

Wie wir gesehen haben, ist die bisher behandelte LAPLACE-Wahrscheinlichkeit für komplexere Anwendungen nicht ausreichend: Mehr als endlich viele mögliche Ergebnisse und fehlende Modelle bzw. Einsichten zur Beurteilung der Gleichwahrscheinlichkeit führen rasch zu einem Versagen des bisherigen mathematischen Modells.

A. N. KOLMOGOROFF zeigte, dass sich alle Eigenschaften des klassischen Wahrscheinlichkeitsbegriffs aus einigen Grundeigenschaften, den *Wahrscheinlichkeitsaxiomen*, ableiten lassen und so eine Verallgemeinerung des Wahrscheinlichkeitsbegriffs möglich ist.

Glühlampe: Für deren Lebensdauer kommen unendlich viele Ergebnisse in Frage, es gibt (noch?) kein physikalisches Modell, das eine Beurteilung der wahrscheinlichen Lebensdauer erlaubt – hier hilft nur die Beobachtung!

Axiom: Eine (Grund-) Tatsache, die ohne Beweis (-möglichkeit) als *wahr* angenommen wird.

Axiome der Wahrscheinlichkeitsrechnung:

- *Axiom 1:* Die Wahrscheinlichkeit (für das Eintreten) eines zufälligen Ereignisses E ist eine reelle Zahl $P(E)$ mit

$$0 \leq P(E) \leq 1$$

- *Axiom 2:* Die Wahrscheinlichkeit des Sicheren Ereignisses S ist gleich eins:

$$P(S) = 1$$

- *Axiom 3:* Für *abzählbar unendlich viele* und paarweise unvereinbare Ereignisse E_1, E_2, \dots ergibt sich die Wahrscheinlichkeit für das Eintreten eines dieser Ereignisse als Summe der Einzelwahrscheinlichkeiten:

$$P(E_1 \cup E_2 \cup \dots) = P(E_1) + P(E_2) + \dots$$

Bemerkung: Durch die Axiome wird nichts darüber ausgesagt, *wie* man ein „vernünftiges“ Wahrscheinlichkeitsmaß für ein Zufallsexperiment erhält! Sie ermöglichen nur ein Urteil darüber, ob ein Wahrscheinlichkeitsmaß vorliegen kann.

abzählbar unendlich: Die Ereignisse lassen sich mittels der natürlichen Zahlen (durch-) nummerieren.

Bemerkung: Unendlich viele Ergebnisse führen dazu, dass einzelne Ereignisse die Wahrscheinlichkeit null (bzw. eins) haben können, ohne dass sie deshalb das Unmögliche (bzw. Sichere) Ereignis sein müssen! Auf die damit verbundene Problematik kann und wird hier nicht eingegangen, die im Folgenden behandelten Begriffe (z. B. Zufallsgröße) werden daher auch nicht mit „letzter mathematischer Strenge“ eingeführt.

Jemand soll – zufällig – eine ganze Zahl nennen. Dass dabei gerade 433 ausgewählt wird, ist bestimmt nicht unwahrscheinlich, obwohl die Wahrscheinlichkeit für diese Wahl gleich null ist.

Statistische (bzw. empirische) Wahrscheinlichkeit

Wird ein Zufallsversuch (beliebig) oft durchgeführt, so nähern sich die beobachteten relativen Häufigkeiten $h(E)$ eines Ereignisses E immer mehr einem bestimmten Wert, die Schwankungen werden immer kleiner: *Empirisches Gesetz der großen Zahlen*.

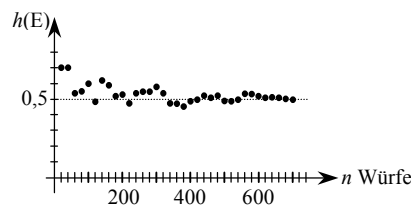
Da dieser (statistische Grenz-) Wert bei LAPLACE-Versuchen der klassischen Wahrscheinlichkeit entspricht, wählt man die relativen Häufigkeiten als *Schätzwerte* für die gesuchte Wahrscheinlichkeit:

$$P(E) \approx h(E)$$

Bemerkung: Umgekehrt interpretiert man daher auch die Wahrscheinlichkeit als Häufigkeit. Hat ein Ereignis die Wahrscheinlichkeit $P(E)$, dann erwartet man, dass dieses Ereignis bei n Versuchen ungefähr $n \cdot P(E)$ – mal auftreten wird.

Beispiel: Werfen einer Münze
E: „Zahl liegt oben“

Bestimmt werden sollen die relativen Häufigkeiten nach 20, 40 usw. Würfeln.

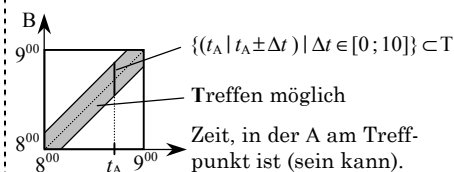


Geometrische Wahrscheinlichkeit

Sind bei einem Zufallsexperiment unendlich viele Ergebnisse möglich, so kann versucht werden, diese auf Kurven, Flächen usw. eindeutig abzubilden. Haben dann alle Punkte dieser Kurve (Fläche) die gleiche Chance, gewählt zu werden, so lassen sich die Wahrscheinlichkeiten durch die Verhältnisse von Teillängen (bzw. Teilflächen) zur Gesamtlänge (Gesamtfläche) angeben.

Beispiel: Anton und Berta möchten sich zwischen 8⁰⁰ und 9⁰⁰ Uhr treffen. Sie vereinbaren, jeweils 10 Minuten auf einander zu warten. Mit welcher Wahrscheinlichkeit kommt es zu einem Treffen, wenn keiner vor 8⁰⁰ erscheinen bzw. nicht länger als bis 9⁰⁰ Uhr warten kann?

T: „A und B treffen einander“, also
 $T = \{(t_A | t_B) | t_A, t_B \in [8; 9] \wedge |t_A - t_B| \leq \frac{1}{6}\}$



→ $P(T) = 1 - \frac{25}{36} = \frac{11}{36}$

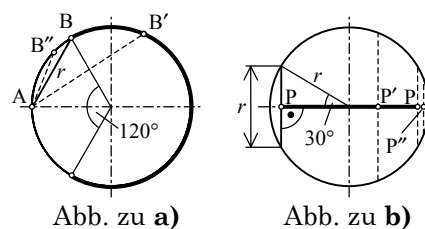
Achtung: Bei geometrischen Wahrscheinlichkeiten kann es sein, dass die Wahrscheinlichkeitsbestimmung nicht eindeutig ist, da durch verschiedene Ansätze, welche als durchaus „vernünftig“ angesehen werden können, unterschiedliche Ergebnisse möglich sind, wie das untenstehende Problem zeigt.

Beispiel: In einem Kreis mit dem Radius r soll eine Sehne gezogen werden. Wie groß ist die Wahrscheinlichkeit dafür, dass die Sehne länger als der Radius ist (Ereignis E)?

Für die Berechnung betrachten wir zwei Möglichkeiten, wie diese Sehne festgelegt werden könnte (jede scheint „richtig“ zu sein):

- a) Die Sehne wird durch einen festen (A) und einen zufällig gewählten Punkt (B, B', B'', ...) gelegt.
- b) Die Sehne wird durch einen Punkt (P, P', P'', ...) eines festgelegten Durchmessers und normal zu diesem gezeichnet.

Die Punkte B', P' liegen günstig, die Punkte B'', P'' ungünstig für das betrachtete Ereignis E. B und P markieren die Grenzlagen.



a) $P_a(E) = \frac{240^\circ}{360^\circ} = \frac{2}{3}$

b) $P_b(E) = \frac{2 \cdot r \cos 30^\circ}{2r} = \frac{\sqrt{3}}{2}$

Es gilt aber $P_a(E) \neq P_b(E) !!$

Neben den Wahrscheinlichkeiten spielen auch deren *Verteilungen* eine große Rolle. Dabei zeigt sich, dass bei Zufallsexperimenten unterschiedlicher Art (z. B. Wägen von Massen – Messen von Längen bzw. Entdecken von Ausschussteilen – Auftreten von Knabengeburt) nicht nur ähnliche Verteilungen, sondern auch ähnliche Fragestellung von Bedeutung sind. Dies führt in einem gewissen Sinn zu einer *Standardisierung* und letztlich zu einer Beschränkung auf einige wenige Verteilungen und deren Kenngrößen (Letztere sind die wahrscheinlichkeitstheoretischen Entsprechungen der in der beschreibenden Statistik verwendeten Begriffe wie z. B. Merkmale, Häufigkeitsverteilungen, Mittelwert oder empirische Standardabweichung).

Für ein Textilunternehmen ist es wichtig zu „wissen“, welche Mengen in den einzelnen Konfektionsgrößen erzeugt werden sollen, um die Nachfrage zu decken, ohne dass es dabei zu Überproduktionen (den späteren „Restposten“) kommt.

Zufallsvariable, Verteilungsfunktion

a) Eine Funktion X , die jedem Ergebnis eines Zufallsversuchs eine reelle Zahl zuordnet, heißt *Zufallsgröße*.

Mit $X=a$ wird ein Ereignis beschrieben, das alle Ergebnisse umfasst, denen durch die Funktion X der Wert a zugeordnet wird. Wir sprechen daher kurz vom „Ereignis $X=a$ “ (Analoges gilt für $X \leq a$ usw.). Die Wahrscheinlichkeit $P(E)$ wird dabei als die „Wahrscheinlichkeit des Ereignisses $X=a$ “ bezeichnet und als $P(X=a)$ geschrieben.

Übereinkunft: Zufallsgrößen bezeichnen wir mit Großbuchstaben ($X, Y \dots$), ihre (Funktions-) Werte mit kleinen Buchstaben.

Bemerkungen:

- Ist b keine Zahl aus dem Wertebereich von X , so ist $X=b$ ein Unmögliches Ereignis und es gilt $P(X=b)=0$.
- Die Werte einer Zufallsgröße bilden eine neue Ereignismenge S_X , die als „Vergrößerung“ der alten aufgefasst werden kann.

b) Als *Verteilungsfunktion* der Zufallsgröße X bezeichnet man die Funktion G mit

$$G : \mathbb{R} \rightarrow [0; 1]$$

$$x \mapsto G(x) := P(X \leq x)$$

Bemerkung: Man beachte, dass Verteilungsfunktionen auf ganz \mathbb{R} erklärt sind!

c) Eine Zufallsgröße heißt *diskret*, wenn sie nur endlich viele oder abzählbar unendlich viele verschiedene Werte annehmen kann.

Eine Zufallsgröße ist *stetig*, wenn ihre Verteilungsfunktion stetig ist.

Beispiel: Werfen von zwei Münzen
Die Funktion X sei so definiert, dass sie den Wurfergebnissen die Anzahl zuordnet, wie oft Zahl (Z) zu sehen ist: $X(WW)=0, X(ZW)=X(WZ)=1, X(ZZ)=2$

$X=1$ bedeutet daher das Ereignis
A: „genau eine Münze zeigt Zahl“,
 $A = \{ZW; WZ\}$

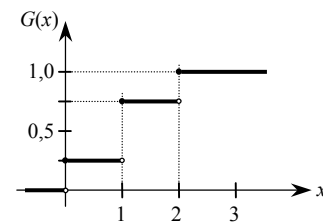
$$\rightarrow P(X=1) = P(A) = \frac{1}{2}$$

$X \geq 1$ beschreibt folgendes Ereignis
B: „auf mindestens einer Münze ist Zahl zu sehen“, $B = \{ZW; WZ; ZZ\}$

$$\rightarrow P(X \geq 1) = P(B) = \frac{3}{4}$$

$$S = \{WW; ZW; WZ; ZZ\}$$

$$S_X = \{0; 1; 2\}$$



X ist eine diskrete Zufallsvariable.

Das Gewicht eines zufällig gewählten Säuglings ist (wäre) eine stetige Zufallsgröße.

Bemerkungen:

- Zufallsgrößen sind Funktionen, daher können mit ihnen auch Vielfache, Summen und Produkte ($k \cdot X, X+Y, X \cdot Y \dots$) gebildet werden, vorausgesetzt, diese sind für die gleiche Ergebnismenge erklärt.
- Diskrete Zufallsgrößen werden vor allem bei (Ab-) Zählvorgängen angewendet, bei Messgrößen benutzt man häufig stetige Zufallsvariablen (genau genommen liegen beim Messen durch die Verwendung gerundeter Messwerte auch diskrete Verteilungen vor; die meist feinen Unterteilungen des Messintervalls liefert jedoch eine hinreichende Begründung für eine (näherungsweise) Beschreibung durch stetige Größen).

Beispiel: Werfen von zwei Würfeln

$X_{1(2)}$: „Augenzahl von Würfel 1 (2)“,

$$\text{z. B. } X_1((3; 5))=3, X_2((3; 5))=5$$

$$Y := 2X_1 \rightarrow Y((3; 5))=6$$

$$Z := X_1 + X_2^2 \rightarrow Z((3; 5))=28$$

Wird z. B. das Gewicht von Säuglingen in Gramm angegeben, so ist die Größe „Gewicht eines zufällig ausgewählten Säuglings“ eine diskret verteilte Zufallsgröße, weil nur endlich viele (Mess-) Werte vorkommen können.

- Verteilungsfunktionen entsprechen der Summenhäufigkeit der beschreibenden Statistik. Bei diskreten Zufallsvariablen werden sie daher häufig durch Treppenkurven oder Histogramme dargestellt.
- Berechnung von (oft benötigten) Wahrscheinlichkeiten mit Hilfe von Verteilungsfunktionen:

$$P(X > a) = 1 - P(X \leq a) = 1 - G(a)$$

$$P(a < X \leq b) = G(b) - G(a)$$

$X > a$ ist das Gegenereignis von $X \leq a$ und damit sind $X \leq a$ und $a < X \leq b$ unvereinbare Ereignisse:

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsdichte

- a) Sei X eine diskrete Zufallsgröße. Als *Wahrscheinlichkeitsfunktion* bzw. *Verteilung* von X bezeichnet man die Funktion g mit

$$g : \mathbb{R} \rightarrow [0; 1]$$

$$x \mapsto g(x) := P(X = x)$$

Bemerkung: Analog zur Summenhäufigkeit folgt für eine diskret verteilte Zufallsgröße X mit den Werten x_i die Darstellung von $G(x)$, $x_n \leq x < x_{n+1}$ durch

$$G(x) = \sum_{i=1}^n g(x_i)$$

- b) Sei X eine stetige Zufallsgröße. Existiert für die Verteilungsfunktion G eine Funktion g von $\mathbb{R} \rightarrow [0; 1]$ mit

$$\int_{-\infty}^{\infty} g(t) \cdot dt = 1$$
 und

$$G(x) = \int_{-\infty}^x g(t) \cdot dt,$$

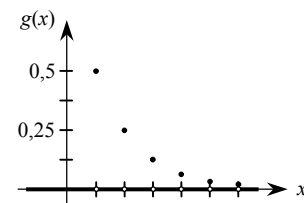
so nennt man g die *Wahrscheinlichkeitsdichte* bzw. *Dichtefunktion* von X .

Bemerkungen:

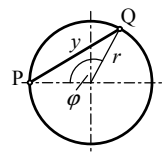
- Da für stetige Zufallsgrößen $P(X = a) = 0$ für alle $a \in \mathbb{R}$ gilt, so gilt auch $P(X \leq a) = P(X < a)$.
- Ein Funktionswert $G(a)$ der Verteilungsfunktion G einer stetigen Zufallsgröße entspricht dem Inhalt der Fläche zwischen der Dichtekurve und der x -Achse im Intervall $[-\infty; a]$. Dementsprechend ergibt sich die Wahrscheinlichkeit $P(a \leq X \leq b)$ als Flächeninhalt im Intervall $[a; b]$.
- Die für stetige Zufallsgrößen naheliegende Flächeninterpretation lässt sich auf diskrete Größen übertragen: Wird die Wahrscheinlichkeitsfunktion g durch ein Histogramm dargestellt, so entsprechen den Werten der Verteilungsfunktion G die Summen der Rechtecksflächen.

Beispiel 1: Werfen einer Münze
 X : „Anzahl der Würfe, bis erstmalig Wappen (W) oben liegt“

$$x_i = 1; 2; 3; \dots \text{ und } P(X = x_i) = (\frac{1}{2})^i$$



Beispiel 2: Gegeben ist ein Kreis mit dem Radius r . Von einem festen Kreispunkt P wird eine Sehne zu einem zufällig ausgewählten Kreispunkt Q gezogen.



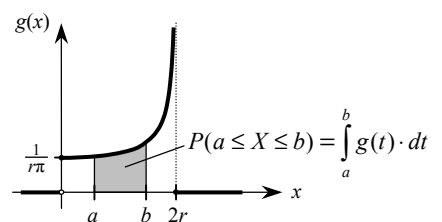
Y : „Länge der Sehne PQ “;

für die Verteilungsfunktion gilt:

$$G(y) = \begin{cases} 0 & y < 0 \\ \frac{2\phi}{2\pi} = \frac{2}{\pi} \cdot \arcsin \frac{y}{2r} & y \in [0; 2r] \\ 1 & y > 2r \end{cases}$$

und damit für die Dichtefunktion:

$$g(y) = \begin{cases} \frac{2}{\pi \cdot \sqrt{4r^2 - y^2}} & y \in [0; 2r[\\ 0 & \text{sonst} \end{cases}$$



Erwartungswert, Varianz und Standardabweichung einer Zufallsgröße

(g ist die Wahrscheinlichkeitsfunktion bzw. -dichte der Zufallsgröße X)

a) Der Erwartungswert $E(X)$ einer Zufallsvariablen X ist

- die Summe aller Produkte $x_k \cdot g(x_k)$, wenn X diskret ist:

$$\mu = E(X) := \sum_{k=1}^n x_k \cdot g(x_k)$$

bzw. der Grenzwert dieser Summe, wenn X abzählbar unendlich viele Werte x_k besitzt.

- bei einer stetigen Größe das uneigentliche Integral

$$\mu = E(X) := \int_{-\infty}^{\infty} t \cdot g(t) \cdot dt$$

b) Der Erwartungswert der Zufallsvariablen $(X - E(X))^2$ ist die Varianz $V(X)$ der Zufallsgröße X .

- Ist X diskret, so gilt

$$\sigma^2 = V(X) := \sum_{k=1}^n (x_k - \mu)^2 \cdot g(x_k) \quad \mu = E(X)$$

bzw. für abzählbar unendlich viele Werte x_k

$$\sigma^2 = V(X) := \lim_{n \rightarrow \infty} \sum_{k=1}^n (x_k - \mu)^2 \cdot g(x_k)$$

- Für eine stetige Größe ergibt sich

$$\sigma^2 = V(X) := \int_{-\infty}^{\infty} (t - \mu)^2 \cdot g(t) \cdot dt \quad \mu = E(X)$$

c) Die Quadratwurzel aus der Varianz $V(X)$ ist die Standardabweichung (bzw. Streuung) σ der Zufallsgröße X :

$$\sigma := \sqrt{V(X)}$$

Fortsetzung von Beispiel 1:

$$E(X) = \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k = \sum_{k=0}^{\infty} k \cdot \left(\frac{1}{2}\right)^k = 2$$

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{i}{x^i} &= x \cdot \sum_{i=1}^{\infty} \frac{i}{x^{i+1}} = -x \cdot \left(\sum_{i=1}^{\infty} \frac{1}{x^i} \right)' = \\ &= -x \cdot \left(\frac{1}{1 - \frac{1}{x}} \right)' = \frac{x}{(x-1)^2} \end{aligned}$$

$|x| > 1$
Grenzwert einer geom. Reihe

Fortsetzung von Beispiel 2:

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} t \cdot g(t) \cdot dt = \lim_{a \rightarrow 2r} \frac{2}{\pi} \int_0^a \frac{t \cdot dt}{\sqrt{4r^2 - t^2}} = \\ &= \lim_{a \rightarrow 2r} \frac{2}{\pi} \left[-\sqrt{4r^2 - t^2} \right]_0^a = \frac{4r}{\pi} \end{aligned}$$

Für die Beispiele 1 und 2:

$$\begin{aligned} V(X) &= \sum_{k=1}^{\infty} (k-2)^2 \cdot \left(\frac{1}{2}\right)^k = \\ &= \sum_{k=1}^{\infty} \frac{k^2}{2^k} - 4 \cdot \sum_{k=1}^{\infty} \frac{k}{2^k} + 4 \cdot \sum_{k=1}^{\infty} \frac{1}{2^k} = \\ &= 6 - 4 \cdot 2 + 4 \cdot 1 = 2 \end{aligned}$$

$$\sum_{i=1}^{\infty} \frac{i^2}{x^i} = \dots = \frac{x + x^2}{(x-1)^3}$$

$$\begin{aligned} V(Y) &= \lim_{a \rightarrow 2r} \frac{2}{\pi} \int_0^a \frac{(t - \frac{4r}{\pi})^2}{\sqrt{4r^2 - t^2}} \cdot dt = \\ &= \lim_{a \rightarrow 2r} \frac{2}{\pi} [\dots]_0^a = 2r^2 \cdot \left(1 - \frac{8}{\pi^2}\right) \end{aligned}$$

$$\int \frac{x^2 dx}{\sqrt{b^2 - x^2}} = \frac{b^2}{2} \arcsin \frac{x}{b} - \frac{x}{2} \sqrt{b^2 - x^2}$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{2}$$

$$\sigma_Y = \sqrt{V(Y)} = 2r \cdot \sqrt{\frac{1}{2} - \frac{2}{\pi^2}}$$

Bemerkungen:

- Erwartungswert, Varianz und Standardabweichung sind die wahrscheinlichkeitstheoretischen Gegenstücke zum arithmetischen Mittelwert und der empirischen Varianz bzw. Standardabweichung.
- Für die Berechnung der Varianz ergibt sich analog zur empirischen Varianz (siehe beschreibende Statistik):

$$V(X) = E(X^2) - \mu^2$$

- Für diskrete Zufallsgrößen mit nur endlich vielen Werten x_1, x_2, \dots, x_n lassen sich die Erwartungswerte und Varianzen immer berechnen.

Ist die Zufallsgröße jedoch stetig oder hat sie zumindest abzählbar unendlich viele Werte x_k , so brauchen Erwartungswert und Varianz nicht unbedingt zu existieren.

- Für linear transformierte Zufallsgrößen ergeben sich vereinfachende Berechnungsmöglichkeiten, wenn die Erwartungswerte und Varianzen der Ausgangsgrößen existieren und bekannt sind ($a, b \in \mathbb{R}$):

$$\begin{aligned} E(a \cdot X + b) &= a \cdot E(X) + b \\ V(a \cdot X + b) &= a^2 \cdot V(X) \end{aligned}$$

- Die dem Vergleich dienende *Standardisierung* ist eine spezielle lineare Transformation der Zufallsgröße X : Die mittels Erwartungswert μ und Varianz σ von X gebildete Größe

$$Z := \frac{X - \mu}{\sigma}$$

ist die zu X gehörige standardisierte Zufallsgröße; für sie gilt:

$$E(Z) = 0 \quad \text{und} \quad V(Z) = 1$$

Beweis:

$$\begin{aligned} E((X - \mu)^2) &= E(X^2 - 2\mu X + \mu^2) = \\ &= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2 \end{aligned}$$

↑
Summen- bzw. Integraleigenschaft

Der Beweis folgt wie vorhin aus den Eigenschaften der Summe bzw. des Integrals.

Beispiel: Werfen von vier Münzen
 X : „Anzahl der geworfenen Wappen“
 $x_i = 0; 1; 2; 3; 4$

Y : „3 Gewinnpunkte pro Wappen“

$$E(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + \dots + 4 \cdot \frac{1}{16} = 2$$

$$V(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + \dots + 16 \cdot \frac{1}{16} - 2^2 = 1$$

$$\rightarrow E(Y) = E(3X) = 3 \cdot E(X) = 6$$

$$V(Y) = V(3X) = 3^2 \cdot V(X) = 9$$

Fortsetzung Münzenbeispiel:

$$Z = \frac{X - 2}{1} \quad \rightarrow \quad z_i = -2; -1; 0; 1; 2$$

Beweis:

$$E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} \cdot E(X) - \frac{\mu}{\sigma} = \frac{1}{\sigma} \cdot \mu - \frac{\mu}{\sigma}$$

$$V\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \cdot V(X) = \frac{1}{\sigma^2} \cdot \sigma^2$$

Binomialverteilung (BERNOULLI-Verteilung)

Ein Zufallsexperiment ist ein BERNOULLI-Experiment, wenn

- nur zwei verschiedene Ergebnisse möglich bzw. von Interesse sind und
- diese einander ausschließen: *Erfolg* E, *Misserfolg* M = \bar{E} .

Die Wahrscheinlichkeit für den Erfolg sei $p = P(E)$, für den Misserfolg $q = P(M) = 1 - p$.

Bei n -maliger Durchführung dieses Experiments spricht man von einem n -stufigen BERNOULLI-Experiment.

a) Betrachtet man die Zufallsgröße X : „Anzahl der Erfolge bei einem n -stufigen BERNOULLI-Versuch“, so heißt die zugehörige Verteilung *Binomialverteilung mit den Parametern n und p* (X ist binomialverteilt).

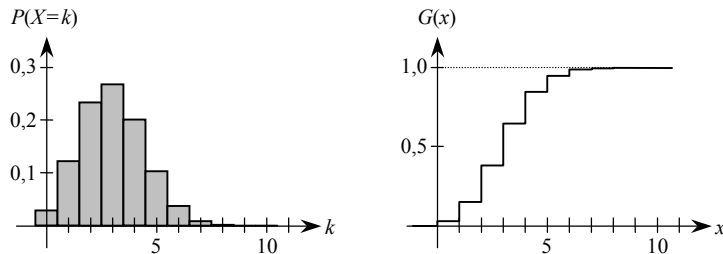
b) Für die Binomialverteilung gilt: Die Wahrscheinlichkeit für genau k Erfolge ist gegeben durch

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

und die Verteilungsfunktion hat die Darstellung

$$G(x) = \begin{cases} 0 & \text{für } x < 0 \\ \sum_{k=0}^m \binom{n}{k} \cdot p^k \cdot q^{n-k} & \text{für } x \geq 0 \end{cases},$$

wobei $m \leq n$ die größte ganze Zahl ist, für die $m \leq x$ gilt.



Wahrscheinlichkeitsverteilung und Verteilungsfunktion für $n = 10$; $p = 0,3$

c) Erwartungswert und Varianz der Binomialverteilung ergeben sich aus den Formeln

$$\begin{aligned} E(X) &= \mu = n \cdot p \\ V(X) &= \sigma^2 = n \cdot p \cdot q \end{aligned}$$

Beispiele:

- Münzwurf (E: „Zahl wird geworfen“)
- Funktionsprüfung eines Apparats (E: „funktionsfähig“)
- Testen auf eine bestimmte Krankheit (E: „nicht infiziert“)

Mehrmaliger Münzwurf, Endabnahme bei einer Produktion, Reihenuntersuchung.

Nach dem Satz über unabhängige Ereignisse gilt für jede beliebige Serie von k Erfolgen bei $n - k$ Misserfolgen die Wahrscheinlichkeit $p^k q^{n-k}$.

Da es $\binom{n}{k}$ derartige (ungeordnete) „Stichproben“ gibt, folgt daraus die Behauptung.

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} = \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k q^{n-k} = \\ &= n \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} q^{n-(k+1)} = \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} = \\ &= np \underbrace{(p+q)}_1^{n-1} \end{aligned}$$

Der Beweis für die Varianz verläuft ähnlich (mit $k^2 = k \cdot (k - 1) + k$).

Bemerkungen:

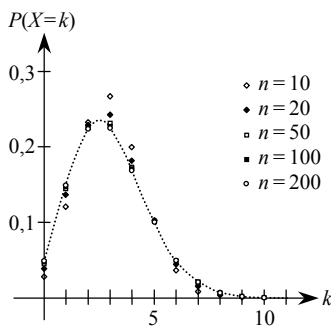
- Die Binomialverteilung wird auch auf Zufallsexperimente angewandt, die nur näherungsweise einem n -stufigen BERNOULLI-Versuch entsprechen, wie dies beispielsweise bei der Verwendung statistischer Wahrscheinlichkeiten der Fall ist (Anzahl von Knabengeburt, Ausschussanteile bei einer Massenfertigung etc.).
- Die Berechnung der Wahrscheinlichkeit bei einem n -stufigen BERNOULLI-Experiment kann rekursiv erfolgen:

$$P(X = k + 1) = \frac{n - k}{k + 1} \cdot \frac{p}{q} \cdot P(X = k) \quad \text{für } k = 0; 1; \dots; n - 1$$

$$P(X = k + 1) = \binom{n}{k + 1} \cdot p^{k + 1} q^{n - (k + 1)} = \frac{n - k}{k + 1} \cdot \binom{n}{k} \cdot p^k q^{n - k} \cdot \frac{p}{q}$$

POISSON-Verteilung (Verteilung für seltene Ereignisse)

Betrachtet man n -stufige BERNOULLI-Experimente, bei denen bei gleichbleibendem Erwartungswert μ die Anzahl n der Wiederholungen sehr groß wird, so zeigt sich, dass die Werte der Binomialverteilung immer besser auf einer bestimmten Kurve liegen. Im Folgenden sei $\mu = 3$:



- a) Als POISSON-Verteilung mit dem Parameter μ bezeichnet man die Wahrscheinlichkeitsverteilung mit

$$P(X = k) = \frac{\mu^k}{k!} \cdot e^{-\mu}, \quad k \in \mathbb{N}_0$$

und der Verteilungsfunktion

$$G(x) = \begin{cases} 0 & \text{für } x < 0 \\ e^{-\mu} \cdot \sum_{k=0}^m \frac{\mu^k}{k!} & \text{für } x \geq 0 \end{cases},$$

wobei m die größte ganze Zahl ist, für die $m \leq x$ gilt.

- b) Erwartungswert und Varianz sind gegeben durch

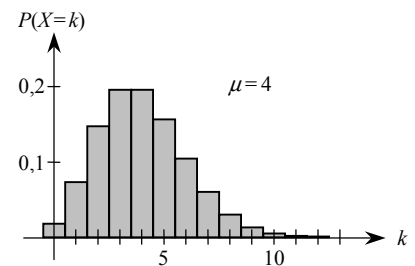
$$\begin{aligned} E(X) &= \mu \\ V(X) &= \sigma^2 = \mu \end{aligned}$$

Wegen $\mu = n \cdot p$ muss die Wahrscheinlichkeit für den Erfolg E sehr klein sein, dieses Ereignis ist also selten.

Grenzkurve ($n \rightarrow \infty$):

$$\begin{aligned} P(X = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1 - p)^{n - k} = \\ &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n - k} = \\ &= \lim_{n \rightarrow \infty} \frac{\mu^k}{k!} \cdot \underbrace{\frac{n \cdot \dots \cdot (n - k + 1)}{(n - \mu)^k}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{\mu}{n}\right)^n}_{\rightarrow e^{-\mu}} \end{aligned}$$

Druckfehler in Büchern, radioaktiver Zerfall, Lackfehler usw. lassen sich in guter Näherung durch die POISSON-Verteilung beschreiben.

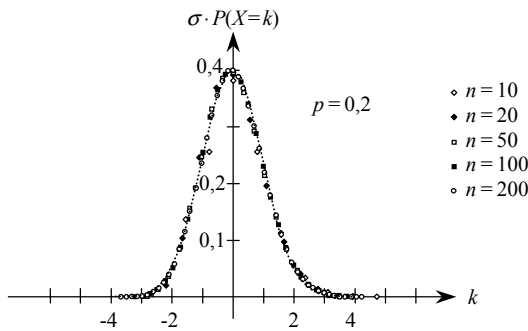


$$\begin{aligned} E(X) &= e^{-\mu} \cdot \sum_{k=0}^{\infty} k \cdot \frac{\mu^k}{k!} = \\ &= e^{-\mu} \cdot \mu \cdot \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \stackrel{\text{TAYLOR-Reihe}}{=} e^{-\mu} \cdot \mu \cdot e^{\mu} \end{aligned}$$

Der Beweis für die Varianz verläuft analog.

Normalverteilung (GAUSS-Verteilung)

Wir betrachten noch einmal n -stufige BERNOULLI-Experimente mit steigender Anzahl n von Wiederholungen, wobei aber diesmal die Wahrscheinlichkeit p konstant bleiben soll. Zum besseren Vergleich werden standardisierte Verteilungen benutzt und zwar in der Form, dass alle Histogramme der Wahrscheinlichkeitsverteilungen den gleichen Flächeninhalt aufweisen sollen (Normieren). Um übersichtlichere Diagramme zu erhalten, werden nur die Rechteckhöhen (über den Seitenmitten) eingezeichnet:



Es zeigt sich, dass mit zunehmender Anzahl n die Punkte immer besser auf einer bestimmten Kurve liegen: GAUSS'sche Glockenkurve. Auf die Herleitung des Bildungsgesetzes dieser Funktion muss verzichtet werden.

- a) Eine stetige Zufallsgröße Z ist *standard-normalverteilt* (mit dem Erwartungswert 0 und der Varianz 1), wenn der Verteilung die GAUSS'sche Dichtefunktion (Glockenkurve)

$$\varphi(z) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$

zu Grunde liegt. Die Verteilungsfunktion ist gegeben durch

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^z e^{-\frac{u^2}{2}} \cdot du$$

Bemerkung: Das Integral kann nur näherungsweise ausgewertet werden, die Werte liegen allerdings in tabellarischer Form vor.

- b) Liegt für eine Zufallsgröße X die Wahrscheinlichkeitsdichte in der Form

$$g(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

vor, so ist X normalverteilt mit dem Erwartungswert μ und der Varianz σ^2 .

Für die standardisierte Zufallsgröße $Z = \frac{1}{\sigma}(X - \mu)$ gilt $z_1 - z_2 = \frac{1}{\sigma}(x_1 - x_2)$, d. h., die Rechteckhöhen müssen für die Normierung mit σ multipliziert werden.

LAPLACE-Bedingung: Gilt für eine binomialverteilte Zufallsvariable

$$\sigma^2 > 9,$$

so können die Wahrscheinlichkeitswerte in guter Näherung mittels der GAUSS'schen Dichtefunktion berechnet werden:

$$P(X = k) \approx \frac{1}{\sigma} \cdot \varphi\left(\frac{k - \mu}{\sigma}\right)$$

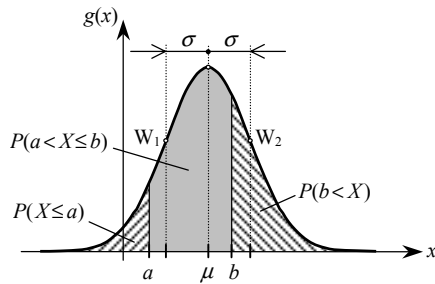
Beispiele normalverteilter Größen: (zufällige) Messfehler, Füllmengen

Wegen der Symmetrie $\varphi(-z) = \varphi(z)$ gilt $\Phi(-z) = 1 - \Phi(z)$ und daher sind in den Tabellen die Werte von φ und Φ ab $z=0$ bis ca. $z=5$ angegeben.

Beweis: Wegen $X = \sigma \cdot Z + \mu$ und der Bemerkung über linear transformierte Zufallsvariablen ergibt sich sofort die Behauptung.

Eigenschaften der Normalverteilung

- a) Wichtige Teilflächen und ihre Berechnung mittels der Verteilungsfunktion:



$$\begin{aligned}
 P(X \leq a) &= G(a) \\
 P(a < X \leq b) &= G(b) - G(a) \\
 P(b < X) &= 1 - G(b)
 \end{aligned}$$

Zum Skizzieren praktisch: die Wendepunkte $W_{1,2}(\mu \pm \sigma | y_w)$, wobei y_w ca. 61% der Maximalhöhe entspricht.

- b) Für eine normalverteilte Zufallsvariable gilt:

$$\begin{aligned}
 P(\mu - \sigma < X \leq \mu + \sigma) &\approx 68,3\% \\
 P(\mu - 2\sigma < X \leq \mu + 2\sigma) &\approx 95,5\% \quad \text{bzw.} \\
 P(\mu - 3\sigma < X \leq \mu + 3\sigma) &\approx 99,7\%
 \end{aligned}$$

$$\begin{aligned}
 P(\mu - 1,96\sigma < X \leq \mu + 1,96\sigma) &\approx 95,0\% \\
 P(\mu - 2,58\sigma < X \leq \mu + 2,58\sigma) &\approx 99,0\% \\
 P(\mu - 3,29\sigma < X \leq \mu + 3,29\sigma) &\approx 99,9\%
 \end{aligned}$$

Bemerkung: Diese Wahrscheinlichkeitsaussagen sind die zur Ungleichung von ČEBYŠEV äquivalenten Formulierungen.

- c) Näherung einer Binomialverteilung ($\mu = n \cdot p$, $\sigma^2 > 9$) durch die Normalverteilung:

$$\begin{aligned}
 P(X \leq k) &\approx \Phi\left(\frac{1}{\sigma} \cdot (k + 0,5 - \mu)\right) \\
 P(k_1 \leq X \leq k_2) &\approx \Phi\left(\frac{1}{\sigma} \cdot (k_2 + 0,5 - \mu)\right) - \Phi\left(\frac{1}{\sigma} \cdot (k_1 - 0,5 - \mu)\right) \\
 P(X = k) &\approx \Phi\left(\frac{1}{\sigma} \cdot (k + 0,5 - \mu)\right) - \Phi\left(\frac{1}{\sigma} \cdot (k - 0,5 - \mu)\right)
 \end{aligned}$$

- d) **Zentraler Grenzwertsatz:** Die Verteilung einer Summe von n unabhängigen Zufallsgrößen X_i nähert sich mit steigender Anzahl n zunehmend einer Normalverteilung an. Für den Erwartungswert und die Varianz gilt dabei:

$$\mu = \sum_{i=1}^n \mu_i \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2$$

- e) Folgerung aus dem Zentralen Grenzwertsatz: Werden aus einer Grundgesamtheit mit dem Mittelwert μ_x und der Standardabweichung σ_x Stichproben mit dem Umfang n gezogen, so sind die Mittelwerte der Stichprobe annähernd normalverteilt mit

$$\mu = \mu_x \quad \sigma^2 = \frac{1}{n} \cdot \sigma_x^2$$

Bemerkung: Die Stichprobenmittelwerte streuen also weniger als die Messwerte um den (gemeinsamen) Mittelwert.

Gebrauch der Tabelle für die standardisierten Normalverteilung: Mit $z_a = \frac{1}{\sigma} \cdot (a - \mu)$ gilt

$$\begin{aligned}
 G(a) &= \Phi(z_a) \quad \text{für } a \geq \mu \\
 G(a) &= 1 - \Phi(-z_a) \quad \text{für } a < \mu
 \end{aligned}$$

Liegen a und $b > a$ symmetrisch zu μ , so gilt

$$P(a < X \leq b) = 2 \cdot \Phi(z_b) - 1$$

Im ersten Fall bedeutet dies zum Beispiel für normalverteilte Messfehler: Abweichungen um mehr als σ vom Mittelwert sind nur bei etwa jeder dritten Messung zu erwarten.

Die zweiten Wahrscheinlichkeiten sind im Zusammenhang mit dem Schätzen von (Verteilungs-) Parametern bzw. Testen von Hypothesen von Bedeutung.

(Die auf MOIVRE und LAPLACE zurückgehende Näherung heißt „Integrale Näherungsformel“.)

Dieser Satz liefert auch die Begründung dafür, dass viele Messgrößen annähernd normalverteilt sind.

Betrachtet man die Stichprobenwerte als Werte von Zufallsvariablen X_i (gleiche Verteilung, gleicher Erwartungswert μ_x , gleiche Varianz σ_x^2), so beschreibt die Größe

$$\bar{X} := \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

das Stichprobenmittel und es gilt:

$$E(\bar{X}) = \frac{1}{n} \cdot n \cdot \mu_x \quad \text{und} \quad V(\bar{X}) = \frac{1}{n} \cdot n \cdot \sigma_x^2$$