

Kapitel 1

Grundlagen der Statistik

1.1 Wahrscheinlichkeit

Grundlegend für statistische Analysen, das heißt der Behandlung von Vorgängen mit zufälligem, unvorhersagbarem Ausgang, ist der Begriff der Wahrscheinlichkeit. Obwohl so grundlegend, wird über die Definition der Wahrscheinlichkeit immer noch, zum Teil sehr emotional, gestritten. Es gibt eine, nicht umstrittene, axiomatische Definition, die die Rechenregeln festlegt, aber offen lässt, wie man tatsächlich Wahrscheinlichkeiten bestimmt. In der Praxis benutzt man meistens eine Definition über die relative Häufigkeit von Ereignissen.

1.1.1 Definition über die Häufigkeit

Wenn man N Versuche macht, bei denen das Ereignis e auftreten kann, und dabei n mal das Ereignis e tatsächlich auftritt, ordnet man dem Ereignis e die Wahrscheinlichkeit $p(e)$ durch die relative Häufigkeit des Auftretens des Ereignisses zu:

$$p(e) = \lim_{N \rightarrow \infty} \frac{n}{N} \quad (1.1)$$

In der Praxis wird der Grenzübergang zu unendlich vielen Versuchen erschlossen oder aus endlichen ‘Stichproben’ abgeschätzt.

1.1.2 Kombinatorische Definition

Wahrscheinlichkeiten können erschlossen werden, wenn man zum Beispiel aus Symmetriebetrachtungen argumentieren kann, dass alle möglichen Ereignisse gleich wahrscheinlich sind, zum Beispiel welche Zahl beim Würfeln erscheint. Dann ist die Wahrscheinlichkeit für jedes einzelne Ereignis durch die Anzahl der möglichen Ereignisse N gegeben:

$$p(e) = \frac{1}{N} \quad (1.2)$$

Zum Beispiel ist die Wahrscheinlichkeit für das Würfeln einer 6 gerade $1/6$ und das Werfen von ‘Zahl’ bei einer Münze $1/2$. Beim Werfen von zwei Würfeln ist jede Kombination von Zahlen gleich wahrscheinlich, also $1/36$ (weil es $6 \cdot 6 = 36$ Kombinationen gibt). Was ist die Wahrscheinlichkeit, dass mindestens eine 6 auftritt?

Dazu muss man die Anzahl der Kombinationen mit mindestens einer 6 abzählen: 1) der erste Würfel hat eine 6 und der andere hat die Zahlen 1 bis 5; 2) dasselbe für die ausgetauschten Würfel; 3) beide haben eine 6. Das sind also $2 \cdot 5 + 1 = 11$ Kombinationen und damit ist die Wahrscheinlichkeit $11/36$.

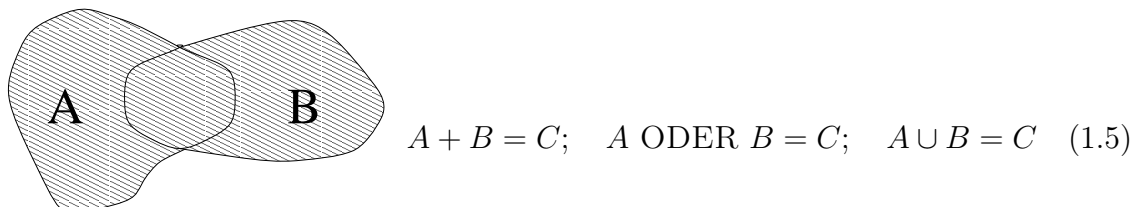
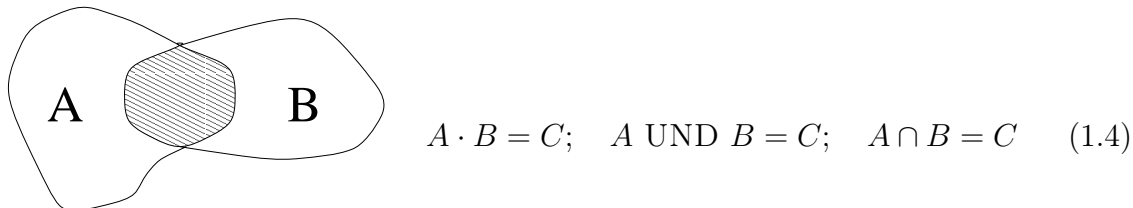
Der Fall, dass alle Möglichkeiten gleich wahrscheinlich sind, hat in der Physik eine besondere Bedeutung: in der Quantentheorie kann ein physikalisches System verschiedene Zustände einnehmen, die alle mit gleicher Wahrscheinlichkeit auftreten.

1.1.3 Axiomatische Definition der Wahrscheinlichkeit

Ereignismenge: Es sei

$$\Omega = \{e_i\} \quad (1.3)$$

die Menge aller möglichen Ereignisse, zum Beispiel die möglichen Resultate eines Experimentes. Für Untermengen $A, B, C \subseteq \Omega$ werden die üblichen Verknüpfungen, Durchschnitt und Vereinigung, definiert:



Durchschnitt \cap und Vereinigung \cup entsprechen den logischen Operationen UND (\cdot) und ODER ($+$).

Weiterhin wird ein **elementares Ereignis**, das **Komplement** \bar{A} von A und das **sichere Ereignis** E definiert (\emptyset ist die leere Menge):

$$A \text{ elementar} \iff A \cdot B = \emptyset \text{ oder } A \cdot B = A \quad \forall B \in \Omega \quad (1.6)$$

Das Nichteintreten von A ist \bar{A} und damit sind

$$A + \bar{A} = E, \quad A \cdot \bar{A} = \emptyset \quad (1.7)$$

das sichere und das unmögliche Ereignis.

Wahrscheinlichkeitsaxiome: Jedem Ereignis $A \in \Omega$ wird eine Zahl $p(A)$ mit folgenden Eigenschaften zugeordnet:

- (1) $0 \leq p(A) \leq 1$
- (2) $p(\Omega) = 1$
- (3) $A \cdot B = \emptyset \implies p(A + B) = p(A) + p(B)$

Offensichtlich erfüllen die beiden oben angegebenen Definitionen für die Wahrscheinlichkeit diese Axiome. Andererseits legen die Axiome nicht fest, wie man tatsächlich Wahrscheinlichkeiten bestimmen soll.

Aus den Axiomen ergibt sich:

- Eine Untermenge A von B hat eine kleinere Wahrscheinlichkeit als B :

$$A \subset B \implies p(A) \leq p(B) \quad (1.8)$$

- Im allgemeinen, falls (3) nicht zutrifft, also $A \cdot B \neq \emptyset$ ist, gilt das **Additionstheorem**:

$$p(A + B) = p(A) + p(B) - p(A \cdot B) \quad (1.9)$$

Bedingte Wahrscheinlichkeiten: Die Wahrscheinlichkeit von A , wenn B gegeben ist, wird mit $p(A|B)$ bezeichnet:

$$p(A|B) = p(A) \text{ gegeben } B \quad (1.10)$$

Zum Beispiel ändert sich die Wahrscheinlichkeit schwarzhaarig zu sein, wenn man die beiden Bedingung betrachtet, dass die Person eine Deutsche oder dass die Person eine Griechin ist. Die bedingte Wahrscheinlichkeit ergibt sich zu:

$$p(A|B) = \frac{p(A \cdot B)}{p(B)} \quad (1.11)$$

Das ist also zum Beispiel die Wahrscheinlichkeit, schwarzhaarig und Grieche zu sein, normiert auf die Wahrscheinlichkeit Grieche zu sein. Mit der Häufigkeitsdefinition würde man also die Anzahl der schwarzhaarigen Griechen durch die Zahl aller Griechen dividieren.

Die Gleichung (1.11) lässt sich nach $p(A \cdot B)$ auflösen:

$$p(A \cdot B) = p(A|B) \cdot p(B) = p(B|A) \cdot p(A) \quad (1.12)$$

Daraus folgt das **Bayes-Theorem**:

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)} \quad (1.13)$$

Beispiel: Eine Krankheit K trete in der gesamten Bevölkerung mit der Häufigkeit $p(K) = 10^{-4}$ auf. Auf diese Krankheit reagiert ein zu derem Nachweis entwickelter Test mit einer Wahrscheinlichkeit von 98% positiv (+), also

$p(+|K) = 0.98$. Allerdings spricht die Gesamtbevölkerung mit einer Wahrscheinlichkeit von 3% ebenfalls positiv an, also $p(+) = 0.03$. Was ist die Wahrscheinlichkeit, die Krankheit zu haben, wenn das Testresultat positiv ist? Die Rechnung ergibt:

$$p(K|+) = \frac{p(+|K) \cdot p(K)}{p(+)} = \frac{0.98 \cdot 10^{-4}}{0.03} \approx 0.003 \quad (1.14)$$

Diese geringe Wahrscheinlichkeit von nur 3 Promille würde zum Beispiel einen schwereren Eingriff, der im Krankheitsfall notwendig würde, nicht rechtfertigen. Obwohl die Effizienz des Tests, die Krankheit nachzuweisen, recht gut ist, ist die Fehlerrate bei Gesunden relativ hoch. Das liegt daran, dass die ‘a priori’ Wahrscheinlichkeit für das Auftreten der Krankheit sehr klein ist. Das gleiche Problem tritt auf, wenn man in Experimenten sehr seltene Ereignisse identifizieren will, die Identifikation aber auch auf die anderen Ereignisse mit einer zwar kleinen aber endlichen Wahrscheinlichkeit anspricht. Abhilfe schaffen hier nur weitere unabhängige Tests, so dass sich die Ansprechwahrscheinlichkeiten multiplizieren.

Unabhängige Ereignisse: Man nennt zwei Ereignisse unabhängig, wenn gilt:

$$A, B \text{ unabhängig} \iff p(A|B) = p(A) \iff p(A \cdot B) = p(A) \cdot p(B) \quad (1.15)$$

Beispiel: Wenn man zwei Würfel wirft, sind die Ergebnisse beider Würfel unabhängig voneinander. Die Wahrscheinlichkeit zweimal 6 zu würfeln ist demnach

$$\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36},$$

wie man auch mit dem kombinatorischen Ansatz durch Abzählen der möglichen Fälle findet.

Allgemeine Form des Bayes-Theorems: Wenn die Gesamtheit der Ereignisse E sich vollständig in unabhängige Ereignisse oder Klassen A_i zerlegen lässt,

$$E = \sum_{i=1}^n A_i, \quad (1.16)$$

dann lässt sich B als Summe der möglichen Klassenzugehörigkeiten darstellen:

$$p(B) = \sum_{i=1}^n p(B|A_i)p(A_i) \quad (1.17)$$

Eingesetzt in (1.13) ergibt sich das Bayes-Theorem in allgemeinerer Form:

$$p(A_j|B) = \frac{p(B|A_j) \cdot p(A_j)}{\sum_{i=1}^n p(B|A_i)p(A_i)} \quad (1.18)$$

Beispiel: In dem obigen Beispiel mit dem Test zum Nachweis einer Krankheit hatten wir $p(+)=0.03$ als die Wahrscheinlichkeit, mit der die Gesamtbevölkerung auf den Test anspricht, angesetzt. Zerlegen wir die Gesamtheit in Kranke und Nichtkranke, K und \bar{K} , dann ist diese Wahrscheinlichkeit:

$$p(+)=p(+|K)p(K)+p(+|\bar{K})p(\bar{K}) \quad (1.19)$$

und Gleichung (1.14) wird:

$$p(K|+)=\frac{p(+|K)\cdot p(K)}{p(+|K)p(K)+p(+|\bar{K})p(\bar{K})} \quad (1.20)$$

Eine solche Darstellung ist sinnvoll, wenn die Testergebnisse für beide Klassen getrennt vorliegen.

1.2 Verteilungen von Zufallsvariablen

Das Ergebnis eines Experimentes wird durch eine Zufallsvariable x oder einen Satz von Zufallsvariablen $\vec{x}=(x_1, x_2, \dots)$ beschrieben. Diese Variablen können diskrete oder kontinuierliche Werte haben.

Bei **diskreten Variablen** n können wir eine Wahrscheinlichkeit $p(n)$ für das Auftreten eines bestimmten Wertes von n angeben. Ein Beispiel ist die Wahrscheinlichkeit für das Auftreten von n Zerfällen eines radioaktiven Präparates in einem festen Zeitintervall Δt . Üblicherweise werden solche Verteilungen diskreter Variablen wie in Abb. 1.1 als Treppenfunktion dargestellt.

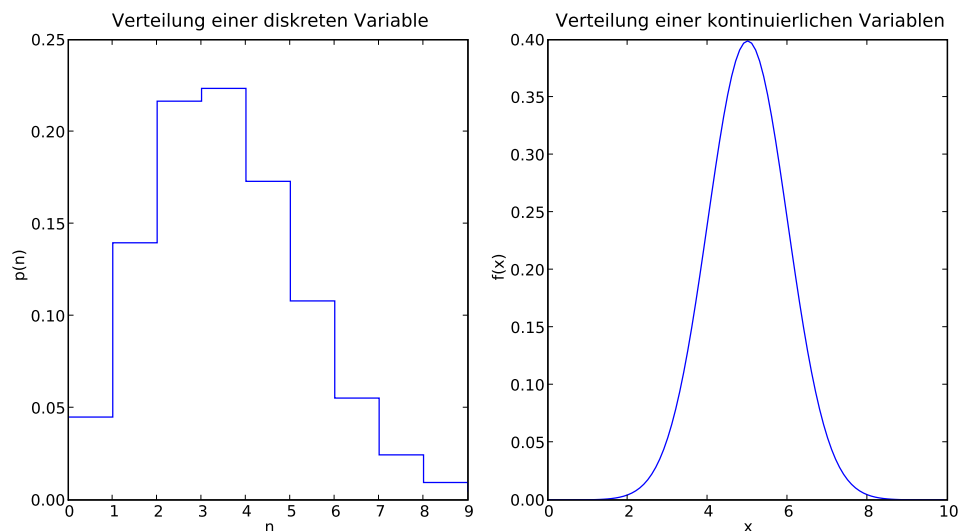


Abbildung 1.1: Beispiele von Wahrscheinlichkeitsverteilungen: diskrete Variable (links); kontinuierliche Variable (rechts).

Bei **kontinuierlichen Variablen** gibt man eine Wahrscheinlichkeit für das Auftreten von x in einem Intervall Δx an:

$$\Delta p(x)=\frac{\Delta p(x)}{\Delta x}\Delta x \quad \xrightarrow{\Delta x \rightarrow 0} \quad dp(x)=\frac{dp(x)}{dx}dx=f(x)dx, \quad (1.21)$$

wobei $f(x)$ **Wahrscheinlichkeitsdichte** genannt wird (mit der Dimension von x^{-1}).

1.2.1 Eigenschaften von Verteilungen

Normierung: Die Wahrscheinlichkeit, irgendeinen möglichen Wert von x bzw. n zu erhalten, muss 1 sein:

$$\begin{aligned} \text{kontinuierliche Variable : } & \int_{-\infty}^{+\infty} f(x) dx = 1 \\ \text{diskrete Variable : } & \sum_{n=0}^{+\infty} p(n) = 1 \end{aligned} \tag{1.22}$$

Die Integrations- oder Summationsgrenzen können auch allgemeiner gewählt werden (x_{min} , x_{max} bzw. n_{min} , n_{max}), zur Vereinfachung benutzen wir im Folgenden aber meistens die Grenzen wie in (1.22).

Beispiel: In der Physik treten häufig Exponentialfunktionen auf, die Wachstum oder Abnahme proportional dem jeweils Vorhandenen und der Intervalllänge dx der Variablen beschreiben. Die physikalische Annahme ist, dass die Wahrscheinlichkeit pro Zeitintervall gleich und unabhängig von der bereits verstrichenen Zeit ist. Für einen Absorptions- oder Zerfallsprozess ergibt sich zum Beispiel:

$$df(x) = -f(x) \lambda dx \tag{1.23}$$

Bekanntlich ergibt sich daraus:

$$f(x) = f_0 e^{-\lambda x} \tag{1.24}$$

Diese Wahrscheinlichkeitsdichte soll im x -Intervall $[0, \infty]$ normiert werden:

$$1 = \int_0^{\infty} f_0 e^{-\lambda x} dx = f_0 \frac{1}{\lambda} \tag{1.25}$$

Daraus folgt:

$$f(x) = \lambda e^{-\lambda x} \tag{1.26}$$

Verteilungsfunktion: Häufig möchte man die Wahrscheinlichkeit, dass x in einem Intervall $[x_1, x_2]$ liegt, bestimmen (Abb. 1.2). Dazu muss man das entsprechende Integral der Wahrscheinlichkeitsdichte auswerten:

$$p(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x) dx = \int_{-\infty}^{x_2} f(x) dx - \int_{-\infty}^{x_1} f(x) dx = F(x_2) - F(x_1) \tag{1.27}$$

Unter anderem kann man hier auch sehen, dass die Wahrscheinlichkeit, einen ganz bestimmten Wert von x zu erhalten, Null ist, weil die Fläche über einem Punkt Null ist. Das bestimmte Integral

$$F(x) = \int_{-\infty}^x f(\xi) d\xi \tag{1.28}$$

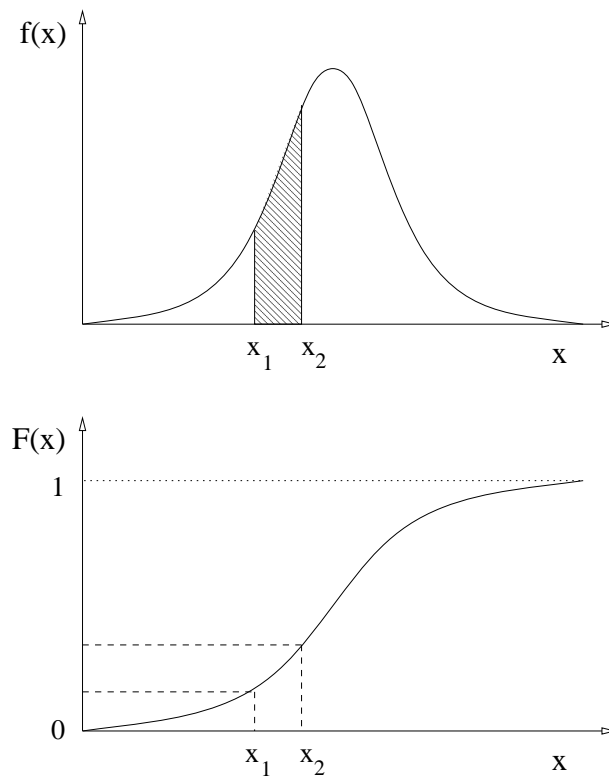


Abbildung 1.2: Wahrscheinlichkeitsdichte (oben) und dazugehörige Verteilungsfunktion (unten).

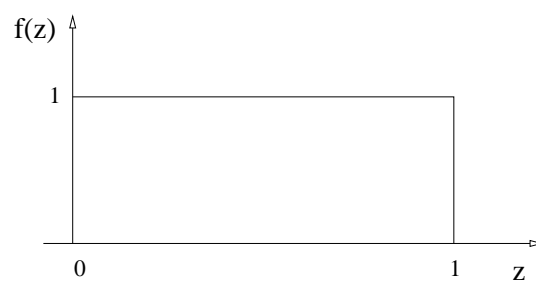


Abbildung 1.3: Wahrscheinlichkeitsdichte einer zwischen 0 und 1 gleichverteilten Variablen.

nennt man die (kumulative) Verteilungsfunktion zu $f(x)$. Der Funktionswert $F(x_0)$ entspricht der Wahrscheinlichkeit, dass x kleiner als x_0 ist:

$$F(x_0) = p(x < x_0). \quad (1.29)$$

Bei diskreten Variablen ergibt sich die Verteilungsfunktion entsprechend:

$$P(n) = \sum_{k=0}^n p(k) \quad (1.30)$$

Für wichtige Verteilungen sind Wahrscheinlichkeitsdichte und Verteilungsfunktion in Statistikbüchern tabelliert zu finden.

Die Zuordnung

$$x \rightarrow F(x) \quad (1.31)$$

bildet die Zufallsvariable x auf eine gleichverteilte Variable $z = F(x)$ zwischen 0 und 1 ab (Abb. 1.3). Das sieht man wie folgt: Wenn z eine gleichverteilte Variable ist, die aber die gleiche Wahrscheinlichkeit um den Punkt z wie um x beschreibt, muss gelten:

$$dp(x) = f(x)dx = dz = dp(z) \quad (1.32)$$

Der Bezug zu der Verteilungsfunktion ergibt sich dann durch Integration beider Seiten in (1.32):

$$F(x) = \int_{-\infty}^x f(\xi)d\xi = \int_0^z d\zeta = z \quad (1.33)$$

Die Normierung von $f(x)$ stellt sicher, dass z im Intervall $[0,1]$ liegt.

Erzeugung von Zufallsvariablen: Computerprogramme haben in der Regel Zugang zu Zufallszahlengeneratoren, die Zufallszahlen im Intervall $[0,1]$ liefern. Wenn die zu der Dichte f gehörende Verteilungsfunktion F eine analytisch invertierbare Funktion ist, ist es besonders einfach, die Zufallsvariable x entsprechend der Dichte $f(x)$ zu würfeln: Man erzeugt sich gleichverteilte Zufallszahlen z_i , $i = 1, \dots, n$ und bestimmt daraus die x_i :

$$F(x_i) = z_i \quad \Rightarrow \quad x_i = F^{-1}(z_i) \quad (1.34)$$

Beispiel: Wir wollen die Variable t mit der Wahrscheinlichkeitsdichte

$$f(t) = \lambda e^{-\lambda t}, \quad (1.35)$$

erzeugen. Dazu ordnen wir t der gleichverteilten Variablen z zu:

$$z = \int_0^t f(\tau)d\tau = 1 - e^{-\lambda t}. \quad (1.36)$$

Die Umkehrung ergibt:

$$t = \frac{1}{\lambda} \ln \frac{1}{1-z}. \quad (1.37)$$

Man sieht, dass zum Beispiel $z = 0$ auf $t = 0$ und $z = 1$ auf $t = \infty$ abgebildet wird.

1.2.2 Erwartungswerte

Eine Funktion $g(x)$ von der Zufallsvariablen x mit der Wahrscheinlichkeitsdichte $f(x)$ hat den Erwartungswert:

$$E(g(x)) = \langle g(x) \rangle = \int_{-\infty}^{+\infty} g(x)f(x)dx \quad (1.38)$$

Entsprechend gilt für den Erwartungswert einer Funktion $q(n)$ der diskreten Variablen n mit der Wahrscheinlichkeitsverteilung $p(n)$:

$$E(q(n)) = \langle q(n) \rangle = \sum_{n=0}^{\infty} q(n)p(n) \quad (1.39)$$

Die Bildung des Erwartungswertes ist eine lineare Operation:

$$E(a \cdot g(x) + b \cdot h(x)) = a \cdot E(g(x)) + b \cdot E(h(x)) \quad (1.40)$$

Im Folgenden behandeln wir spezielle Erwartungswerte, die für die Beschreibung von Verteilungen wichtig sind.

Mittelwert: Der Erwartungswert der Zufallsvariablen x selbst, heisst der Mittelwert der Verteilung:

$$\mu = E(x) = \int_{-\infty}^{+\infty} x f(x)dx \quad (1.41)$$

Zum Beispiel ergibt sich für das Zerfallsgesetz

$$f(t) = \lambda e^{-\lambda t}, \quad (1.42)$$

eine mittlere Lebensdauer $\langle t \rangle = 1/\lambda$.

Varianz: Der Erwartungswert der quadratischen Abweichung vom Mittelwert heisst mittlere quadratische Abweichung oder Varianz:

$$\sigma^2 = E((x - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx \quad (1.43)$$

Die Wurzel aus der Varianz, σ , heisst Standardabweichung. Für die praktische Berechnung der Varianz ist folgende Relation nützlich:

$$\sigma^2 = E((x - \mu)^2) = E(x^2 - 2\mu x + \mu^2) = E(x^2) - 2\mu E(x) + \mu^2 = E(x^2) - \mu^2 \quad (1.44)$$

Dabei ist die Linearität des Operators E und $\mu = E(x)$ benutzt worden.

Momente einer Verteilung: Allgemein nennt man die Erwartungswerte von Potenzen von x oder $x - \mu$ Momente der Verteilung:

$$\begin{aligned} \mu'_n &= E(x^n) && n - \text{tes algebraisches Moment} \\ \mu_n &= E((x - \mu)^n) && n - \text{tes zentrales Moment} \end{aligned} \quad (1.45)$$

Spezielle Momente:

- $\mu'_1 =$ Mittelwert,
- $\mu_2 =$ Varianz
- $\beta = \mu_3/\sigma^3 =$ Schiefe (=0 für symmetrische Verteilungen)

Mittelwert, Varianz und Schiefe werden benutzt, um Verteilungen zu charakterisieren. Häufig sind diese Größen Parameter von speziellen Verteilungen, die experimentell zu bestimmen sind. Zum Beispiel ist die Gaussverteilung durch Mittelwert und Varianz gegeben; die Wahrscheinlichkeitsverteilung für einen Zerfall nach (1.42) ist durch die mittlere Zerfallszeit $\tau = 1/\lambda$ gegeben.

Eine Wahrscheinlichkeitsdichte kann nach Momenten entwickelt werden, entsprechend einer Taylor-Entwicklung.

Charakteristische Funktion Die charakteristische Funktion einer Wahrscheinlichkeitsdichte ist deren Fourier-Transformierte, was dem Erwartungswert einer komplexen Exponentialfunktion entspricht:

$$\phi(t) = E(e^{itx}) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx; \quad (1.46)$$

entsprechend für diskrete Verteilungen:

$$\phi(t) = E(e^{itx}) = \sum_0^{+\infty} e^{itx} p(k). \quad (1.47)$$

Die Eigenschaften einer Fourier-Transformation können vorteilhaft für Rechnungen mit Verteilungen genutzt werden (zum Beispiel wird die Berechnung von Momenten dadurch sehr erleichtert). Allerdings wollen wir es hier im wesentlichen bei der Erwähnung charakteristische Funktionen belassen und im Folgenden auf deren Einsatz verzichten.

1.2.3 Wahrscheinlichster Wert und Median

Zur Charakterisierung von Verteilungen werden auch andere Größen herangezogen:

Wahrscheinlichster Wert: Bei diesem Wert der Variablen hat die Wahrscheinlichkeitsdichte ein Maximum.

Median: Bei diesem Wert der Variablen hat die Verteilungsfunktion gerade 0.5 erreicht, $F(x_m) = 0.5$. Eine Verallgemeinerung sind Quantile, bei der die Verteilungsfunktion einen bestimmten Wert erreicht, zum Beispiel 0.9 (benutzt zur Angabe von Vertrauensbereichen).

Bei asymmetrischen Verteilungen fallen Mittelwert, wahrscheinlichster Wert und Median nicht zusammen.

1.2.4 Stichproben und Schätzwerte

Bei einer Messung entnimmt man meistens der Gesamtheit aller möglichen Werte einer oder mehrerer Zufallsvariablen eine endliche Stichprobe (die Gesamtheit kann endlich oder unendlich sein).

Beispiel: Eine Länge x wird n -mal gemessen. Die Messwerte x_1, \dots, x_n sind eine Stichprobe aus den unendlich vielen möglichen Messungen (Abb. 1.4).



Abbildung 1.4:

Eine Stichprobe benutzt man dann, um auf das Verhalten der Zufallsvariablen zurückzuschließen. Dabei reduziert man die Daten auf wesentliche Informationen, die dann Rückschlüsse auf die ursprünglichen Verteilungen, zum Beispiel über die Bestimmung der Parameter der Verteilungen, erlauben. Die aus einer Stichprobe gewonnenen Parameter von Verteilungen nennt man Schätzwerte. Schätzwerte von Erwartungswerten werden häufig durch Mittelung der entsprechenden Größe über die Stichprobe gebildet.

Schätzung der Verteilung: Die Wahrscheinlichkeitsdichte kann nur gemittelt über endliche Intervalle der Zufallsvariablen geschätzt werden. Falls es sich um eine kontinuierliche Variable handelt, wird man Messwerte in endliche Intervalle ('Bins') zusammenfassen, 'histogrammieren'.

Beispiel: Bei der Messung des Zerfalls einer radioaktiven Probe seien N_0 Zerfälle mit jeweils $N(t_i)$ Zerfällen in Zeitintervallen Δt um t_i gemessen worden (Abb. 1.5). Eine Abschätzung der Wahrscheinlichkeitsdichte erhält man aus:

$$\hat{f}(t_i) = \frac{N(t_i)}{N_0} \quad (1.48)$$

Wie man leicht sieht, ist die Normierung

$$\sum_i \hat{f}(t_i) = 1 \quad (1.49)$$

sichergestellt.

Mittelwert: Den Schätzwert für den Mittelwert einer Verteilung erhält man durch Mittelung der Messwerte. Aus n Messwerten x_1, \dots, x_n erhält man als Schätzwert \bar{x} des Erwartungswertes $\langle x \rangle$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.50)$$

Beispiel: In dem vorigen Beispiel würde man die mittlere Zerfallszeit $\tau = 1/\lambda$ (nach Gleichung (1.42)) durch Mittelung über die Messintervalle bestimmen:

$$\hat{\tau} = \frac{1}{N_0} \sum_i t_i N(t_i) = \sum_i t_i \hat{f}(t_i). \quad (1.51)$$

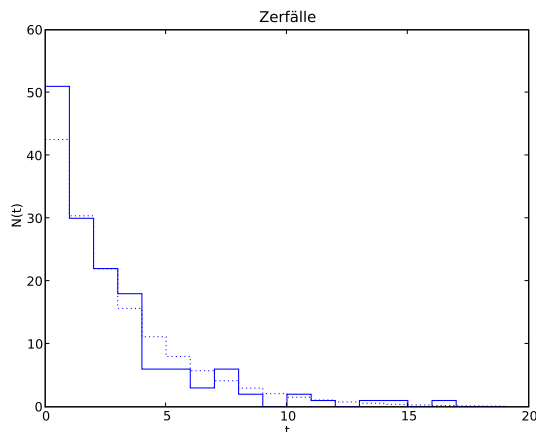


Abbildung 1.5: Histogramm der Anzahl von Zerfällen pro Zeitintervall. Die Messwerte (durchgezogen) und die exakte Verteilung (gepunktet) werden verglichen.

Varianz: Als Schätzwert der Varianz definiert man:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.52)$$

Mit der Division durch $n-1$ statt n erhält man eine bessere Abschätzung, wie wir später noch bei der Diskussion der optimalen Eigenschaften von Schätzwerten sehen werden.

1.3 Simulation von Verteilungen

Computer-Simulationen sind ein wichtiges Hilfsmittel in verschiedensten Bereichen geworden, wie zum Beispiel in Wissenschaft, Technik, Wirtschaft. So werden Wetter- und Klimamodelle, Optimierungen von Auto- und Flugzeugformen, Bestimmung von Nachweiswahrscheinlichkeiten von Teilchenreaktionen oder Lösungen von komplizierten Integralen mit Simulationen nach dem Zufallsprinzip (Monte-Carlo-Methode) berechnet. Die Idee ist, repräsentative Stichproben zu erzeugen, die von einem Satz Zufallsvariablen abhängen. Für jedes erzeugte ‘Ereignis’ werden die Variablen entsprechend ihrer Wahrscheinlichkeitsverteilung ‘gewürfelt’.

In der Regel geht man von einem Zufallszahlengenerator aus, der bei jedem Aufruf eine neue Zahl z , die im Intervall $[0, 1]$ gleichverteilt ist, zurückgibt. Die Frage ist dann, wie man eine Variable in einem beliebigen Intervall und mit einer beliebigen Verteilung erzeugt.

1.3.1 Umkehrung der Verteilungsfunktion

Eine Methode haben wir bereits in Abschnitt 1.2.1 kennengelernt: Die Verteilungsfunktion $F(x)$ zu einer Wahrscheinlichkeitsdichte ist gleichverteilt zwischen 0 und 1. Wir können also

$$z = F(x) \quad (1.53)$$

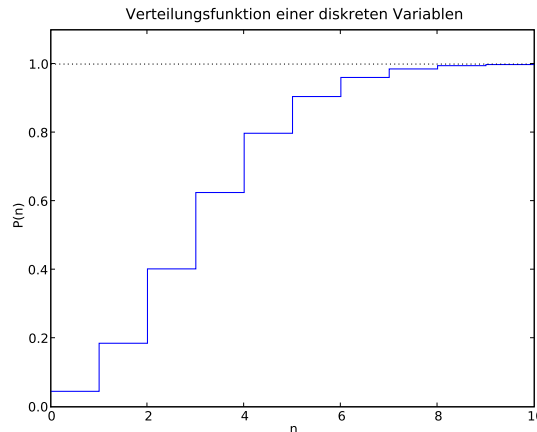


Abbildung 1.6: Verteilungsfunktion einer diskreten Variablen.

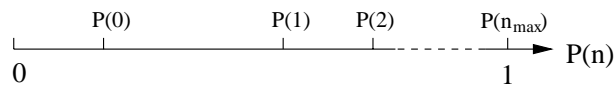


Abbildung 1.7: Abbildung der Verteilungsfunktion einer diskreten Variablen auf das Einheitsintervall.

setzen und erhalten, wenn die Umkehrfunktion F^{-1} existiert, zu jeder gewürfelten Zahl z die entsprechende Zufallszahl x mit der gewünschten Verteilung:

$$x = F^{-1}(z) \quad (1.54)$$

Beispiel: Ein Beispiel ist bereits für die Lebensdauerverteilung gegeben worden (Gleichungen (1.35 - 1.37)).

Bei diskreten Verteilungen ist die Verteilungsfunktion eine Stufenfunktion (Abb. 1.6):

$$P(n) = \sum_{k=0}^n p(k). \quad (1.55)$$

Wenn man die Werte $P(0), P(1), \dots, P(n)$ als Einteilung des Intervalles $[0, 1]$ benutzt (Abb. 1.7) entspricht der Länge jedes Abschnitts gerade eine Wahrscheinlichkeit $p(k)$, beginnend bei $p(0)$ und endend bei $p(n)$. Einer gewürfelten Zufallszahl z ordnet man dann die diskrete Zufallszahl k zu, wenn gilt:

$$\begin{aligned} P(k-1) < z \leq P(k), & \quad k \neq 0 \\ 0 \leq z \leq P(0), & \quad k = 0 \end{aligned} \quad (1.56)$$

Wenn man zu der Verteilungsfunktion einer kontinuierlichen Variablen x keine Umkehrfunktion findet, kann man die Variable diskretisieren, zum Beispiel in Intervalle Δx um diskrete Werte x_i aufteilen zu denen Wahrscheinlichkeiten $f(x_i) \cdot \Delta x$ gehören (siehe das Beispiel in Abb. 1.5). Verteilungen, die sich bis $+\infty$ oder $-\infty$ ausdehnen, aber in der Regel mit fallenden Wahrscheinlichkeiten, schneidet man bei geeigneten Grenzen ab. Als Maß benutzt man dafür häufig die Standardabweichung σ (zum Beispiel $\pm 5\sigma$ um den Mittelwert).

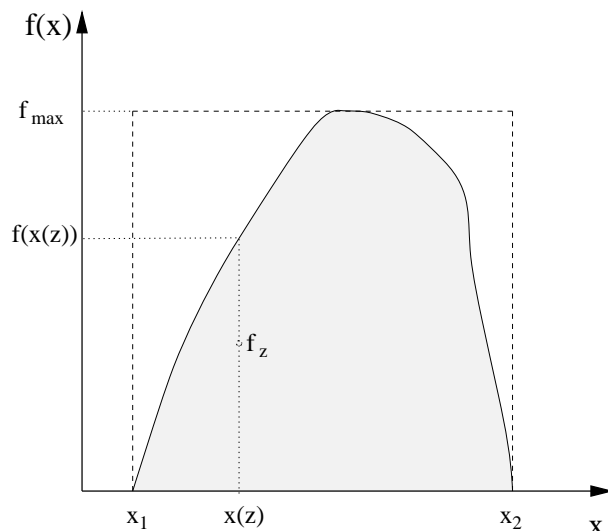


Abbildung 1.8: Zur Erklärung der ‘Hit and Miss’ Methode.

1.3.2 ‘Hit and Miss’ Methode

Wenn die Wahrscheinlichkeitsdichte sehr unübersichtlich wird, insbesondere bei Abhängigkeit von mehreren Variablen oder wenn man davor zurückschreckt, analytische Berechnungen zu machen, kann man Ereignisse nach der ‘Hit and Miss’ Methode erzeugen.

Sei x eine Zufallsvariable mit der Wahrscheinlichkeitsdichte $f(x)$ (Abb. 1.8). Sowohl x als auch $f(x)$ sollte in einem endlichen Intervall liegen:

$$\begin{aligned} x_1 &\leq x \leq x_2 \\ 0 &\leq f(x) \leq f_{max} \end{aligned} \quad (1.57)$$

Falls das nicht gegeben ist, kann man sich häufig auf relevante Bereiche beschränken, siehe oben. Der ‘Hit and Miss’ Algorithmus lautet dann:

- (i) Erzeuge x gleichverteilt im Intervall $[x_1, x_2]$;
- (ii) erzeuge einen Wert f_z gleichverteilt im Intervall $[0, f_{max}]$;
- (iii) akzeptiere x falls $f_z \leq f(x)$;
- (iv) wiederhole.

Es werden also Punkte $x(z), f(x(z))$ gleichverteilt in der Box (1.57) erzeugt. Ein Punkt wird als Treffer gezählt, wenn er unterhalb der Kurve $f(x)$ liegt. Die so erzeugten Treffer x folgen der Verteilung $f(x)$ normiert auf das eventuell beschränkte Intervall.

Die benötigte Transformation einer Gleichverteilung im Einheitsintervall $[0, 1]$ auf eine beliebige Gleichverteilung zum Beispiel in $[x_1, x_2]$ ergibt sich aus der entsprechenden Umkehrfunktion:

$$z = \frac{\int_{x_1}^x dx}{\int_{x_1}^{x_2} dx} = \frac{x - x_1}{x_2 - x_1} \implies x = x_1 + z \cdot (x_2 - x_1) \quad (1.58)$$

Die ‘Hit and Miss’ Methode ist nicht sehr effizient, wenn sehr große Werte der Wahrscheinlichkeitsdichte $f(x)$ in sehr kleinen x -Intervallen auftreten ($f(x) \rightarrow \infty$ ist möglich, solange das Integral über $f(x)$ endlich bleibt). Dann benutzt man andere Verfahren, die wir teilweise in einem späteren Kapitel besprechen werden.

