der Verteilung, die dem Wurf eines einzelnen Würfels entspricht.

Ganz allgemein gilt für alle Zufallsexperimente, die wiederholt durchgeführt werden, wobei die Wiederholungen jeweils unabhängig von allen Vorgängern und Nachfolgern sind:

Der Erwartungswert für das arithmetische Mittel aller Ergebnisse bei der Durchführung von n gleichen voneinander unabhängigen Zufallsexperimenten ist immer identisch mit dem Erwartungswert für die Durchführung eines Einzelexperiments:

$$E(\overline{X}) = E(X)$$

Die Varianz für das arithmetische Mittel aller Ergebnisse bei der Durchführung von n gleichen voneinander unabhängigen Zufallsexperimenten ist der n-te Teil der Varianz für die Durchführung eines Einzelexperiments:

$$VAR(\overline{X}) = \overline{\sigma}^2 = \frac{VAR(X)}{n}$$

Um die Standardabweichung des arithmetischen Mittelwerts aller Ergebnisse bei der Durchführung von n gleichen voneinander unabhängigen Zufallsexperimenten zu erhalten, muss man die Standardabweichung des Einzelexperiments durch die Wurzel der Anzahl der Wiederholungen dividieren:

$$\overline{\sigma} = \sqrt{VAR(\overline{X})} \quad = \frac{\sigma}{\sqrt{n}}$$

Daraus folgt unmittelbar:

Je öfter man ein Zufallsexperiment wiederholt, desto geringer wird die Streuung des arithmetischen Mittels aller Ereignisse. Mit anderen Worten: Je öfter man ein Experiment wiederholt, desto geringer wird die relative Abweichung vom Erwartungswert.

Wenn die Anzahl der Wiederholungen (n) gegen unendlich strebt, muss σ^2 gegen 0 streben:

$$\lim_{n \to \infty} \frac{VAR(X)}{n} = 0 = \overline{\sigma}^2$$

Betrachten wir unter diesem Gesichtspunkt nochmals die Tschebyschowsche Ungleichung:

$$P(|X-\mu|<k) \;\; \geq \;\; 1-\frac{\sigma^2}{k^2}$$

Wenn n gegen unendlich strebt, strebt der Bruch $\frac{\sigma^2}{k^2}$ gegen 0, egal wie klein wir k wählen. Der Ausdruck rechts vom Ungleichheitszeichen strebt also gegen 1.

Da aber (zweites Kolmogorow-Axiom) keine Wahrscheinlichkeit größer als 1 sein kann, wird dann aus dem Ungleichheitszeichen ein Gleichheitszeichen. Wir erhalten also:

$$\lim_{n\to\infty} P(|X-\mu|<k) \;\; = \;\; 1$$

Weil das für jedes noch so kleine k gilt, führt das zum ...

Gesetz der großen Zahlen:

Wenn ein beliebiges Zufallsexperiment oft genug wiederholt wird, wobei die Wiederholungen voneinander unabhängig sein müssen, wird sich das arithmetische Mittel der Ergebnisse der Experimente beliebig nahe an den Erwartungswert annähern. Es gilt für jede noch so kleine Zahl ε:

$$\lim_{n\to\infty} P\left(\left|\overline{X}-\mu\right|<\varepsilon\right) \;\; = \;\; 1$$

»Mit wachsendem Stichprobenumfang konvergiert der Mittelwert der Ergebnisse der Zufallsexperimente (das ist $\overline{X}$) stochastisch gegen den Erwartungswert μ.«

Beachte, dass das streng genommen nur gilt, wenn die Anzahl der Experimente unendlich groß wird. Bei einer endlichen Anzahl an Experimenten hat jeder mögliche Wert für die Differenz $\overline{X}-\mu$ eine Wahrscheinlichkeit, die größer als 0 ist. Was der Satz aussagt, ist folgendes:

Je mehr gleiche und voneinander unabhängige Experimente man wiederholt, desto geringer wird die Wahrscheinlichkeit, dass der gemessene Mittelwert der Ergebnisse (das ist $\overline{X}$) weit vom Erwartungswert μ entfernt ist.

6.1 Satz von Bernoulli

Das Gesetz der großen Zahlen geht auf Jakob I Bernoulli[5] zurück, der es aber selbst ein wenig anders formuliert hat:

Satz von Bernoulli

Man führt ein Zufallsexperiment durch bei dem das Ereignis A mit der Wahrscheinlichkeit p eintritt, und wiederholt das Experiment sehr oft. Es sei f_n die relative Häufigkeit für das Eintreffen von A nach n Wiederholungen. Dann gilt für jedes noch so kleine ε:

$$\lim_{n\to\infty} P(|f_n - p| < \varepsilon) \;=\; 1$$

Je öfter man ein Experiment durchführt, desto größer ist die Wahrscheinlichkeit dafür, dass der Anteil der erfolgreichen Durchführungen nahe an der Wahrscheinlichkeit für das Eintreten von A ist.

6.2 Fundamentalsatz der Statistik

Betrachtet man diesen Zusammenhang von einem anderen Blickwinkel aus, erkennt man etwas, das unsere intuitive Vorstellung von Wahrscheinlichkeiten prägt:

Nicht immer kann man aus theoretischen Überlegungen direkt auf eine absolut genaue Wahrscheinlichkeitsverteilung schließen. Daher versucht man die Wahrscheinlichkeiten für die einzelnen Realisierungen durch empirische Untersuchungen zu ermitteln. Das heißt: Man führt das Zufallsexperiment, dessen Wahrscheinlichkeitsverteilung man ermitteln möchte, so oft wie möglich durch und notiert die relativen Häufigkeiten, mit denen die Realisierungen auftreten.

Man erhält dadurch eine relative Häufigkeitsverteilungsfunktion, die man »empirische Verteilungsfunktion« nennt. Unsere Intuition sagt uns, dass diese empirische Verteilungsfunktion umso besser mit der wahren, aber prinzipiell unbekannten Wahrscheinlichkeitsverteilungsfunktion übereinstimmt, je mehr Zufallsexperimente wir durchführen.

---

[5] Jakob I Bernoulli (1655-1705) war ein Schweizer Mathematiker und Physiker, der wegweisende Arbeiten zur Wahrscheinlichkeitstheorie und Variationsrechnung verfasst hat. Gemeinsam mit seinem Bruder Johann I Bernoulli hat er Leibnitz's Infinitesimalrechnung erweitert und bekannt gemacht. Ebenfalls sehr bekannt ist sein Neffe Daniel Bernoulli (1700-1782), den man vor allem für seine Arbeiten zur Strömung von Gasen und Flüssigkeiten kennt (Bernoulli-Gleichung). Wenn es um Wahrscheinlichkeiten geht, ist mit »Bernoulli« aber stets Jakob I Bernoulli gemeint. (Es gibt auch noch einen Jakob II Bernoulli und einen Johann II Bernoulli, die ebenfalls Schweizer Mathematiker waren.)

Der Satz von Bernoulli bzw. das Gesetz der großen Zahlen erlauben es nun, diesen intuitiven Zusammenhang auch mathematisch exakt zu formulieren:

Fundamentalsatz der Statistik (Satz von Gliwenko-Cantelli):

Wenn ein Zufallsexperiment mit der tatsächlichen Verteilungsfunktion $F(x)$ n-mal wiederholt durchgeführt wird, wobei alle Durchführungen unabhängig von allen anderen sind, erhält man eine empirische Verteilungsfunktion $\overline{F}(x)$. Wenn dabei die Anzahl der Wiederholungen über alle Grenzen wächst, gilt für jedes x und für jedes noch so kleine ε:

$$\lim_{n\to\infty} P\left(\left|\overline{F}(x) - F(x)\right| < \varepsilon\right) \; = \; 1$$

»Die empirische Verteilungsfunktion $\overline{F}(x)$ konvergiert stochastisch gegen die tatsächlichen Verteilungsfunktion $F(x)$.«

Erst dieser Fundamentalsatz macht es überhaupt möglich, dass man anhand empirischer Untersuchungen Schätzungen für Verteilungsfunktionen erzeugen kann.