

Stochastik für Informatiker

SS 2008

M. Schäl

Vorwort

Alle Vorgänge der Natur enthalten eine zufällige Komponente. Das Wirken zufälliger Faktoren zu beschreiben und daraus praktische Folgerungen zu ziehen, ist Aufgabe der Stochastik, die sich aus der Wahrscheinlichkeitstheorie (kurz W–Theorie) und der mathematischen Statistik zusammensetzt. Die Stochastik ist in den letzten Jahrzehnten eines der zentralen Gebiete der angewandten Mathematik geworden, die in wechselseitigen Beziehungen zu vielen anderen mathematischen Fachrichtungen steht. Sie spielt gleichzeitig eine zentrale Rolle bei der Mathematisierung verschiedener angewandter Wissenschaften. In der ferneren Vergangenheit zerfiel die Stochastik weitgehend in eine kombinatorische Behandlung von Glücksspielen und in die sogenannte Fehlertheorie rund um die Normalverteilung. (Die letztere ist durch die Gaußsche Glockenkurve charakterisiert, die etwa auf dem 10 DM–Schein abgebildet war.)

Die historische Entwicklung der Stochastik ist von einer intensiven und äußerst fruchtbaren Wechselwirkung zwischen Theorie und Anwendungen geprägt. Ihre Methoden sind heute unentbehrlich für so unterschiedliche Anwendungsbereiche wie Versicherungswesen (Prämienkalkulation), Wirtschaftswissenschaften (Portfolio–Analyse, Optionsbewertung), Physik (Quantentheorie), Naturwissenschaften (Planung und Auswertung von Versuchen), Medizin (Vergleich verschiedener Therapien), Epidemiologie (Ausbreitung von Krankheiten), Informatik (Auslastung in Telefon– und Daten–Netzen, Antwortzeiten im Rechner), Verkehrswesen (Warteschlangensysteme), Meinungsforschung (repräsentative Stichproben),

Die Stochastik ist gleichzeitig eine große mathematische Disziplin [mit all deren Kennzeichen: reizvolle gelöste und ungelöste Probleme, interessante Methoden, strenge Begründungen und umfassende kohärente Theorien]. Durch die einzigartige Kombination von konkreten und anschaulichen Ideen mit tiefliegenden und oft abstrakten Theorien übt sie eine besondere Anziehungskraft aus.

Die W–Theorie befaßt sich mit stochastischen Vorgängen, die also vom Zufall beeinflußt sind und somit nicht vorhersagbare Ergebnisse hervorbringen. Im Gegensatz zu deterministischen mathematischen Modellen ist die Wahl eines passenden Modells für stochastische Vorgänge oft weniger offensichtlich. Zeitweise haben Mathematiker sogar geglaubt, es läge im Wesen der zufälligen Erscheinungen, daß sie sich nicht mathematisieren ließen. So hat es bis in dieses Jahrhundert hinein gedauert, daß man eine gesicherte axiomatische Grundlegung gegeben hat.

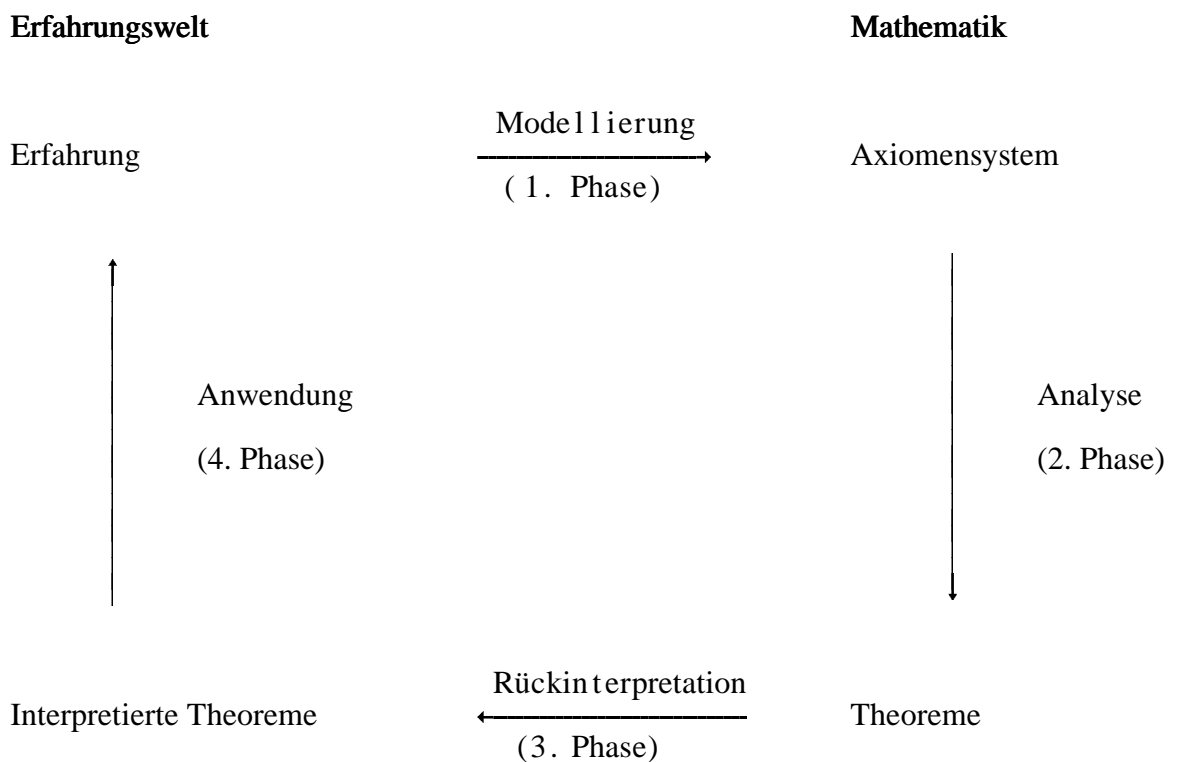
Die moderne W–theorie arbeitet mit analytischen Hilfsmitteln, die in der ersten Hälfte des letzten Jahrhunderts geschaffen wurden. Als Folge kam es kurze Zeit später zu einer Erweiterung der W–Theorie durch die Theorie der stochastischen Prozesse. Diese dienen als mathematisches Modell für stochastische Vorgänge, die in einem Zeitintervall ablaufen.

Literatur: Ulrich Krenzel: Einführung in die Wahrscheinlichkeitstheorie und Statistik
vieweg studium Aufbaukur Mathematik

Kapitel I Diskrete Wahrscheinlichkeitsräume

§0 Einführung

Eines der Grundziele der Mathematik besteht darin, Erfahrungen des Menschen über Vorgänge der Natur in Modelle, d. h. Axiome, umzusetzen, diese mathematisch zu analysieren, also Eigenschaften und Gesetzmäßigkeiten (Theoreme) aus ihnen abzuleiten und durch Rückinterpretation in die Erfahrungswelt nutzbar zu machen. Dabei zeigt es sich häufig, daß Modifizierungen am mathematischen Modell nötig werden, die wiederum Anlaß zu erneuter Analyse und Rückinterpretation sind, u. s. w. . Das folgende 4 – Phasen – Schema verdeutlicht diesen Kreislauf.



Ein wichtiges Ziel dieser Vorlesung ist die Entwicklung eines Basismodells zur Beschreibung und mathematischen Analyse des " Zufalls ". Der Begriff "zufällig" wird auch durch die Vokabel "**stochastisch**" ausgedrückt werden. Die Bezeichnung "**Stochastik**" wird als Sammelbegriff für die Gebiete Wahrscheinlichkeitstheorie und Mathematische Statistik verwendet.

Die Wahrscheinlichkeitstheorie (kurz **W–Theorie**) befaßt sich mit der Analyse (Beschreibung und Gesetzmäßigkeiten) von stochastischen Vorgängen. Anstelle des Begriffes " zufälliger Vorgang " führen wir den Terminus technicus "**Zufallsexperiment**" (kurz **Experiment**) ein. Der Begriff Zufallsexperiment steht für eine Situation, die ein vom Zufall beeinflusstes, also nicht vorhersagbares Ergebnis hervorbringt.

Der Übergang von der Wirklichkeit zum Modell ist dabei nie rein logisch begründbar. Er setzt in starkem Maß Erfahrung über die Natur des Experiments voraus. Das ist keine Besonderheit der Modelle für Zufallsexperimente. So genügt z. B. das ebene Modell der Erdoberfläche vollauf, wenn man eine Landkarte des Rhein–Sieg–Kreises herstellen will. Für feine geophysikalische Betrachtungen ist selbst das Modell der Erdkugel zu grob und man betrachtet im feineren Modell Abplattungen. Wir sehen daran gleich, daß die Wahl des Modells von der Zielsetzung mitbestimmt wird.

Wir halten fest, daß es keinen **prinzipiellen** Unterschied zwischen Modellen in der Geometrie und der W–Theorie gibt. **Praktisch** scheint es aber Unterschiede zu geben, weil das passende Modell für Zufallsexperimente oft weniger offensichtlich ist. Häufig läßt sich die Richtigkeit eines Modells nur empirisch prüfen, und das ist stets mit Unsicherheiten behaftet, die mit dem zufälligen Ausgang der Experimente zusammenhängen. So braucht man zur Überprüfung, ob ein Würfel unverfälscht ist, eine sehr große Zahl von Würfeln und mit **absoluter** Sicherheit läßt sich auch dann noch nicht die Unverfälschtheit belegen. Man täuscht sich leicht darüber, ob ein Versuch wirklich hinreichend viele Symmetrien enthält, um die Annahme zu rechtfertigen, alle Versuchsausgänge seien gleichwahrscheinlich.

Zeitweise haben Mathematiker sogar geglaubt, es läge im Wesen der zufälligen Erscheinungen, daß sie sich nicht mathematisieren ließen. Jedenfalls hat es – im Gegensatz z. B. zur Geometrie – bis in das 20. Jahrhundert hinein gedauert, bis man eine gesicherte axiomatische Grundlegung gegeben hat. Andererseits macht gerade diese Tatsache, daß man über Zufallsereignisse mathematisch rigorose Resultate beweisen kann, einen Reiz dieses Gebiets aus.

Hilbert stellte 1900 eine Liste der 23 wichtigsten damals ungelösten mathematischen Probleme auf. Das 6. Problem fragte gerade nach einer Begründung der Wahrscheinlichkeitstheorie, die den neuen, strengen Ansprüchen der Mathematik gerecht wird.

Der Beginn der Wahrscheinlichkeitsrechnung liegt im 17. Jahrhundert. (Dabei wird der Begriff W–Rechnung anstelle von W–Theorie verwendet, wenn die Theorie ohne Maßtheorie entwickelt wird.) Die heute allgemein akzeptierte axiomatische Einführung der Wahrscheinlichkeit wurde im Jahre 1933 durch A. N. Kolmogoroffs "Grundbegriffe der Wahrscheinlichkeitsrechnung " gegeben.

Die W–Theorie ist besonders wertvoll zur Beschreibung von zufälligen Experimenten, die oft und unter gleichen Bedingungen und im Idealfall ohne gegenseitige Beeinflussung wiederholt werden. Dazu zählen etwa die Notierung von Lebenszeiten oder Schadensfällen in einem Versicherungsunternehmen, die Beobachtung des Heilungserfolges nach Verabreichung eines medizinischen Medikaments oder die Notierung der Nachfrage nach einer bestimmten Ware an verschiedenen Tage in einem Wirtschaftsunternehmen.

0.1 Ein Finanzmarktmodell

Es werden **einperiodige** Modelle zugrundegelegt (also mit den Zeitpunkten $t = 0, 1$). Der Markt wird durch die folgende **Situation** beschrieben:

Es bestehen 2 Anlagemöglichkeiten jeweils zum Zeitpunkt $t = 0$:

Eine der Anlagemöglichkeiten ist das *Sparbuch/Bankkonto (bond)*, das durch den *Zinssatz* r beschrieben wird. Dabei wird angenommen, daß der gleiche Zinssatz r gilt, wenn man Geld anlegt und wenn man Geld leiht. Dies ist natürlich eine Modelleinschränkung. Eine Anlage von η Euro z.Zt. $t=0$ hat z.Zt. $t=1$ den Wert $\eta \cdot (1+r)$.

Die andere Anlagemöglichkeit ist ein risikobehaftete *Wertpapier (stock)* [z.B. Aktien oder Devisen, etwa Dollar]; sein Preis oder Kurs z.Zt. $t=0$ sei S_0 und der z.Zt. $t=1$ sei S_1 . Eine Anlage von $\xi \in \mathbb{R}$ Anteilen in das kursabhängige Wertpapier kostet z.Zt. $t=0$ $\xi \cdot S_0$ Euro und hat z.Zt. $t=1$ den Wert $\xi \cdot S_1$. Die Anlage (η, ξ) kostet also z.Zt. $t=0$ $\eta + \xi \cdot S_0$ Euro und hat z.Zt. $t=1$ den Wert $\eta \cdot (1+r) + \xi \cdot S_1$.

[Negative Anteile ξ entsprechen sogenannten *Wertpapierleerverkäufen*. D.h. etwa, daß man Aktien verkauft, die man noch nicht besitzt, die man also einem anderen Investor schuldet.] Die Größen S_0 und S_1 werden stets als positiv angenommen.

Der Unsicherheit über die Marktentwicklung soll dadurch Rechnung getragen werden, daß mehrere Möglichkeiten (Szenarios) zugelassen werden. In dem hier gewählten Modell sollen genau zwei Möglichkeiten betrachtet werden. Wir machen folgende Annahme:

Es sei $S_0 > 0$ ein bekannter, gegebener Wert. Die beiden möglichen Ergebnisse werden durch die Symbole u (up) und d (down) repräsentiert und in einem zweielementigen sogenannten Ergebnisraum Ω zusammengefaßt

$$(0.1) \quad \Omega = \{u, d\}.$$

Dabei wird es also die Möglichkeit geben, daß der Kurs in einer bestimmten Weise steigt oder in einer bestimmten Weise fällt.

$$\begin{array}{c}
 S_1(u) \\
 \swarrow \\
 S_0 \\
 \searrow \\
 S_1(d)
 \end{array}
 \quad
 \text{Beispiel: }
 \begin{array}{c}
 1,02 \cdot S_0 \\
 \swarrow \\
 S_0 \\
 \searrow \\
 \frac{1}{1,02} S_0
 \end{array}$$

Funktionen auf Ω bezeichnen wir als Zufallsvariablen (Zva). Es sei hier $S_1(\omega)$ eine Zva mit

$$(0.2) \quad S_1(u) > S_0 \cdot (1+r) > S_1(d).$$

Eine Investition in das Wertpapier ist also entweder günstiger oder ungünstiger als eine Investition auf das Sparkonto.

Es bezeichne $\xi \in \mathbb{R}$ immer die Anteile, die der Investor in der Periode hält.

Gibt man die *Anfangsinvestition* x vor, so ist der Stand auf dem Sparbuch z.Z. 1 bestimmt durch η und η wiederum durch die *Budgetgleichung*:

$$(0.3) \quad \eta + \xi \cdot S_0 = x.$$

0.4 Definition. $V_{(x,\omega)}^{\xi}$ ist der **Endwert** z.Zt. $t=1$ zu ξ und x mit

$$(0.5) \quad V_{(x,\omega)}^{\xi} := \eta \cdot (1+r) + \xi \cdot S_1(\omega) ,$$

wobei η jeweils durch (0.3) festgelegt wird.

Also ergibt sich

$$(0.6) \quad V_{(x,\omega)}^{\xi} = x \cdot (1+r) + \xi \cdot [S_1(\omega) - (1+r) \cdot S_0] .$$

Im Mittelpunkt wird die *Bewertung von Derivaten* stehen. Dies sind Verträge, die eine *Zahlung* $X(\omega)$ zusichern, der sich aus der Kursentwicklung S_0, S_1 herleiten läßt in der Weise, daß etwa $X(\omega) = \psi(S_1(\omega))$ gilt.

Für diesen Vertrag muß z. Zt. 0 eine *Prämie* (ein *Preis*) bezahlt werden, die gerade die **Bewertung** des Finanztitels darstellt.

0.7 Def. Ein **Zahlungsanspruch** (*contingent claim*) ist eine Funktion (Zva) $X(\omega)$ auf Ω .

Bei Optionen z.B. wird dieser Zahlungsanspruch X in jedem Fall nichtnegativ sein, sodaß der Vertragsunterzeichner (Käufer) also ohne die Prämie in jedem Fall einen Gewinn erzielen würde.

Im klassischen Fall würde

$$E\left[\frac{X}{1+r}\right] := p \cdot \frac{X(u)}{1+r} + (1-p) \cdot \frac{X(d)}{1+r}$$

als faire Prämie angesehen werden, wenn u mit der relativen Häufigkeit p und d mit der relativen Häufigkeit $1-p$ eintritt. Die Diskontierung ist nötig, weil die Zahlung X erst in $t=1$ erfolgt.

Hier wird aber eine andere Antwort gegeben werden. Eine faire Prämie hat zwar die obige Gestalt, aber mit einem künstlichen p^* , das unabhängig von der auf dem Markt beobachteten Häufigkeit p ist, die nicht bekannt zu sein braucht.

Als Beispiel dazu betrachten wir die folgende Situation:

Wenn der Verkäufer durch einen Vertrag eine Zahlung $X = a \cdot S_1$ zusichert, so ist $a \cdot S_0$ eine *faire Prämie*; denn offenbar kann der Verkäufer die Prämie von $a \cdot S_0$ sofort in das Wertpapier investieren und hat dann z. Zt. $t=1$ den auszahlenden Betrag $X = a \cdot S_1$ zur Verfügung.

Der Verkäufer geht dann also kein Risiko ein. Ebenso kann der Käufer anstelle des Vertrages selbst $a \cdot S_0$ z.Zt. $t=0$ in das kursabhängige Wertpapier investieren. In dem Sinn sind also die Prämie und der Vertrag gleichwertig. Dies gilt noch in allgemeineren Situationen.

0.8 Definition. Ein Zahlungsanspruch X heißt *erreichbar* (attainable) oder *duplizierbar* [durch (x, ξ)], wenn gilt: $V_{(x, \xi)}^{\xi}(x, \omega) = X(\omega) \forall \omega$ für gewisse Zahlen ξ und $x \in \mathbb{R}$.

0.9 Definition. In der Situation $V_{(x, \xi)}^{\xi}(x, \omega) = X(\omega) \forall \omega$ ist x eine *faire Prämie* für den Zahlungsanspruch X .

In der Situation von 0.9 kann sich der Verkäufer nämlich gegen den Zahlungsanspruch X absichern (hedging), indem er die Prämie x benutzt, um gemäß (x, ξ) zu investieren. Dann hat er z. Zt. $t=1$ den Betrag $V_{(x, \xi)}^{\xi}(x) = X$ zur Verfügung. Das gleiche kann jeder andere Marktteilnehmer tun. Im vorliegenden speziellen Modell wird noch gezeigt, daß jeder Zahlungsanspruch erreichbar ist und somit abgesichert werden kann.

0.10 Beispiel. Optionen. Eine *europäische (Kauf-) Option* [call option] ist ein Vertrag, der dem Käufer das Recht einräumt, zum *Fälligkeitstermin* $t=1$ (maturity time, exercise time) für einen festen *Wahrnehmungs- / Basispreis* K (exercise/striking price), unabhängig von dem vorliegenden Kurs S_1 , a Anteile des Wertpapiers zu kaufen [, z.B.

$K = a \cdot (1+r) \cdot S_0$]. Dann ist der Gewinn des Käufers und damit der Verlust des Verkäufers

$$X(\omega) := \psi(S_1(\omega)) := [a \cdot S_1(\omega) - K]^+ = \max(0, a \cdot S_1(\omega) - K).$$

Es ist also günstig für den Käufer, wenn der Kurs steigt; es ist günstig für die Bank, wenn der Kurs fällt. Um den Verlust bei steigendem Kurs abzusichern (hedging), kann die Bank selbst Wertpapiere kaufen . Dadurch kann es auch günstig für den Verkäufer werden, wenn der Kurs steigt. \square

Es soll wieder ein Modell betrachtet werden, in dem sich der Kurs in einer bestimmten Weise nach unten oder noch oben bewegen kann mit (0.2).

X sei der Zahlungsanspruch zu einem Derivat. Es soll gezeigt werden, daß im vorliegenden Fall X stets erreichbar ist. Gesucht sind also gemäß (0.6) x und ξ mit

$$\begin{aligned} (0.11) \quad & (1+r) \cdot x + \xi \cdot [S_1(\omega) - (1+r) \cdot S_0] = X(\omega) \quad \text{für } \omega = d, u \\ \Leftrightarrow & x + \xi \cdot [S_1(\omega)/(1+r) - S_0] = x_{\omega} \quad \text{für } \omega = d, u \quad \text{mit } x_{\omega} := \frac{1}{1+r} X(\omega) \\ \Leftrightarrow & x + \xi \cdot R(\omega) = x_{\omega} \quad \text{für } \omega = d, u \quad \text{mit } R(\omega) := S_1(\omega)/(1+r) - S_0. \end{aligned}$$

0.12 Lemma. Das Gleichungssystem $x + \xi \cdot R(\omega) = x_{\omega}$, für $\omega = d, u$, in $(x, \xi) \in \mathbb{R}^2$ mit $R(d) < 0 < R(u)$ hat eine eindeutige Lösung; dabei ist:

$$x = p_d x_d + p_u x_u, \quad \xi = (x_u - x_d) / (R(u) - R(d)), \quad \text{mit} \\ p_d := \frac{R(u)}{R(u) - R(d)}, \quad p_u := 1 - p_d, \quad \text{und} \quad 0 < p_{\omega} < 1.$$

Würde S_1 mehr als zwei Werte annehmen, so hätte man für die beiden Unbekannten x und ξ mehr als zwei Bestimmungsgleichungen. Der **Beweis** von 0.12 ist einfach.

X ist also erreichbar durch (x, ξ) . Ferner ist p_ω eine W -Funktion (Zähldichte) gemäß

0.13 Definition. Eine Funktion $P(\omega)$ auf Ω mit $P(\omega) \geq 0$ und $\sum_{\omega \in \Omega} P(\omega) = 1$ heißt *Wahrscheinlichkeitsfunktion* (W -Funktion)/*Zähldichte*.

Beschreibt Ω die Menge aller möglichen Ergebnisse bei einem (wiederholbaren) zufallsabhängigen Experiment [und ist Ω endlich], so soll $P(\omega)$ ein Maß für die relative Häufigkeit des Eintretens von ω sein. Besteht das Experiment in dem Werfen einer Münze und beschreibt u Zahl, d Wappen, so wird man $P(\omega) = \frac{1}{2}$ wählen.

0.14 Bemerkung. Wählt man eine künstliche W -Funktion P^* auf Ω mit $P^*(\omega) = p_\omega$ wie in 0.12, so gilt:

$$(0.15) \quad x = E^* \left[\frac{X}{1+r} \right] := p_u \cdot \frac{X(u)}{1+r} + p_d \cdot \frac{X(d)}{1+r},$$

$$(0.16) \quad E^* \left[\frac{S_1}{1+r} \right] := p_u \cdot \frac{S_1(u)}{1+r} + p_d \cdot \frac{S_1(d)}{1+r} = S_0 \quad \text{wegen}$$

$$(0.17) \quad p_u \cdot R(u) + p_d \cdot R(d) = 0.$$

P^* ist die einzige Wahrscheinlichkeitsfunktion mit $(0.16) \Leftrightarrow E^* \left[\frac{S_1}{S_0 \cdot (1+r)} \right] = 1$. Dies bedeutet, daß unter P^* das Verhältnis einer Anlage von S_0 auf dem Bankkonto und einer Anlage von S_0 in das Wertpapier im Mittel gleich bleibt. P^* heißt deswegen *risiko-neutral*. Die Eigenschaft (0.16) besagt gerade, daß $S_0, S_1/(1+r)$ ein sogenanntes *Martingal* ist.

0.18 Beispiel. Im Falle einer europäischen Kaufoption mit $a=1$, also für

$$X(\omega) = [S_1(\omega) - K]^+ \quad \text{mit} \quad S_1(d) \leq K < S_1(u)$$

gilt $\xi \cdot S_0 > x$, also

$$\eta = x - \xi_0 \cdot S_0 < 0;$$

zur Absicherung von X muß also zusätzlich Geld geliehen werden, um in das kursabhängige Wertpapier gemäß ξ zu investieren. Wegen $X(d) = 0$ hat man

$$x = p_u \cdot X(u)/(1+r)$$

$$\xi \cdot S_0 > x. \quad \square$$

Oft wird auch die Modellannahme mit einer Wahrscheinlichkeitsfunktion P formuliert, das wie P^* aussieht mit einem anderen Parameter p anstelle von P^* . Dann beschreibt $P(\omega)$ die relative Häufigkeit des Eintretens von ω . Der Wechsel von P zu P^* bedeutet gerade einen Wechsel vom Parameter p zu p^* . Offenbar wird diese Wahrscheinlichkeitsfunktion P hier aber nicht benötigt.

Gemäß 0.8 und 0.12/0.15 ist jeder Zahlungsanspruch X erreichbar mit einer Anfangsinvestition

$$x = E^* \left[\frac{X}{1+r} \right].$$

0.2 Ergebnisräume

Wir beschäftigen uns also mit Experimenten, die nicht vorhersagbar sind, die also zu verschiedenen Versuchsausgängen/Ergebnissen/Szenarien führen können. Im ersten Schritt überlegen wir uns die Menge aller bei einem Versuch möglichen Ergebnisse. Die Gesamtheit dieser Ergebnisse wird als **Ergebnismenge**–**raum** bezeichnet; als Symbol verwendet man oft Ω . Die Elemente von Ω , also die möglichen Versuchsausgänge, heißen **Ergebnisse**; man verwendet das Symbol $\omega \in \Omega$. Die Teilmengen von Ω heißen **Ereignisse**. Wir identifizieren dabei $A \subset \Omega$ mit dem Ereignis, daß ein $\omega \in A$ der beobachtete Versuchsausgang ist. Diese Konvention erlaubt es, mengentheoretische Notationen zu gebrauchen. Die mengentheoretischen Operationen und Beziehungen lassen sich dabei in aussagenlogische Verknüpfungen übersetzen:

Dies soll an einem Beispiel erklärt werden. Wir wollen das Experiment betrachten, das aus dem zweimaligen Werfen eines Würfels besteht. Jedes Ergebnis kann durch ein Paar $\omega = (i,k)$ beschrieben werden. So bedeutet das Paar (i,k) gerade, daß beim ersten Wurf die Zahl i und beim zweiten Wurf die Zahl k gewürfelt wurde. Wir wählen also

$$\Omega := \{ \omega = (i,k); 1 \leq i, k \leq 6 \} = \mathfrak{X} \times \mathfrak{X} = \{ (i,k); i \in \mathfrak{X}, k \in \mathfrak{X} \} \quad \text{mit } \mathfrak{X} := \{1, \dots, 6\}.$$

6
5
4
3
2
1
	1	2	3	4	5	6

Die Wahl von Ω als ein Produktraum / kartesisches Produkt ist typisch für mehrstufige Experimente. Es ist vorteilhaft, die Ergebnisse der einzelnen Würfe durch Funktionen (hier Projektionen) auf Ω zu beschreiben. Wir setzen

$$X_1(\omega) = X_1((i,k)) := i \hat{=} \text{Resultat des ersten Wurfes}$$

$$X_2(\omega) = X_2((i,k)) := k \hat{=} \text{Resultat des zweiten Wurfes.}$$

Diese Funktionen sind Beispiele für den wichtigen Begriff einer **Zufallsvariablen**, der später eingeführt werden wird. Nun betrachten wir das Ereignis

$$A := \{ \omega \in \Omega; X_1(\omega) + X_2(\omega) \geq 10 \} = \{(6,6), (6,5), (5,6), (6,4), (4,6), (5,5)\}.$$

Dabei ist die Anzahl der Elemente von Ω , also $\text{card}(\Omega) = 6^2$, $\text{card}(A) = 6$ und somit $\text{card}(A)/\text{card}(\Omega) = 1/6$. Da ein symmetrischer / unverfälschter Würfel vorliegen soll, wird jede Zahl auf lange Sicht gleich häufig auftreten. Somit wird, aus diesen Symmetriegründen, in einem Sechstel der Fälle auf lange Sicht das Ereignis A eintreten. Wir werden in einem mathematischen Modell dem Ereignis A die Wahrscheinlichkeit $1/6$ zuordnen.

Es empfiehlt sich, wie auch später, die folgende verkürzte und anschauliche Notation einzuführen:

$$A =: \{X_1 + X_2 \geq 10\}.$$

Das **Komplement** von A wird mit $A^c = \{\omega \in \Omega; \omega \notin A\}$ bezeichnet. A^c steht für das Ereignis, daß A nicht eintritt. Hier ist also $A^c = \{X_1 + X_1 < 10\}$.

Sei nun

$$B := \{X_1 = X_2\} = \{\omega \in \Omega; X_1(\omega) = X_2(\omega)\} = \{(6,6), (5,5), (4,4), (3,3), (2,2), (1,1)\}.$$

Dann ist $\text{card}(B) = 6$ und $\text{card}(B)/\text{card}(\Omega) = 1/6$. Somit wird wieder in einem Sechstel der Fälle auf lange Sicht das Ereignis B eintreten.

Für den **Durchschnitt** von A und B gilt:

$$A \cap B = \{\omega \in \Omega; \omega \in A, \omega \in B\} = \{\omega \in \Omega; \omega \in A \text{ und } \omega \in B\}.$$

Dann steht also $A \cap B$ für das Ereignis, daß sowohl A als auch B eintreten.

Hier ist also $A \cap B = \{X_1 + X_2 \geq 10; X_1 = X_2\} = \{(6,6), (5,5)\}$.

Wir setzen noch

$$C := \{X_1 + X_2 = 9\} = \{(3,6), (4,5), (5,4), (6,3)\}.$$

Für die **Vereinigung** von A und C gilt:

$$A \cup C = \{\omega \in \Omega; \omega \in A \text{ oder } \omega \in C\}.$$

Dann steht also $A \cup C$ für das Ereignis, daß A oder C eintreten (im nicht ausschließenden Sinne).

Hier gilt also $A \cup C = \{X_1 + X_2 \geq 9\}$.

Zeichne A, B, C, $A \cap B$, $A \cup C$ ein:

6
5
4
3
2
1
	1	2	3	4	5	6

Die folgenden Distributivgesetze sind bekannt:

1. Distributivgesetz: $(A \cup C) \cap B = (A \cap B) \cup (C \cap B)$

2. Distributivgesetz: $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

Die **Regeln von de Morgan** besagen:

$$(A \cup C)^c = A^c \cap C^c; \quad (A \cap B)^c = A^c \cup B^c.$$

Eine Vereinigung ist also als Durchschnitt und ein Durchschnitt als Vereinigung darstellbar:

$$A \cup C = (A^c \cap C^c)^c; \quad A \cap B = (A^c \cup B^c)^c.$$

Die **mengentheoretische Differenz** ist definiert gemäß:

$$A \setminus B := A \cap B^c = \{\omega \in \Omega; \omega \in A \text{ und } \omega \notin B\},$$

dann gilt also $A^c = \Omega \setminus A$ sowie $A \cup B = A \cup (B \setminus A)$.

Im Falle $A \cap C = \emptyset$ sagt man die Mengen A und C sind **disjunkt**; die Ereignisse A und C sind **unvereinbar**.

Man spricht von einer **disjunkten Zerlegung** von Ω , falls:

$$\Omega = \bigcup A_i, \quad A_i \cap A_j = \emptyset \text{ f\u00fcr } i \neq j, \text{ [d.h. die } (A_i) \text{ sind paarweise disjunkt].}$$

Im Beispiel haben wir: $\Omega = \bigcup_{j=1}^6 \{X_1 = j\}$ sowie

$$\Omega = \bigcup_{j=2}^{12} C_j, \text{ mit } C_j = \{X_1 + X_2 = j\}, \text{ also } C = C_9.$$

Wir haben dann $C_{10} \subset A$, also $\{X_1 + X_2 = 10\} \subset \{X_1 + X_2 \geq 10\}$, C_{10} impliziert A,

$$A = \bigcup_{j=10}^{12} C_j = \{\omega \in \Omega; \omega \in C_j \text{ f\u00fcr mindestens ein } j \text{ mit } 10 \leq j \leq 12\} \\ = \{\omega \in \Omega; \exists j \in \{10 \leq j \leq 12\} \text{ mit } \omega \in C_j\}.$$

An diesem Beispiel sieht man auch, da\u00df ein Vereinigung $\cup_j C_j$ auch die logische Beziehung \exists ausdr\u00fccken kann ($\exists j$ soda\u00df C_j eintritt).

Eine disjunkte Zerlegung ist nicht nur f\u00fcr Ω zweckm\u00e4\u00dfig. So k\u00f6nnen wir auch C zerlegen gem\u00e4\u00df

$$C = C \cap \Omega = C \cap \left[\bigcup_{j=1}^6 \{X_1 = j\} \right] = \bigcup_{j=1}^6 C \cap \{X_1 = j\} \\ = \bigcup_{j=1}^6 \{X_1 + X_2 = 9, X_1 = j\}, \text{ also} \\ \{X_1 + X_2 = 9\} = \bigcup_{j=1}^6 \{X_1 = j, X_2 = 9-j\}.$$

Setzt man $A_j := \{X_1 + X_2 \geq j\}$ (also $A = A_{10}$), dann gilt: $A_j \supset A_{j+1}$;

$$A_{12} = \{(6,6)\} = \bigcap_{j=2}^{12} A_j = \{\omega \in \Omega; \omega \in A_j \text{ f\u00fcr alle } j\}.$$

Setzt man noch

$$\tilde{A} := \{\max(X_1, X_2) \geq 5\} = \{\omega \in \Omega; \max(X_1(\omega), X_2(\omega)) \geq 5\}, \text{ so gilt:}$$

$$\tilde{A} = \{X_1 \geq 5\} \cup \{X_2 \geq 5\} \text{ und } \tilde{A}^c = \{\max(X_1, X_2) \leq 4\} = \{X_1 \leq 4\} \cap \{X_2 \leq 4\} \text{ (de Morgan).}$$

Offenbar ist $\text{card}(\tilde{A}^c) = 4^2 = 16$ und $\text{card}(\tilde{A}) = 36 - 16 = 20$, $\text{card}(\tilde{A})/\text{card}(\Omega) = 20/36 = 5/9$.

Somit wird wieder in 5/9 aller F\u00e4lle auf lange Sicht das Ereignis \tilde{A} eintreten.

Zeichne \tilde{A} , $\{X_j \geq 5\}$ ein:

6
5
4
3
2
1
	1	2	3	4	5	6

Wir merken uns:

$A^c \hat{=} \text{das Ereignis tritt nicht ein}$

$A \cap B \hat{=} \text{das Ereignis A und das Ereignis B treten ein}$

$A \cup B \hat{=} \text{das Ereignis A oder das Ereignis B treten ein}$

$A \subset B \hat{=} \text{das Ereignis A impliziert das Ereignis B}$

\emptyset heißt **unmögliches Ereignis**;

Ω heißt **sicheres Ereignis**.

Standardmodell.

Besteht das Gesamtexperiment aus n Einzelexperimenten (oben ist $n=2$), so wird ein Ergebnis durch ein n -Tupel dargestellt, wobei die m -te Komponente das Ergebnis des m -ten Einzelperiments beschreibt. Dies ist das wichtigste Modell, daß gerade ein Gesamtexperiment beschreibt, daß aus der n -fachen Wiederholung eines Einzelperiments besteht. Ist die **endliche Menge** \mathfrak{X} die Menge aller möglichen Ergebnisse des Einzelperiments, so wählen wir für ein Modell für die n -fache Wiederholung

$$\Omega = \mathfrak{X} \times \dots \times \mathfrak{X} = \mathfrak{X}^n.$$

Die Funktionen $X_m : \Omega \rightarrow \mathfrak{X}$ seien wieder die Projektionen auf die m -te Koordinate, also $X_m(\omega) := x_m$ für $\omega = (x_1, \dots, x_n)$. Dann schreiben wir wieder

$$\begin{aligned} \{X_m = x_m\} &:= \{\omega \in \Omega; X_m(\omega) = x_m\} \quad \text{für } x_m \in \mathfrak{X} \text{ oder} \\ \{X_\ell = x_\ell, X_m = x_m\} &:= \{\omega \in \Omega; X_\ell(\omega) = x_\ell, X_m(\omega) = x_m\} = \{X_\ell = x_\ell\} \cap \{X_m = x_m\}. \end{aligned}$$

Für ein beliebiges Ereignis $B \subset \Omega$ ergibt sich eine Zerlegung:

$$B = \bigcup_{x \in \mathfrak{X}} B \cap \{X_1 = x\} \quad \text{wegen } \Omega = \bigcup_{x \in \mathfrak{X}} \{X_1 = x\}$$

insbesondere für eine andere Funktion (Zva) $Y : \Omega \rightarrow \mathcal{Y}$, wobei \mathcal{Y} wieder eine endliche Menge ist (in der Regel $\mathcal{Y} \subset \mathbb{R}$):

$$\{Y = y\} = \bigcup_{x \in \mathfrak{X}} \{Y = y, X_1 = x\}.$$

Für $A_m \subset \mathfrak{X}$ ergibt sich die Zerlegung:

$$\begin{aligned} \{X_1 \in A_1, \dots, X_n \in A_n\} &= \{(X_1, \dots, X_n) \in A_1 \times \dots \times A_n\} \\ &= \bigcup_{(x_1, \dots, x_n) \in A_1 \times \dots \times A_n} \{(X_1, \dots, X_n) = (x_1, \dots, x_n)\} \\ &= \bigcup_{x_1 \in A_1} \dots \bigcup_{x_n \in A_n} \{X_1 = x_1, \dots, X_n = x_n\}. \end{aligned}$$

Dann ergibt sich wieder

$$\{Y = y\} = \bigcup_{x_1 \in \mathfrak{X}} \cdots \bigcup_{x_n \in \mathfrak{X}} \{Y = y, X_1 = x_1, \dots, X_n = x_n\}.$$

Im Beispiel $Y(\omega) = S(\omega) = X_1(\omega) + \dots + X_n(\omega)$ ergibt sich dann

$$\{S = k\} = \bigcup_{x_1 \in \mathfrak{X}, \dots, x_n \in \mathfrak{X}, x_1 + \dots + x_n = k} \{X_1 = x_1, \dots, X_n = x_n\}.$$

Dabei besteht $\{X_1 = x_1, \dots, X_n = x_n\}$ nur aus der einelementigen Menge $\{\omega\} = \{(x_1, \dots, x_n)\}$.

§1 Modelle für Zufallsexperimente, Abzählmethoden.

1.1 Endliche Wahrscheinlichkeitsräume

Sei Ω eine endliche Menge. Eine Abbildung $P(\omega)$ auf Ω heißt **Wahrscheinlichkeitsfunktion** (W–funktion), wenn die folgenden Eigenschaften gelten:

$$(1.11) \quad P(\omega) \geq 0 \text{ für alle } \omega \in \Omega, \quad \sum_{\omega \in \Omega} P(\omega) = 1.$$

Durch eine W–funktion wird jedes ω mit einem Gewicht $P(\omega)$ versehen, daß man sich als **Maß für die relative Häufigkeit** $h_n(\omega)$ vorstellen kann, mit der das Ergebnis ω bei einer großen Zahl n von Versuchen auftritt. In unserem Würfelbeispiel setzen wir aus Symmetriegründen $P(\omega) = 1/36$, da keine Zahl ausgezeichnet vor den anderen ist.

Die Bedeutung der W–Theorie gründet sich auf eine wesentliche Erfahrungstatsache (**Stabilität relativer Häufigkeiten, empirisches Gesetz der großen Zahlen**): $h_n(\omega)$ tendiert bei verschiedenen langen Versuchsreihen zu annähernd gleichen Werten. Dies führt dazu, daß man ω eine Wahrscheinlichkeit $P(\omega) \in [0,1]$ zuordnen möchte, die unabhängig von der speziellen Versuchsreihe ist, verbunden mit der Prognose: In $P(\omega) \cdot 100\%$ aller Fälle ist das Ergebnis eines Einzelexperiments gleich ω .

Offenbar besitzt die Funktion $h_n(\omega)$ die Eigenschaften von $P(\omega)$. [Bezeichnet nämlich $k_n(\omega)$ die absolute Häufigkeit bei einer n –maligen Wiederholung eines Experiments, also die Zahl der Einzelexperimente, bei den das Ergebnis ω eingetreten ist, so ist $h_n(\omega) = k_n(\omega)/n$ und somit

$$\sum_{\omega \in \Omega} h_n(\omega) = \frac{1}{n} \sum_{\omega \in \Omega} k_n(\omega) = \frac{n}{n} = 1.]$$

Nun soll auch Ereignissen A eine Wahrscheinlichkeit $P(A)$ zugeordnet werden. Die Menge der Ereignisse ist mengentheoretisch die Potenzmenge $\mathfrak{P}(\Omega)$, d.h. die Menge aller Teilmengen von Ω . Wir müssen also ein Funktion $P(A)$ für $A \subset \Omega$ ($\Leftrightarrow A \in \mathfrak{P}(\Omega)$) definieren. Dabei orientieren wir uns an der relativen Häufigkeit $h_n(A)$. Dann gilt offenbar

$$h_n(A) = \frac{1}{n} \sum_{\omega \in A} k_n(\omega) = \sum_{\omega \in A} h_n(\omega) \text{ für } A \subset \Omega.$$

Wir setzen nun

$$(1.10) \quad P(A) = \sum_{\omega \in A} P(\omega) \text{ für } A \subset \Omega.$$

Dies ist wieder ein Gewicht, daß man sich als Maß für die relative Häufigkeit vorstellen kann, mit der das Ereignis A bei einer großen Zahl von Versuchswiederholungen eintritt.

[Dabei identifizieren wir genau genommen $P(\{\omega\})$ und $P(\omega)$.]

Die Abbildung $P : \mathfrak{P}(\Omega) \rightarrow [0,1]$ heißt **Wahrscheinlichkeitsmaß** (W–Maß) [oder Wahrscheinlichkeitsverteilung] und hat die folgenden Eigenschaften:

$$(1.1) \quad P(\Omega) = 1 \text{ (Normiertheit);}$$

$$(1.2) \quad P(A) \geq 0 \text{ für alle } A \text{ (Positivität);}$$

$$(1.3) \quad P(A \cup B) = P(A) + P(B) \text{ für alle disjunkten } A, B \text{ (Additivität).}$$

Die Eigenschaft (1.3) folgt aus

$$\sum_{\omega \in A \cup B} P(\omega) = \sum_{\omega \in A} P(\omega) + \sum_{\omega \in B} P(\omega) \quad \text{für } A \cap B = \emptyset.$$

Das Paar (Ω, P) heißt der dem Experiment zugeordnete **Wahrscheinlichkeitsraum** (W–Raum).

[In der Regel geht man anders vor und definiert zunächst ein W–Maß und dann eine W–Funktion. Dann nennt man die Beziehungen (1.1) – (1.3) auch manchmal Axiome.]

Der Beweis der folgenden Eigenschaften von P ist sehr einfach.

Eigenschaften von P:

$$(1.4) \quad P(A^c) = 1 - P(A),$$

speziell $P(\emptyset) = 0$, (Nulltreue);

$$(1.5) \quad B \subset A \Rightarrow P(B) \leq P(A), \text{ (Isotonie);}$$

$$(1.6) \quad P(A \setminus B) = P(A) - P(B), \text{ für } B \subset A, \text{ (Subtraktivität);}$$

$$(1.7) \quad P\left[\bigcup_{i=1}^n A_i\right] = \sum_{i=1}^n P(A_i) \quad \text{falls } A_1, \dots, A_n \text{ (paarweise) disjunkt, (Additivität);}$$

$$(1.8) \quad P\left[\bigcup_{i=1}^n A_i\right] \leq \sum_{i=1}^n P(A_i) \quad \text{für beliebige } A_1, \dots, A_n \text{ (Subadditivität);}$$

$$(1.9) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \quad \text{für beliebige } A, B.$$

Offenbar impliziert (1.9) die Aussage (1.8) für $n=2$.

Das Würfelbeispiel ist ein Spezialfall der folgenden Situation.

(1) *Kombinatorische Bestimmung* von $P(\omega)$.

Geht man aus Symmetrie–Erwägungen davon aus, daß alle $\omega \in \Omega$ gleichwahrscheinlich sind, so heißt das W–Maß P **Gleichverteilung** auf Ω . (Ω, P) heißt dann ein **Laplacescher W–Raum**.

Bezeichnet $\text{card}(A)$ wieder die Anzahl der Elemente von A, so ist in diesem Fall

$$P(\omega) = 1/\text{card}(\Omega) \quad \text{für alle } \omega \in \Omega;$$

$$(1.12) \quad P(A) = \text{card}(A)/\text{card}(\Omega) \quad \text{für alle } A \subset \Omega.$$

Spricht man davon, daß ein Element einer Menge Ω *rein zufällig* ausgewählt wird, meint man in der Regel, daß man dieses Experiment durch einen Laplaceschen W–raum beschreiben kann.

(2) *Statistische Schätzung*.

Es stelle Ω etwa eine Menge von Ausprägungen einer Eigenschaften dar (Alter x , Körpergröße x , Konfektionsgröße x , Sympathie für eine Partei x , ...). Man befragt nun bei einer repräsentativen Umfrage n Leute nach einer dieser Eigenschaften, etwa ihrer Konfektionsgröße, und erhält dann eine sogenannte Stichprobe (x_1, \dots, x_n) .

Auf Grund dieser Stichprobe kann man dann das Experiment "Befragung einer zufällig eintreffenden Person nach ihrer Konfektionsgröße" für zukünftige Berechnungen durch eine W-Funktion P beschreiben, die gerade durch die relative Häufigkeit gegeben ist gemäß:

$$P(\omega) := h_n(\omega) = \text{card}(\{i; 1 \leq i \leq n, x_i = \omega\}) / n.$$

Dann ergibt sich $P(A) = h_n(A) = \text{card}(\{i; 1 \leq i \leq n, x_i \in A\}) / n$.

Als Wahrscheinlichkeit, daß eine zufällig eintreffende Person eine Konfektionsgröße ω verlangt, wählt man (für zukünftige Berechnungen) also die relative Häufigkeit, mit der in der Stichprobe (x_1, \dots, x_n) diese Konfektionsgrößen aufgetreten sind. [Hier gilt einmal nicht $\omega = (x_1, \dots, x_n)$!]

Beispiel.

Werfen wir eine Reißzwecke, so landet sie entweder mit der Spitze nach oben (up) oder nach unten (down). Bezeichnen wir diese beiden möglichen Ergebnisse mit u bzw. d , so kann man $\Omega = \{d, u\}$ wählen. Dieses ist offenbar ein Fall, der **nicht** mit einem Laplaceschen W-Raum beschrieben werden sollte. Offenbar ist eine W-Funktion hier bereits durch den Wert $P(u) =: p$ festgelegt wegen $P(d) = 1 - p$. Eine statistische Schätzung mit $n > 25\,000$ hat einmal den Wert $p = 0,5652$ ergeben. \square

Wir werden noch weitere Möglichkeiten kennenlernen, um zu Ansätzen für die Wahl einer geeigneten W-Funktion zu kommen. Zunächst sollen Laplacesche W-Räume untersucht werden. Dabei müssen offenbar Mengen abgezählt werden. Hilfsmittel dazu bietet die Kombinatorik.

1.2 Kombinatorik, einfache Urnenmodelle

Wir stellen uns die folgende Situation vor: In einer Urne sind N Kugeln, die wir uns mit $1, 2, \dots, N$ nummeriert denken. Nacheinander werden n Kugeln zufällig gezogen. Mit dieser Situation lassen sich auch andere Experimente beschreiben (Entnahme eines Objekts aus einer Gesamtheit von N verschiedenen Objekten).

I) Ziehung in Reihenfolge mit Zurücklegen.

Wird nach jeder Ziehung die gezogene Kugel zurückgelegt und wird neu gemischt, so erhält man nach n Ziehungen als Ergebnis ein n -Tupel $\omega = (x_1, \dots, x_n)$, wobei $x_i \in \{1, \dots, N\}$ die Nummer der i -ten gezogenen Kugel angibt. Hierbei soll es also auf die Reihenfolge der gezogenen Kugeln ankommen. Als Ergebnisraum Ω_I , also als Menge der möglichen Ergebnisse, erhält man:

$$\begin{aligned} \Omega_I &= \{ \omega = (x_1, \dots, x_n); x_i \in \{1, \dots, N\} \text{ für } i = 1, \dots, n \} \\ &= \{1, \dots, N\}^n = \{1, \dots, N\} \times \dots \times \{1, \dots, N\}. \end{aligned}$$

Offenbar gilt: $\text{card}(\Omega_I) = N^n$.

Beispiel: $\text{card}(\{1, \dots, 1000\}) = \text{card}(\{0, \dots, 999\}) = \text{card}(\{0, \dots, 9\}^3) = 10^3$. Allgemein gibt es 10^n n -stellige nicht-negative ganze Zahlen.

II) Ziehung in Reihenfolge ohne Zurücklegen.

Wird nach jeder Ziehung die gezogene Kugel **nicht** zurückgelegt, aber neu gemischt, so erhält man nach n Ziehungen als Ergebnis wieder ein n -Tupel $\omega = (x_1, \dots, x_n)$, wobei $x_i \in \{1, \dots, N\}$ die Nummer der i -ten gezogenen Kugel angibt. Hierbei soll es also wieder auf die Reihenfolge der gezogenen Kugeln ankommen. Als Ergebnisraum Ω_{II} erhält man jetzt:

$$\Omega_{II} = \{ \omega = (x_1, \dots, x_n); x_i \in \{1, \dots, N\}, x_i \neq x_j \text{ für } i \neq j, 1 \leq i, j \leq n \}.$$

Zur Bestimmung von $\text{card}(\Omega_{II})$ [und auch $\text{card}(\Omega_I)$] benutzen wir das folgende Abzählprinzip:

Abzählprinzip: Wir betrachten ein Gesamtexperiment, das aus n Einzelexperimenten besteht. Das Ergebnis ist dann ein n -Tupel $\omega = (x_1, \dots, x_n) \in \Omega$, wobei x_i das Ergebnis des i -ten Einzelexperimentes ist. Für das erste Einzelexperiment gebe es k_1 mögliche Ausgänge. Für jedes dieser Ausgänge gebe es beim zweiten Einzelexperiment dann k_2 mögliche Ausgänge. (Für $n=1$ gilt also $\text{card}(\Omega) = k_1$ und für $n=2$ $\text{card}(\Omega) = k_1 \cdot k_2$.) Für das i -te Einzelexperiment gebe es für jede fest gewählte "Vorgeschichte" (x_1, \dots, x_{i-1}) gerade k_i mögliche Ausgänge.

Dann ergibt sich leicht durch vollständige Induktion:

$$\text{card}(\Omega) = k_1 \cdot k_2 \cdot \dots \cdot k_n.$$

Bei der **Anwendung** zur Bestimmung von $\text{card}(\Omega_{II})$ haben wir für ω_1 zunächst N mögliche Ausgänge. Vor der i -ten Ziehung liegt (x_1, \dots, x_{i-1}) fest und es bleiben nur noch $N-(i-1)$ Kugeln zur Auswahl, also ist $k_i = N-i+1$. Es folgt also

$$\text{card}(\Omega_{II}) = N \cdot (N-1) \cdot \dots \cdot (N-n+1) = N! / (N-n)! = n! \cdot \binom{N}{n}, \quad 1 \leq n \leq N.$$

Im *Spezialfall* $N=n$ sind zum Schluß alle Kugeln gezogen und Ω_{II} besteht gerade aus der Menge der Permutationen von $\{1, \dots, N\}$, und man erhält das bekannte Ergebnis

$$\text{card}(\Omega_{II}) = N! \quad \text{für } N=n.$$

III) Ziehung ohne Reihenfolge ohne Zurücklegen.

Bei diesem Experiment sollen die Kugeln wieder **nicht** zurückgelegt werden; darüberhinaus soll nicht mehr notiert werden, in welcher Reihenfolge die Kugeln gezogen wurden, sondern nur noch die Menge der gezogenen Kugeln. Die gezogenen Kugeln sind bei einer Ziehung ohne Zurücklegen natürlich wieder alle verschieden. Wir haben also

$$\Omega_{III} = \{ \omega = \{x_1, \dots, x_n\} \subset \{1, \dots, N\}, x_i \neq x_j \text{ für } i \neq j, 1 \leq i, j \leq n \}.$$

Offenbar gilt dann:

$$\Omega_{III} = \{ \omega \subset \{1, \dots, N\}; \text{card}(\omega) = n \}.$$

Also ist $\text{card}(\Omega_{III})$ die **Anzahl der n -elementigen Teilmengen der N -elementigen Menge $\{1, \dots, N\}$.**

[Es ist ungewohnt, daß ω als Ergebnis (und nicht als Ereignis) eines Experiment eine Menge ist. Man kann dies umgehen, in dem man jedes $\omega \in \Omega_{III}$ durch ein geordnetes Tupel darstellt. Man wählt anstelle von $\omega = \{x_1, \dots, x_n\}$ dann die Darstellung $\omega = (x_1, \dots, x_n)$ mit $x_1 < x_2 < \dots < x_n$ und

$$\Omega_{III} = \{ \omega = (x_1, \dots, x_n); x_i \in \{1, \dots, N\} \forall i, 1 \leq x_1 < x_2 < \dots < x_n \leq N \}.$$

Zur Berechnung von $\text{card}(\Omega_{\text{III}})$ ist es zweckmäßig, eine Äquivalenzrelation in Ω_{II} einzuführen. Wir sagen $\omega = (x_1, \dots, x_n) \sim \omega' = (x'_1, \dots, x'_n)$, wenn ω' aus ω durch eine Permutation der Komponente hervorgeht. Jede Äquivalenzklasse ist dann durch die Menge $\{x_1, \dots, x_n\}$ festgelegt. Da jede Äquivalenzklasse $n!$ Elemente hat, gilt: $\text{card}(\Omega_{\text{II}}) = n! \cdot \text{card}(\Omega_{\text{III}})$ und damit:

$$(1.14) \quad \text{card}(\Omega_{\text{III}}) = \frac{N!}{n! (N-n)!} = \binom{N}{n} = \binom{N}{N-n}, \quad 1 \leq n \leq N.$$

Wir setzen wie üblich $\binom{N}{n} = 1$ für $n=0$ und $\binom{N}{n} = \frac{N \cdot (N-1) \cdot \dots \cdot 1 \cdot 0 \cdot \dots \cdot (N-n+1)}{n!} = 0$ für $n > N$.

Eine alternative Interpretation:

Wir fragen nach der Anzahl der Möglichkeiten, n (nicht unterscheidbare) Objekte auf N Plätze $\{1, \dots, N\}$ zu verteilen. Die Anzahl der belegten Plätze kann man nun mit der Menge der gezogenen Kugeln bei einer Ziehung ohne Reihenfolge ohne Zurücklegen identifizieren. Also ist die gesuchte Anzahl gerade $\binom{N}{n}$.

Die Ziehung ohne Reihenfolge mit Zurücklegen soll hier nicht behandelt werden.

1.3 Anwendungsbeispiele

Geburtstagszwillinge. Wie groß ist die W. p , daß mindestens zwei von n Studierenden in einer Vorlesung (oder von n rein zufällig ausgewählten Personen) am gleichen Tag Geburtstag haben. Als Ergebnisraum kann man gerade wählen:

$$\Omega_{\text{I}} = \{ \omega = (x_1, \dots, x_n); x_i \in \{1, \dots, N\} \text{ für } i = 1, \dots, n \} \quad \text{mit } N = 365.$$

Das Ergebnis $\omega = (x_1, \dots, x_n)$ bedeutet, daß der Studierende i am Tag x_i Geburtstag hat. Wie es oft vorkommt, ist es günstiger zu fragen, ob das interessierende Ereignis A nicht eintritt. Das Ereignis A^c , daß alle Studierenden an verschiedenen Tagen Geburtstag haben, wird dann gerade durch

$$A^c = \{ \omega = (x_1, \dots, x_n); x_i \in \{1, \dots, N\}, x_i \neq x_j \text{ für } i \neq j, 1 \leq i, j \leq n \} = \Omega_{\text{II}} \quad \text{mit } N = 365$$

beschrieben. Wählt man als mathematisches Modell nun den Laplaceschen W–raum (Ω_{I}, P) , so ist

$$P(A^c) = \text{card}(\Omega_{\text{II}}) / \text{card}(\Omega_{\text{I}}) = N \cdot (N-1) \cdot \dots \cdot (N-n+1) / N^n,$$

$$p = P(A) = 1 - \prod_{k=0}^{n-1} \left(1 - \frac{k}{N}\right).$$

Damit ergeben sich die folgenden Werte:

$\frac{n}{p}$	25	30	40	50
	0,44	0,71	0,89	0,97

. Man kann zeigen, daß sich die W. p noch erhöht, wenn man keine Gleichverteilung zu Grunde legt.

Zahlenlotto. Es werden $n=6$ Kugeln aus $N = 49$ Kugeln ohne Zurücklegen gezogen. Da die Reihenfolge der gezogenen Kugeln keine Rolle spielen soll, beschreiben wir das Experiment durch den Laplaceschen W–raum (Ω_{III}, P) mit

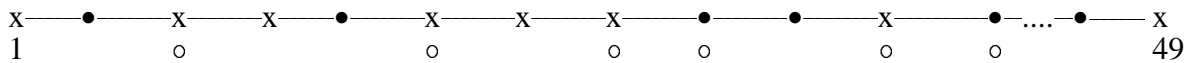
$$\Omega_{\text{III}} = \{ \omega \subset \{1, \dots, 49\}, |\omega| = 6 \}.$$

Dann ist die W. für jedes Ergebnis $\omega = \{x_1, \dots, x_6\}$ gleich

$$P(\omega) = 1 / \binom{49}{6} = 7,1511 \cdot 10^{-8} =: p_6.$$

Dies ist gerade die W. für "6 Richtige".

Es soll nun die W. p_4 für "genau 4 Richtige" ermittelt werden bei einem Tip $\bar{\omega} = \{\bar{x}_1, \dots, \bar{x}_6\}$.



• $\hat{=}$ getippte Zahlen, $\circ \hat{=}$ gezogene Zahlen

Dann läßt sich das Ereignis $A_4 \hat{=}$ "4 Richtige" beschreiben durch

$$A_4 = \{\omega \in \Omega_{\text{III}}; \text{card}(\bar{\omega} \cap \omega) = 4\}.$$

Die Anzahl der Elemente von A_4 soll mit dem Abzählprinzip ermittelt werden (mit $n=2$ Einzelexperimenten). Wir können eine zu A_4 gehörende Menge ω von Kugeln dadurch beschreiben, daß wir zunächst 4 Kugeln aus $\bar{\omega}$ (erstes Einzelexperiment) und dann zwei Kugeln aus $\{1, \dots, 49\} \setminus \bar{\omega}$ festlegen (zweites Einzelexperiment). Für das erste Einzelexperiment gibt es

$\binom{6}{4}$ Möglichkeiten und für das zweite gerade $\binom{43}{2}$ Möglichkeiten. Somit erhalten wir nun $\text{card}(A_4) = \binom{6}{4} \cdot \binom{43}{2}$; damit gilt

$$p_4 = P(A_4) = \binom{6}{4} \cdot \binom{43}{2} / \binom{49}{6}.$$

1.4 Die hypergeometrische Verteilung.

Das Beispiel vom Zahlenlotto läßt sich verallgemeinern. Wir gehen wieder von einem Tip $\bar{\omega} = \{\bar{x}_1, \dots, \bar{x}_6\}$ aus und denken uns diese Kugeln schwarz angemalt, während die übrigen weiß sind.

Dann haben wir danach gefragt, daß genau s schwarze Kugeln gezogen werden mit $s = 6$ und 4.

Allgemein betrachten wir jetzt die Situation, daß wir S schwarze und W weiße Kugeln, also insgesamt $N = S + W$ Kugeln in einer Urne haben. Es werden nun $n \leq N$ Kugeln **ohne Zurücklegen** ohne Reihenfolge gezogen. Dann soll gezeigt werden: Die W. dafür, daß genau s schwarze und $n-s = w$ weiße Kugeln gezogen werden, ist gleich

$$h(s; n, N, S) = \binom{S}{s} \cdot \binom{W}{w} / \binom{N}{n}, \quad 0 \leq s \leq n, \quad n-s = w, \quad N = S + W, \quad n \leq N.$$

Die Funktion $s \mapsto h(s; n, N, S)$ ist bei festen Parametern n, N, S offenbar eine W -funktion auf $\mathfrak{X} := \{0, \dots, n\}$. Sie ist positiv genau dann, wenn gilt $s \leq S$, $w \leq W$, d.h. $(n-N+S)^+ \leq s \leq n \wedge S$, wegen

$$\binom{S}{s} = \frac{S \cdot (S-1) \cdot \dots \cdot (S-s+1)}{s!} = 0 \quad \text{für } s \geq S+1.$$

Wenn man bei gegebenem n für alle Parameter N, S den gleichen Raum $\mathfrak{X} = \{0, \dots, n\} \ni s$ haben möchte, ist es günstig, solche Werte s mit verschwindender W -funktion zuzulassen. Das W -maß auf $\{0, \dots, n\}$ mit der W -funktion $s \mapsto h(s; n, N, S)$ heißt **hypergeometrische Verteilung**.

Zum Beweis denken wir uns die Kugeln wieder numeriert mit $1, \dots, N$; dabei stehe etwa $\Gamma \subset \{1, \dots, N\}$, z.B. $\Gamma = \{1, \dots, S\}$, für die Menge der schwarzen Kugeln, also $\text{card}(\Gamma) = S$.

Wir arbeiten jetzt wieder mit dem Raum

$$\Omega_{\text{III}} = \{ \omega = \{x_1, \dots, x_n\} \subset \{1, \dots, N\}, (\text{o.E.}) x_1 < x_2 < \dots < x_n \}, \text{ also } \text{card}(\Omega_{\text{III}}) = \binom{N}{n},$$

und der Gleichverteilung P auf Ω_{III} .

Es sei $X(\omega) := \text{card}(\omega \cap \Gamma)$ die Anzahl der gezogenen schwarzen Kugeln. Die Anzahl der Möglichkeiten, genau s schwarze und $n-s = w$ weiße Kugeln zu ziehen, ist also $\text{card}(\{X = s\})$.

Sie soll mit dem Abzählprinzip ermittelt werden (mit 2 Einzelexperimenten).

Es gibt $\binom{S}{s}$ Möglichkeiten, s schwarze Kugeln aus S vorhandenen schwarzen Kugeln ohne Zurücklegen zu ziehen (erstes Einzelexperiment). Für jede Wahl von den s schwarzen Kugeln gibt

es $\binom{W}{w}$ Möglichkeiten, w weiße Kugeln aus den verbleibenden Kugeln zu ziehen (zweites Einzelexperiment). Somit gibt es gerade $\binom{S}{s} \cdot \binom{W}{w}$ Möglichkeiten, genau s schwarze und $n-s = w$ weiße Kugeln zu ziehen. Also gilt mit $\{X = s\} := \{ \omega \in \Omega_{\text{III}}; X(\omega) = s \}$ wie in § 0.2

$$P'(s) = P(\{X = s\}) = h(s; n, N, S) \quad \text{und} \quad X : \Omega_{\text{III}} \mapsto \mathfrak{X}.$$

[Mit dem Raum (Ω_{III}, P) haben wir also einen einfacheren Ergebnisraum \mathfrak{X} , der nur einen Teil der Informationen enthält und auf dem die W -Funktion $P'(s) := h(s; n, N, S)$ erklärt ist.

Diesen Sachverhalt werden wir im Zusammenhang mit dem Begriff Zufallsvariable und der Verteilung genauer studieren.]

Anwendung (Qualitätskontrolle). Wir stellen uns eine Lieferung von N gleichartigen Objekten vor, von denen S defekt und damit $N-S$ intakt sind. Als Stichprobe greift man n Objekte heraus und überprüft diese. Dann ist $h(s; n, N, S)$ die W. dafür, daß s der überprüften Objekte defekt sind.

In der **Statistik** faßt man S als unbekanntem Parameter auf und versucht von der Stichprobe s Rückschlüsse auf S zu ziehen.

Beispiel 1.2 (Skatenspiel). Beim Skat erhält jeder der drei Spieler zehn Karten aus 32 gemischten Karten, und zwei Karten (der Skat) werden zunächst verdeckt beiseite gelegt. Es gibt 4 Buben. Wie groß ist die W. für das Ereignis A_I , daß ein Spieler I genau 3 Buben erhält? Wir haben die obige Situation vorliegen mit $S = 4$, $W = 28$, $n = 10$ und $s = 3$. Die Buben entsprechen den schwarzen Kugeln. Die gesuchte W. ist also: $\binom{4}{3} \binom{28}{7} / \binom{32}{10} = 66 / 899$.

Die W. (in einem geeigneten W -Raum vgl. Übung) für das Ereignis $A_I \cup A_{\text{II}} \cup A_{\text{III}}$, daß mindestens einer der drei Spieler 3 Buben erhält, ist dreimal so groß, da die Ereignisse $A_I, A_{\text{II}}, A_{\text{III}}$, daß Spieler I, II bzw. III drei Buben erhält, unvereinbar (disjunkt) sind.

1.5 Vereinigung von Ereignissen

Nach der Formel (1.9) gilt:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Diese Formel läßt sich verallgemeinern für die Vereinigung von n Ereignissen $P(A_1 \cup \dots \cup A_n)$. Dies soll hier aber nicht geschehen. Verwendet man die mengentheoretische Differenz, so gilt:

$$A \cup B = A \cup (B \setminus A).$$

Dabei handelt es sich auf der linken Seite um eine disjunkte Vereinigung und hat also:

$$P(A \cup B) = P(A) + P(B \setminus A).$$

Allgemein kann man auf diese Weise aus einer beliebigen Vereinigung eine disjunkte Vereinigung machen. Für drei Mengen gilt z.B.

$$A \cup B \cup C = A \cup B \setminus A \cup C \setminus (A \cup B) = A \cup B \cap A^c \cup C \cap A^c \cap B^c \text{ und}$$

$$P(A \cup B \cup C) = P(A) + P(B \setminus A) + P(C \setminus (A \cup B)) = P(A) + P(B \cap A^c) + P(C \cap A^c \cap B^c).$$

Nach den Regeln von de Morgan ergibt sich auch die folgende nützliche Formel:

$$P(A_1 \cup \dots \cup A_n) = 1 - P((A_1 \cup \dots \cup A_n)^c) = 1 - P(A_1^c \cap \dots \cap A_n^c).$$

Bei Vorliegen einer disjunkten Zerlegung von Ω läßt sich oft die folgende **Formel von der totalen Wahrscheinlichkeit** (einfache Form) anwenden:

$$(1.23) \quad P(A) = \sum_i P(A \cap B_i) \\ \text{falls } \Omega = \bigcup_i B_i, \{B_i\} \text{ paarweise disjunkt; wegen } A = \bigcup_i A \cap B_i.$$

Als **Beispiel** betrachten wir das Experiment "zweimaliges Würfeln" aus § 0, das wir jetzt durch den Laplaceschen W-Raum (Ω, P) beschreiben mit $\Omega = \{1, \dots, 6\}^2$. Es sei wieder X_m das Ergebnis des m -ten Wurfes. Dann kann

$A = \{X_1 + X_2 = j\}$, $B_i = \{X_1 = i\}$ (mit fest gewähltem j und variablem i): gewählt werden.

1.6 Multinomialkoeffizienten

Der Binomialkoeffizient $\binom{n}{k}$ beschreibt, wieviele Möglichkeiten es gibt, k Kugeln aus n unterscheidbaren Kugeln auszuwählen. Wir können auch sagen, $\binom{n}{k}$ beschreibt, wieviele Möglichkeiten es gibt, n unterscheidbare / numerierte Kugeln in zwei Gruppen aufzuteilen (in die gezogenen und die nicht gezogenen), sodaß die erste Gruppe k Elemente hat und die zweite dann $n-k$ Element. Diese Fragestellung läßt sich nun verallgemeinern.

Wir fragen: Wieviele Möglichkeiten gibt es, die Zahlen $1, \dots, n$ in r Gruppen aufzuteilen, sodaß die erste Gruppe k_1 Elemente hat, die zweite k_2 und die r -te Gruppe k_r Elemente hat mit $n = k_1 + \dots + k_r$. Dabei kommt es auf die Reihenfolge der Gruppen an; aber innerhalb der Gruppen soll die Reihenfolge keine Rolle spielen.

Wir verwenden wieder das Abzählprinzip mit r Einzelexperimenten. In ersten Einzelexperiment ordnen wir k_1 Zahlen der ersten Gruppe zu; dafür gibt es $\binom{n}{k_1}$ Möglichkeiten. In zweiten Einzelexperiment ordnen wir zu jeder Wahl dieser k_1 Zahlen die verbleibenden $n-k_1$ Zahlen der zweiten Gruppe zu; dafür gibt es $\binom{n-k_1}{k_2}$ Möglichkeiten. Setzt man das Verfahren fort, so bleiben im r -ten Einzelexperiment für die r -te Gruppe nur noch k_r Kugeln übrig.

Insgesamt ergibt sich also für die Anzahl der Möglichkeiten:

$$\binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \dots \cdot \binom{n-k_1-\dots-k_{r-2}}{k_{r-1}} = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!} \quad \text{wegen } (n-k_1-\dots-k_{r-2}) - k_{r-1} = k_r.$$

Wir setzen wie üblich $\left[\begin{matrix} n \\ k_1, k_2, \dots, k_r \end{matrix} \right] := \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!}$ und sprechen vom

Multinomialkoeffizienten. Für $r=2$ ist dann also $\binom{n}{k} = \left[\begin{matrix} n \\ k, n-k \end{matrix} \right]$.

Beispiel 1.2 (*Skatspiel, Fortsetzung*). Beim Skat erhält jeder der drei Spieler zehn Karten aus 32 gemischten Karten, und zwei Karten (der Skat) werden zunächst verdeckt beiseite gelegt. Wir fragen jetzt nach der Anzahl der Möglichkeiten, 32 Karten auf drei Spieler und auf den Skat aufzuteilen. Dann haben wir also vier Gruppen; dabei sollen drei Gruppen je zehn Karten erhalten ($k_1 = k_2 = k_3 = 10$) und die vierte die restlichen zwei ($k_4 = 2$). Die Anzahl der möglichen

Aufteilungen ist also $\left[\begin{matrix} 32 \\ 10, 10, 10, 2 \end{matrix} \right] = \frac{32!}{10! \cdot 10! \cdot 10! \cdot 2!} \cdot$

§2 Bedingte Wahrscheinlichkeit und Unabhängigkeit

2.1 Definition und Eigenschaften bedingter Wahrscheinlichkeiten

Als **Beispiel** wollen wir zunächst das folgende Zufallsexperiment betrachten: "Beobachtung des Lebensalters (von Geburt bis Tod) einer rein zufällig ausgewählten (männlichen) Person". Wir wählen $\Omega = \{0,1,\dots,N\}$ mit $N > 100$ und stellen uns vor, daß das W–maß P auf Grund einer statistischen Schätzung gewonnen wurde, wie das auch bei *Sterbetafeln* der Fall ist. Ist (x_1, \dots, x_n) die zugrundeliegende Stichprobe und setzt man $k_n(\omega) = \text{card}(\{x_i; 1 \leq i \leq n, x_i = \omega\})$, so hat man:

$$P(\omega) = h_n(\omega) = k_n(\omega) / n.$$

Es interessiert das Ereignis, daß die Person älter als 65 wird, also $A = \{66, \dots, N\}$. Man hat:

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega > 65} \frac{1}{n} k_n(\omega) \quad (\text{vgl. (1.10)}).$$

Die Situation ändert sich, wenn bekannt ist, daß die Person 30 Jahre alt ist (etwa bei Abschluß einer Versicherung). Jetzt wird man bei der Schätzung nur noch die x_i benutzen mit $x_i \geq 30$. Setzt man $B = \{30, \dots, N\}$, so interessiert also:

$$\text{card}(\{x_i; 1 \leq i \leq n, x_i \in A \cap B\}) / \text{card}(\{x_i; 1 \leq i \leq n, x_i \in B\}) = h_n(A \cap B) / h_n(B) = P(A \cap B) / P(B),$$

$$\text{mit } P(A \cap B) / P(B) = \sum_{\omega \geq 66} P(\omega) / \sum_{\omega \geq 30} P(\omega) =: P(A | B).$$

Somit ist hier also $P(A | B) > P(A) = \sum_{\omega \geq 66} P(\omega)$ wegen $A \subset B$.

Wir nehmen diese Überlegungen als Vorbild für die folgende **Definition**: Für einen beliebigen W–raum (Ω, P) und für beliebige Ereignisse B mit $P(B) > 0$ heißt

$$(2.2) \quad P(A | B) := P(A \cap B) / P(B)$$

die **bedingte Wahrscheinlichkeit von A bei gegebenem B**.

Beispiel. (*zweimaliges Werfen einer Münze*). Es ist $\Omega = \{W, Z\}^2$, wobei W für Wappen und Z für Zahl steht. Es ist sinnvoll mit den Laplaceschen W–raum (Ω, P) zu arbeiten. Wir interessieren uns für die folgenden Ereignisse:

$$A_m \hat{=} \text{"der } m\text{-te Wurf ergibt Wappen"}, m=1,2, \text{ also } A_1 = \{(W,W), (W,Z)\};$$

$$A_1 \cap A_2 = \{(W,W)\} \hat{=} \text{"beide Würfe ergeben jeweils Wappen"},$$

$$A_1 \cup A_2 = \{(W,W), (W,Z), (Z,W)\} \hat{=} \text{"mindestens ein Wurf ergibt Wappen"}.$$

$$\text{Dann ist } P(A_1) = P(A_2) = 1/2, P(A_1 \cap A_2) = 1/4,$$

$$P(A_1 \cup A_2) = 1 - P(A_1^c \cap A_2^c) = 1 - P((Z,Z)) = 1 - 1/4 = 3/4,$$

$$A_1 \cap A_2 \subset A_1 \subset A_1 \cup A_2.$$

Also erhalten wir: $P(A_1 \cap A_2 | A_1) = 1/2$ mit $P(A_1 \cap A_2 | A_1) = P(A_2 | A_1)$, also

$$P(A_2 | A_1) = P(A_2) \quad \text{und} \quad P(A_1 \cap A_2 | A_1 \cup A_2) = 1/3. \quad \square$$

Satz 2.2 (i) Sei $P(B) > 0$, dann wird durch $\tilde{P}(A) := P(A|B)$ wieder ein W -maß \tilde{P} auf Ω definiert

mit $P(B|B) = 1$. Ist $A \subset B^c$, so ist $P(A|B) = 0$. Ist $A \subset B$, so ist $P(A|B) = P(A)/P(B) \geq P(A)$.

(iii) (**Formel von Bayes**, einfache Form) Sei $P(A), P(B) > 0$, dann gilt: $P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}$.

Beweis: Offenbar ist $\tilde{P}(\Omega) = \tilde{P}(B) = 1$, $\tilde{P}(A) \geq 0$. Zum Nachweis der Additivität seien A und A' disjunkt; dann ist $P((A \cup A') \cap B) = P((A \cap B) \cup (A' \cap B)) = P(A \cap B) + P(A' \cap B)$.

Die Formel (iii) ist klar. \square

2.2 Unabhängigkeit

Im obigen Beispiel des zweimaligen Werfens einer Münze hatten wir für die Ereignisse $A_m \hat{=} \text{"der } m\text{-te Wurf zeigt Wappen"}$, $m=1,2$: $P(A_2|A_1) = P(A_2)$. Dies können wir so interpretieren, daß das Ergebnis des ersten Wurfes keinen Einfluß auf das Ergebnis des zweiten Wurfes hat. Die Eigenschaft ist offenbar äquivalent zu $P(A_2 \cap A_1) = P(A_2) \cdot P(A_1)$.

Die letzte Eigenschaft ist auch definiert, wenn $P(A_1) = 0$ ist, und ist zudem noch symmetrisch in A_1 und A_2 . Wir nehmen diese deswegen für die folgende **Definition**:

*Zwei Ereignisse A, B heißen **unabhängig**, wenn gilt: $P(A \cap B) = P(A) \cdot P(B)$.*

Zur **Interpretation** betrachten wir das Problem: Wie hängen die Schulnoten in Mathematik und Deutsch eines zufällig herausgegriffenen Schülers voneinander ab ?

Ein Untersuchung bei n Schülern ergibt als Stichprobe ein Tupel

$$x_1, \dots, x_n \in \{1, \dots, 6\} \times \{1, \dots, 6\}.$$

Das Ereignis "mindestens eine Zwei in Mathematik" beschreiben wir durch $A = \{1, 2\} \times \{1, \dots, 6\}$ und das Ereignis "mindestens eine Zwei in Deutsch" beschreiben wir durch $B = \{1, \dots, 6\} \times \{1, 2\}$.

Dann ist wieder

$$\text{card}(\{x_i; 1 \leq i \leq n, x_i \in A \cap B\}) / \text{card}(\{x_i; 1 \leq i \leq n, x_i \in B\}) = h_n(A \cap B) / h_n(B)$$

ein Maß für $P(A \cap B)/P(B)$. Bei Unabhängigkeit würde man also erwarten, daß das Verhältnis der Anzahl der Schüler, die sowohl in Mathematik als auch in Deutsch mindestens ein Zwei haben, zu der Anzahl der Schüler, die in Deutsch mindestens ein Zwei haben, gleich dem Verhältnis ist der Anzahl der Schüler, die in Mathematik mindestens ein Zwei haben, zu der Gesamtzahl n der Schüler ist. Die Beschränkung auf Schüler mit einer Deutschnote von mindestens Zwei hat also keinen Einfluß auf das Verhältnis. Dies wird man aber wohl hier nicht erwarten.

Dies wäre anders, wenn die beiden Schulnoten durch zweimaliges Würfeln ermittelt würden. Dann würde man eine solche Unabhängigkeit sicherlich erwarten.

Dieser Sachverhalt gilt auch allgemein. In den Anwendungen wird Ω oft eine geringere Rolle spielen, vielmehr interessieren der Bildraum \mathfrak{X} und die W-Funktion von X .

Definition. In der Situation 3.1 versteht man unter der **W-Funktion von X** die W-Funktion $P'(x)$ auf \mathfrak{X} mit

$$(3.0a) \quad P'(x) := P(\{X = x\}) = \sum_{\omega \in \{X=x\}} P(\omega) = \sum_{\omega \in \Omega, X(\omega) = x} P(\omega) =: P(X=x), \quad x \in \mathfrak{X}.$$

Unter der **Verteilung von X** versteht man das W-Maß

$$P'(A) := P(\{X \in A\}) = \sum_{\omega \in \{X \in A\}} P(\omega) =: P(X \in A), \quad A \subset \mathfrak{X}.$$

Notation. Will man die Abhängigkeit der W-Funktion von X notieren, so schreibt man auch $P' = P_X$. Beispiele für die obigen Kurzschreibweisen sind z.B. $P(X = 3)$, $P(X \geq 0)$. Sind mehrere Zva auf (Ω, P) gegeben, etwa X und Y , so verwendet man die Notation $\{X \in A, Y \in B\} = \{X \in A\} \cap \{Y \in B\}$ sowie $P(X \in A, Y \in B)$ in naheliegender Weise [z.B. $P(X \geq 0, Y = 3)$].

Bemerkung. $P'(A)$ ist in der Tat ein W-Maß auf \mathfrak{X} mit der W-Funktion $P'(x)$. Denn es gilt wie im Beispiel $P'(x) \geq 0$, $\sum_{x \in \mathfrak{X}} P'(x) = 1$, sowie wie in (1.10)

$$P'(A) = \sum_{x \in A} P'(x) \quad \text{für } A \subset \mathfrak{X}, \text{ d.h.}$$

$$(3.0b) \quad P(X \in A) = \sum_{x \in A} P(X = x) \quad \text{für } A \subset \mathfrak{X}.$$

wegen $P(X \in A) = P(\bigcup_{x \in A} \{X = x\}) = \sum_{x \in A} P(X = x)$. \square

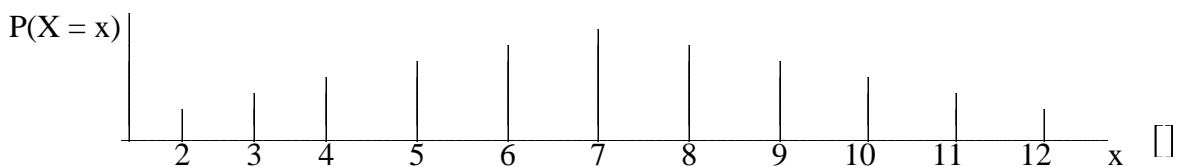
Beispiel (zweimaliges Werfen eines Würfels, Fortsetzung).

Die W-funktion der Summe X von zwei Würfeln ist gegeben durch

$$P(X=x) = \sum_{(i,k) \in \Omega, i+k=x} P((i,k)) = \frac{1}{36} \text{card}(\{(i,k); i+k=x; 1 \leq i \leq 6, x-6 \leq i \leq x-1\}), \text{ also:}$$

x	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Als sogenanntes Stabdiagramm für $P(X=x)$ erhält man:



Dabei hatten wir die **Formel von der totalen W.** benutzt. Diese liefert jetzt für Zva X, Y :

$$(3.0c) \quad P(B) = \sum_{x \in \mathfrak{X}} P(B \cap \{X=x\}), \quad P(Y=y) = \sum_{x \in \mathfrak{X}} P(Y=y, X=x).$$

Beispiel (*hypergeometrische Verteilung*, vgl. §1.4). Eine Urne enthalte S schwarze und $N-S$ weiße Kugeln. Es werden n Kugeln ohne Zurücklegen gezogen. Sei $1 \leq S < N$, und $n \leq N$,

$$\Omega = \{ \omega \subset \{1, \dots, S, S+1, \dots, N\}; \text{card}(\omega) = n \} = \Omega_{\text{III}};$$

hier sind die ω also (n elementige) Teilmengen von $\{1, \dots, N\}$. Wir haben $\text{card}(\Omega) = \binom{N}{n}$ und wählen P als Gleichverteilung auf Ω .

Dann haben wir uns für die Zva X interessiert, die die Anzahl der gezogenen schwarzen Kugeln (d.h. die Kugeln mit Nummer $\leq S$) beschreibt, also für

$$X(\omega) = \text{card}(\omega \cap \{1, \dots, S\}); \quad X : \Omega \longmapsto \mathfrak{X} = \{0, \dots, n\}.$$

In §1.4 hatten wir die W -funktion von X berechnet zu

$$P(X=x) = h(x; n, N, S) = \binom{S}{x} \cdot \binom{W}{n-x} / \binom{N}{n}, \quad 0 \leq x \leq n, \quad n-x = w, \quad N = S + W, \quad n \leq N.$$

Die Verteilung von X ist hier also die hypergeometrische Verteilung! Man sagt auch:

X hat eine hypergeometrische Verteilung. \square

Beispiel 3.2 (*2-maliges Werfen eines Würfels*). Es ist $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$ und P die Gleichverteilung, d.h. (Ω, P) ein Laplacescher W -raum.

Dann beschreibe die Zva X_m wieder das Resultat des m -ten Wurfes ($m=1,2$), also z.B.

$X_1(\omega) = X_1((i,k)) = i$. Dann gilt:

$$P(X_1=i, X_2=k) = P((i,k)) = 1/6^2 \quad \text{und}$$

$$P(X_1=i) = \sum_{k=1}^6 P(X_1=i, X_2=k) = 6/6^2 = 1/6 = P(X_2=k),$$

also haben wir:

- (i) die Verteilung von X_m ist die Gleichverteilung auf $\mathfrak{X} = \{1, \dots, 6\}$;
- (ii) $P(X_1=i, X_2=k) = P(X_1=i) \cdot P(X_2=k)$,
also sind die Ereignisse $\{X_1=i\}$ und $\{X_2=k\}$ unabhängig $\forall i, k$.

Man sagt auch: X_1 und X_2 sind unabhängig und haben eine Gleichverteilung. Insbesondere sind X_1 und X_2 auch **identisch verteilt**, d.h. die Verteilung von X_1 ist gleich der Verteilung von X_2 . \square

Beispiel. Eine Urne enthalte S schwarze und $W := N - S$ weiße Kugeln, $N \geq 2$. Es werden $n=2$ Kugeln ohne Zurücklegen gezogen. Dieses Mal soll die **Reihenfolge notiert** werden.

Dann wählen wir

$$\Omega = \{(i,k); i,k \in \{1,\dots,N\}, i \neq k\}.$$

Dann gilt $\text{card}(\Omega) = N \cdot (N-1)$. P sei die Gleichverteilung auf Ω .

Die Zva X_m ($m = 1, 2$) sollen angeben, ob bei der m -ten Ziehung eine schwarze Kugel gezogen worden ist ($X_m=1$) oder nicht ($X_m=0$).

$$\left[\text{Also } X_1((i,k)) = \begin{cases} 1 & \text{für } i \leq S \\ 0 & \text{für } i > S \end{cases}, \quad X_2((i,k)) = \begin{cases} 1 & \text{für } k \leq S \\ 0 & \text{für } k > S \end{cases} \right]$$

Dann gilt nach dem Abzählprinzip (vgl. Übung):

$$(a) \quad P(X_1 = x, X_2 = y) = P(X_2 = x, X_1 = y);$$

$$P(X_1 = x) = P(X_2 = x)$$

(also sind die W-Funktionen und die Verteilungen von X_1 und X_2 identisch; man sagt:

X_1 und X_2 sind identisch verteilt);

$$P(X_1 = x, X_2 = y) \neq P(X_1 = x) \cdot P(X_2 = y) \text{ (man sagt: } X_1 \text{ und } X_2 \text{ sind abhängig)}$$

Aus (a) folgt wie man erwartet (vgl. Übung):

$$(b) \quad P(X_2 = 1 | X_1 = 0) = \frac{S}{N-1}, \quad P(X_2 = 1 | X_1 = 1) = \frac{S-1}{N-1},$$

$$P(X_1 = 1 | X_2 = 0) = \frac{S}{N-1}, \quad P(X_1 = 1 | X_2 = 1) = \frac{S-1}{N-1}.$$

3.2 Unabhängigkeit

Wir sagen zwei Zufallsvariablen X, Y auf einem Ω -Raum (Ω, P) mit Werten in \mathfrak{X} bzw. \mathfrak{Y} sind unabhängig, wenn je zwei Ereignisse $\{X=x\}, \{Y=y\}$ unabhängig sind, d.h. wenn gilt:

$$P(X=x, Y=y) = P(X=x) \cdot P(Y=y) \quad \text{für } x \in \mathfrak{X}, y \in \mathfrak{Y}.$$

Definition 3.3. Die Zva X_1, \dots, X_n auf (Ω, P) mit Werten in $\mathfrak{X}_1, \dots, \mathfrak{X}_n$ heißen **unabhängig**, falls:

$$(3.2) \quad P(X_1=x_1, \dots, X_n=x_n) = P(X_1=x_1) \cdot \dots \cdot P(X_n=x_n) \quad \forall x_m \in \mathfrak{X}_m, 1 \leq m \leq n.$$

Das Beispiel 3.2 zeigt, daß wir mit den Ergebnissen X_m beim m -ten Würfeln schon vorher unabhängige Zva gesehen haben.

Lemma. Seien X_1, \dots, X_n unabhängige Zva auf (Ω, P) mit Werten in $\mathfrak{X}_1, \dots, \mathfrak{X}_n$. Dann gilt:

$$(3.3) \quad P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdot \dots \cdot P(X_n \in A_n) \quad \forall A_m \subset \mathfrak{X}_m, 1 \leq m \leq n.$$

Offenbar enthält die Eigenschaft (3.3) die Eigenschaft (3.2) als Spezialfall.

Beweis: Wir wollen die Formel (3.0b) " $P(X \in A) = \sum_{x \in A} P(X=x)$ für $A \subset \mathfrak{X}$ " verwenden für die Zva $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ auf Ω mit Werten in $\mathfrak{X}_1 \times \dots \times \mathfrak{X}_n$.

Sind insbesondere die X_m reelle Zva, so ist X ein (n -dimensionaler) Zufallsvektor.

Für $A \subset \mathfrak{X}_1 \times \dots \times \mathfrak{X}_n$ und $A = A_1 \times \dots \times A_n$ ergibt sich aus (3.0b):

$$(3.0d) \quad P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) = \sum_{(x_1, \dots, x_n) \in A_1 \times \dots \times A_n} P((X_1, \dots, X_n) = (x_1, \dots, x_n))$$

$$(3.0e) \quad P((X_1, \dots, X_n) \in A) = \sum_{(x_1, \dots, x_n) \in A} P((X_1, \dots, X_n) = (x_1, \dots, x_n)).$$

Der Ausdruck (3.0d) läßt sich auch schreiben als :

$$\begin{aligned} P(X_1 \in A_1, \dots, X_n \in A_n) &= \\ &= \sum_{x_1 \in A_1, \dots, \sum_{x_n \in A_n} P((X_1, \dots, X_n) = (x_1, \dots, x_n)) \\ &= \sum_{x_1 \in A_1, \dots, \sum_{x_n \in A_n} P(X_1=x_1) \cdot \dots \cdot P(X_n=x_n) \quad [\text{gemäß (3.2)}] \\ &= \sum_{x_1 \in A_1} P(X_1=x_1) \sum_{x_2 \in A_2} P(X_2=x_2) \dots \sum_{x_n \in A_n} P(X_n=x_n) \\ &= P(X_1 \in A_1) \cdot \dots \cdot P(X_n \in A_n) \quad (\text{gemäß (3.0b)}. \quad \square \end{aligned}$$

Sind die Zva X_1, \dots, X_n unabhängig, so sind auch X_1, \dots, X_d unabhängig für $d \leq n$. Dies folgt aus dem folgenden Satz mit $I = \{1, \dots, d\}$.

Satz 3.4. Seien X_1, \dots, X_n unabhängige Zva auf (Ω, P) mit Werten in $\mathfrak{X}_1, \dots, \mathfrak{X}_n$.

Dann gilt für alle $I \subset \{1, \dots, n\}$:

$$(3.3)_I \quad P\left[\bigcap_{m \in I} \{X_m = x_m\}\right] = \prod_{m \in I} P(X_m = x_m) \quad \forall x_m \in \mathfrak{X}_m, 1 \leq m \leq n.$$

Beweis. Nach dem Lemma gilt (3.3)_I mit $A_m = \{x_m\}$ für $m \in I$ und $A_m = \mathfrak{X}_m$ für $m \notin I$, wegen $\{X_m \in \mathfrak{X}_m\} = \Omega$, $P(X_m \in \mathfrak{X}_m) = 1$. \square

Wir wollen jetzt zeigen, daß Funktionen von unabhängigen Zva wieder unabhängig sind.

Wie üblich sei $Y := f \circ X = f(X)$ die Zva mit $Y(\omega) = f \circ X(\omega) := f(X(\omega))$, $\Omega \xrightarrow{X} \mathfrak{X} \xrightarrow{f} \mathfrak{Y}$.

Satz 3.5. Seien X_1, \dots, X_n unabhängige Zva auf (Ω, P) mit Werten in $\mathfrak{X}_1, \dots, \mathfrak{X}_n$.

- (a) Es sind $Y_m = f_m \circ X_m$, $1 \leq m \leq n$, unabhängige Zva, wobei f_m Funktionen auf \mathfrak{X}_m sind.
- (b) Es sind $f(X_1, \dots, X_d)$ und $g(X_{d+1}, \dots, X_n)$ unabhängige Zva für $1 \leq d < n$, wobei f Funktion auf $\mathfrak{X}_1 \times \dots \times \mathfrak{X}_d$ und g Funktion auf $\mathfrak{X}_{d+1} \times \dots \times \mathfrak{X}_n$.

Beispiel: In der Situation von Satz 3.5(b) mit $d = 1$ sind also etwa $X_1 \left[=: f(X_1) \right]$ und $X_2 + X_3 \left[=: g(X_2, X_3) \right]$ unabhängig. \square

Beweis von Satz 3.5. (a) Es ist $\{f_m \circ X_m = y_m\} = \{X_m \in A_m\}$ mit

$A_m = \{x_m \in \mathfrak{X}_m; f(x_m) = y_m\} = f_m^{-1}(\{y_m\})$. Gemäß (3.3) kann man nun faktorisieren

(b) Der Beweis soll hier nur für den Spezialfall des obigen Beispiels geführt werden.

Es ist $\{X_2 + X_3 = y\} = \{(X_2, X_3) \in A\} = \bigcup_{(x_2, x_3) \in A} \{X_2 = x_2, X_3 = x_3\}$ mit

$A := \{(x_2, x_3) \in \mathfrak{X}_2 \times \mathfrak{X}_3; x_2 + x_3 = y\}$; damit gilt offenbar:

$$P(X_1 = x, X_2 + X_3 = y) = \sum_{(x_2, x_3) \in A} P(X_1 = x, X_2 = x_2, X_3 = x_3)$$

$$= \sum_{(x_2, x_3) \in A} P(X_1 = x) \cdot P(X_2 = x_2) \cdot P(X_3 = x_3)$$

$$= P(X_1 = x) \cdot \sum_{(x_2, x_3) \in A} P(X_2 = x_2, X_3 = x_3).$$

Nun ist mit (3.0d):

$$\sum_{(x_2, x_3) \in A} P(X_2 = x_2, X_3 = x_3) = P((X_2, X_3) \in A) = P(X_2 + X_3 = y). \quad \square$$

2.3 Produktexperimente

Wir betrachten die folgende **typische Situation**:

Es sind $(\mathfrak{X}_1, P_1), \dots, (\mathfrak{X}_n, P_n)$ W-Räume, die n sich nicht beeinflussende Einzelexperimente beschreiben.

Unser **Ziel** ist es nun, diese Einzelexperimente durch unabhängige Zva X_1, \dots, X_n zu beschreiben mit Werten in $\mathfrak{X}_1, \dots, \mathfrak{X}_n$ und W-Funktionen oder Verteilungen P_1, \dots, P_n .

Als Beispiel können wir an das n-malige Werfen eines Würfels, einer Münze oder einer Reißzwecke denken.

Das Ziel soll folgendes bedeuten:

Es existieren ein W-Raum (Ω, P) und Zva X_1, \dots, X_n auf (Ω, P) mit

$$\begin{aligned} P(X_m = x_m) &= P_m(x_m); \\ P(X_1 = x_1, \dots, X_n = x_n) &= P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n), \quad \forall x_m \in \mathfrak{X}_m, 1 \leq m \leq n. \end{aligned}$$

Bei der Konstruktion wählt man

$$(2.6) \quad \begin{aligned} \Omega &= \mathfrak{X}_1 \times \dots \times \mathfrak{X}_n; \\ X_m : \Omega &\rightarrow \mathfrak{X}_m \text{ als Projektion, d.h. } X_m(\omega) = X_m((x_1, \dots, x_n)) := x_m; \end{aligned}$$

$$(2.7) \quad P(\omega) = P((x_1, \dots, x_n)) = P_1(x_1) \cdot \dots \cdot P_n(x_n).$$

Dann gilt offenbar $P(X_1 = x_1, \dots, X_n = x_n) = P((x_1, \dots, x_n)) = P_1(x_1) \cdot \dots \cdot P_n(x_n)$, sowie nach (3.0c)

$$\begin{aligned} P(X_1 = x_1) &= \sum_{(x_2, \dots, x_n) \in \mathfrak{X}_2 \times \dots \times \mathfrak{X}_n} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= \sum_{x_2 \in \mathfrak{X}_2} \dots \sum_{x_n \in \mathfrak{X}_n} P_1(x_1) \cdot \dots \cdot P_n(x_n) = P_1(x_1) \cdot \sum_{x_2 \in \mathfrak{X}_2} P_2(x_2) \sum_{x_n \in \mathfrak{X}_n} P_n(x_n) = P_1(x_1) \end{aligned}$$

wegen $\sum_{x_m \in \mathfrak{X}_m} P_m(x_m) = 1$. Ebenso erhält man $P(X_m = x_m) = P_m(x_m)$.

Damit haben wir auch schließlich $P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n)$. Ebenso zeigt man übrigens:

$$P(\Omega) = \sum_{(x_1, \dots, x_n) \in \mathfrak{X}_1 \times \dots \times \mathfrak{X}_n} P((x_1, \dots, x_n)) = 1, \text{ daß also } P \text{ wirklich ein W-maß ist.}$$

Ein Standardmodell in der W-theorie:

Handelt es sich um n Wiederholungen eines Einzelexperimentes unter gleichen Bedingungen, so wird man ansetzen: $\mathfrak{X}_1 = \dots = \mathfrak{X}_n = \mathfrak{X}$ und $P_1 = \dots = P_n = P'$.

Dann haben die Zva X_1, \dots, X_n dieselbe W-Funktion und dieselbe Verteilung; man sagt:

die Zva X_1, \dots, X_n sind **identisch verteilt** mit Verteilung P' (und unabhängig)

[und **nicht** "gleich verteilt"; X_1, \dots, X_n sind gleichverteilt bedeutet: $P(X_m = x_m) = \text{card}(\mathfrak{X}_m)^{-1}$].

2.4 Einige Verteilungen

Die Binomialverteilung: Wir betrachten ein Experiment, daß aus n sich nicht beeinflussenden Wiederholungen eines Einzelexperiment besteht mit zwei Ausgängen 0 und 1 (etwa). Dieses beschreiben wir durch unabhängige identisch verteilte Zva X_1, \dots, X_n auf (Ω, P) mit Werten in $\mathfrak{X} = \{0, 1\}$. [Die Existenz ist nach 2.3 klar.] Das Modell ist dann vollständig festgelegt durch den Parameter:

$$p = P(X_m = 1) \text{ mit } P(X_m = 0) = 1-p.$$

Sieht man das Eintreten von $X_m = 1$ als Erfolg an [und $X_m = 0$ als Mißerfolg], so heißt p die **Erfolgsw.** Die Verteilung von X_m , also das W–maß P' auf $\{0, 1\}$ mit $P'(1) = p$ heißt **Bernoulli–Verteilung** mit Parameter p und das Einzelexperiment heißt **Bernoulli–Experiment**. Dann gilt:

$$(2.8a) \quad P(X_1=x_1, \dots, X_n=x_n) = p^k \cdot (1-p)^{n-k} \text{ mit } k := \text{card}(\{m; x_m = 1\}) = \sum_{m=1}^n x_m.$$

Betrachten wir jetzt die Zva $S := \sum_{m=1}^n X_m$, die die Anzahl der Erfolge angibt. Dann gilt: $\{S = k\} = \{(X_1, \dots, X_n) \in A\}$ für eine geeignete Menge A . Somit folgt mit (3.0e):

$$(2.8b) \quad P(S=k) = \sum_{(x_1, \dots, x_n); x_1 + \dots + x_n = k} P(X_1=x_1, \dots, X_n=x_n) = \binom{n}{k} p^k \cdot (1-p)^{n-k}.$$

Dabei ist $\binom{n}{k}$ die Anzahl der Möglichkeiten, k Einsen auf n Plätze zu verteilen. Wir setzen

$$(2.9) \quad b_{n,p}(k) := \binom{n}{k} p^k \cdot (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

Dabei ist $k \mapsto b_{n,p}(k)$ offenbar eine W–funktion auf $\{0, 1, \dots, n\}$. Das zugehörige W–maß auf $\{0, 1, \dots, n\}$ heißt **Binomialverteilung (Bernoulli–Verteilung für $n=1$)** mit den Parametern n und p oder $b_{n,p}$ –Verteilung. Man setzt oft $q := 1-p$. Anstelle von "die Verteilung von S ist eine Binomialverteilung", sagt man auch

" S ist eine **binomial–verteilte** (bzw. **Bernoulli–verteilte**) Zva".

Lemma Sind X_1, \dots, X_n identisch verteilt, so auch $Y_1 := f \circ Y_1, \dots, Y_n := f \circ Y_n$ für jede Funktion $f: \mathfrak{X} \mapsto \mathfrak{Y}$.

Beweis. Wie im Beweis von Satz 3.5 hängt $P(Y_m = y) = P(X_m \in A)$ nicht von m ab mit $A = \{x \in \mathfrak{X}; f(x) = y\} = f^{-1}(\{y\})$. \square

Beispiele

(1) (*Geburtstagszwillinge*) Wir fragen jetzt: Wie groß ist die W. p , daß mindestens zwei von n Studierenden in einer Vorlesung (oder von n rein zufällig ausgewählten Personen) an einem bestimmten Tag, etwa dem 31. Dezember, Geburtstag haben. Als Ergebnisraum kann man wieder wählen: $\Omega = \{\omega = (x_1, \dots, x_n); x_i \in \{1, \dots, N\} \text{ für } i = 1, \dots, n\}$ mit $N = 365$.

Wählt man als mathematisches Modell den Laplaceschen W -raum (Ω, P) und beschreiben die Zva $X_m(\omega) := x_m$ das Ergebnis der Befragung der m -ten Person, so sind X_1, \dots, X_n unabhängige, identisch verteilte Zva und die Verteilung von X_m ist die Gleichverteilung auf $\{1, \dots, N\}$ [vgl. Beispiel 3.2 n -maliges Werfen eines Würfels, oder Übungen].

Wenn wir mit diesem Modell nicht zufrieden sind, weil wir Zweifel an der Annahme der Gleichverteilung haben, wissen wir jetzt, wie wir eine allgemeinere Situation beschreiben können. Wir setzen voraus:

X_1, \dots, X_n sind unabhängige, identisch verteilte Zva auf einem W -raum (Ω, P) mit Werten in $\{1, \dots, N\}$ und einer beliebigen Verteilung P' .

Sei nun $X'_m := f(X_m) = 0$ für $X_m \neq N$ und $X'_m := f(X_m) = 1$ für $X_m = N$.

Dann sind die X'_1, \dots, X'_n nach Satz 3.5a und dem obigen Lemma wieder unabhängige, identisch verteilte Zva jetzt mit einer Bernoulli-Verteilung. Die Erfolgsw. ist $p := P(X_m = N)$, also die W. dafür, daß eine befragte Person am 31. Dezember Geburtstag hat (z.B. $p = 1/N$ bei einer Gleichverteilung). Also hat $\sum_{m=1}^n X'_m$ eine $b_{n,p}$ -Verteilung.

Gefragt ist jetzt nach der W.: $P(\sum_{m=1}^n X'_m \geq 2) = 1 - P(\sum_{m=1}^n X'_m < 2)$.

Wir wissen nun:

$$P(\sum_{m=1}^n X'_m < 2) = b_{n,p}(0) + b_{n,p}(1) = (1-p)^n + n \cdot p \cdot (1-p)^{n-1} = [1 + (n-1) \cdot p] \cdot (1-p)^{n-1} \\ \approx [1 + (n-1) \cdot p] \cdot e^{-(n-1) \cdot p}. \quad \square$$

(2) (*Geburten*) Ist p die Häufigkeit von Mädchengeburten und nimmt man an, daß sich aufeinanderfolgende Geburten nicht beeinflussen, so berechnet sich in einem mathematischen Modell, die W., daß in einer Familie mit 4 Kindern 2 Jungen und zwei Mädchen vorkommen, zu:

$$\binom{4}{2} p^2 (1-p)^2 \quad [\approx 0,374 \text{ für } p = 0,51 \text{ oder } 0,49 \text{ bzw. } = 0,375 \text{ für } p = \frac{1}{2}].$$

Die Multinomialverteilung. Wir betrachten jetzt den allgemeinen Fall, daß bei einem Einzelexperiment endlich viele, etwa r mögliche Ausgänge beobachtet werden und nicht nur zwei (Erfolg, Mißerfolg). Seien dazu X_1, \dots, X_n unabhängige, identisch verteilte Zva auf einem W -raum (Ω, P) mit Werten in $\mathcal{X} = \{\alpha_1, \dots, \alpha_r\}$ und mit $p_j := P(X_m = \alpha_j)$, $1 \leq j \leq r$.

Seien Zahlen k_j vorgegeben mit $k_j \geq 0$, $k_1 + \dots + k_r = n$. Wir betrachten nun das Ereignis

$$A_{k_1, \dots, k_r} = \{\omega \in \Omega; k_j \text{ der Werte } X_1(\omega), \dots, X_n(\omega) \text{ sind gleich } \alpha_j, 1 \leq j \leq r\}.$$

Dann gilt (vgl. Übung):

$$P(A_{k_1, \dots, k_r}) = \binom{n}{k_1, k_2, \dots, k_r} p_1^{k_1} \cdot \dots \cdot p_r^{k_r} = \frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!} p_1^{k_1} \cdot \dots \cdot p_r^{k_r}.$$

Dabei gilt z.B. $p_1^{k_1} \cdot \dots \cdot p_r^{k_r} = P(X_1 = \alpha_1, \dots, X_{k_1} = \alpha_1, X_{k_1+1} = \alpha_2, \dots, X_{k_1+k_2} = \alpha_2, \dots)$.

Die (abgebrochene/gestutzte) **Geometrische Verteilung**. Seien X_1, \dots, X_n unabhängige, identisch verteilte Zva auf einem W-Raum (Ω, P) mit Werten in $\{0, 1\}$ und mit $p := P(X_m = 1)$ und

$$N(\omega) := \inf \{m; 1 \leq m \leq n, X_m(\omega) = 1\} \text{ mit } \inf \emptyset := \infty.$$

Dann ist N eine Zva mit $N : \Omega \mapsto \{1, \dots, n, \infty\}$.

Dann beschreibt N den Zeitpunkt des ersten Erfolges und $N-1 : \Omega \mapsto \{0, \dots, n-1, \infty\}$ die **Wartezeit** oder die Anzahl der Mißerfolge bis zum ersten Erfolg. Dann erhalten wir

$$P(N=k) = P(X_1=0, \dots, X_{k-1}=0, X_k=1) = P(X_1=0) \dots P(X_{k-1}=0)P(X_k=1), \text{ also}$$

$$P(N=k) = p \cdot (1-p)^{k-1} \text{ für } 1 \leq k \leq n \text{ und } P(N=\infty) = (1-p)^n,$$

$$P(N-1=k) = p \cdot (1-p)^k \text{ für } 0 \leq k \leq n-1.$$

Diese W. sind für $k < n$ unabhängig von n . Wir werden in §2.5 noch den Fall $n \rightarrow \infty$ behandeln.

2.5 Diskrete Wahrscheinlichkeitsräume

Wir kommen auf das Beispiel des letzten Abschnitts zurück. Dann ist $p \cdot (1-p)^k$ die W. dafür, daß die Wartezeit bis zum ersten Erfolg gleich k ist bei mehr als k Bernoulli-Versuchen. Nun gilt:

$$\sum_{k=0}^{\infty} p \cdot (1-p)^k = 1 .$$

Also ist es naheliegend $\omega \mapsto P(\omega) = p \cdot (1-p)^\omega$ als W-funktionen auf $\mathbb{N}_0 = \{0,1,2,\dots\} =: \Omega$ aufzufassen. Definiert man $P(A)$ für $A \subset \Omega$ wie in §1 gemäß

$$(1.10) \quad P(A) = \sum_{\omega \in A} P(\omega) \quad \text{für } A \subset \Omega,$$

so gilt offenbar wieder

$$(2.11) \quad P(\Omega) = 1;$$

$$(2.12) \quad P(A) \geq 0 \quad \text{für } A \subset \Omega;$$

$$(2.13)' \quad P(A \cup B) = P(A) + P(B) \quad \text{für alle disjunkten } A, B \text{ (Additivität).}$$

Es gilt sogar noch eine stärkere Eigenschaft, die sogenannte σ -Additivität:

$$(2.13) \quad P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P(A_i) \quad \text{für Folgen } A_1, A_2, \dots \text{ von disjunkten Ereignissen.}$$

Dabei ist disjunkt in Sinne von paarweise disjunkt zu verstehen.

Diese Eigenschaft folgt aus (1.10) mit dem sogenannten großen Umordnungssatz.

In endlichen Räumen sind bei einer Folge von disjunkten Ereignissen schließlich alle (alle bis auf endlich viele) leer; somit ist dann die σ -Additivität nicht stärker als die Additivität. Für ein abzählbares Ω ist diese Eigenschaft aber echt stärker.

Definition. Wir sprechen von einem **diskreten W-raum** (Ω, P) , wenn Ω eine beliebige endliche oder abzählbare Menge, P eine reelle Funktion auf Ω mit $\sum_{\omega \in \Omega} P(\omega) = 1$ ist. P heißt dann wieder **W-Funktion**.

Wird $P(A)$ gemäß (1.10) definiert, dann gelten die Eigenschaften (2.11) – (2.13).

P heißt dann wieder ein **W-maß**.

Ein **Beispiel** ist nun:

$\Omega = \{0,1,2,\dots\}$; $P(k) = p \cdot (1-p)^k$, $k \in \Omega$. P heißt dann **geometrische Verteilung** auf $\{0,1,2,\dots\}$.
[Man definiert manchmal auch die geometrische Verteilung P^* auf $\{1,2,\dots\}$ mit $P^*(k) = P(k-1)$.]

Ein weiteres wichtiges Beispiel wird die Poisson-Verteilung werden.

Die bisherigen Definitionen und bisher bewiesenen Sätze übertragen sich sofort auf diese allgemeinere Situation. Anstelle von Summen (mit endlich vielen Summanden) treten nun absolut konvergente Reihen. Bekanntlich kann man mit diesen ebenso rechnen wie mit Summen.

Die Lage wird erst schwieriger, wenn man überabzählbare W-Räume studiert, etwa

$$\Omega = [0, \infty), \quad \Omega = \mathbb{R} \quad \text{oder} \quad \Omega = \{0,1\} \times \{0,1\} \times \dots .$$

5.4 Die Faltung und die Poisson-Approximation.

Wir beginnen mit einer Erinnerung an die Formel von der totalen W. in der Form (3.0c):

$$P(Z = k) = \sum_{i \in \mathfrak{X}} P(X = i, Z = k).$$

Ist Ω (höchstens) abzählbar und $X : \Omega \rightarrow \mathfrak{X}$ Zva, so nimmt X auch (höchstens) abzählbar viele Werte an. Wir können also auch \mathfrak{X} als abzählbar voraussetzen.

Satz 5.6. Sind X und Y unabhängige Zva mit Werten in $\mathbb{N}_0 := \{0, 1, 2, \dots\}$, so ist

$$P(X+Y = k) = \sum_{i=0}^k P(X = i) \cdot P(Y = k-i).$$

Beweis. Es ist $P(X+Y = k) = \sum_{i=0}^k P(X = i, X+Y = k)$
 $= \sum_{i=0}^k P(X = i, Y = k-i) = \sum_{i=0}^k P(X = i) \cdot P(Y = k-i). \quad \square$

Definition. Sind $P'(i), P''(i)$ zwei W-Funktionen auf \mathbb{N}_0 , so nennt man die W-Funktion $P^*(k)$ mit

$$(5.19) \quad P^*(k) = \sum_{i=0}^k P'(i) \cdot P''(k-i)$$

die **Faltung** $P^* = P' * P''$ von P' und P'' .

Die Faltung ist offenbar wieder eine W-Funktion wegen $P^*(k) \geq 0$ und

$$\begin{aligned} \sum_{k=0}^{\infty} P^*(k) &= \sum_{k=0}^{\infty} \sum_{i=0}^k P'(i) \cdot P''(k-i) \\ &= \sum_{i=0}^{\infty} P'(i) \sum_{k=i}^{\infty} P''(k-i) = \sum_{i=0}^{\infty} P'(i) \sum_{j=0}^{\infty} P''(j) = 1 \cdot 1 = 1. \end{aligned}$$

In der Situation von Satz 5.6 gilt also:

Korollar. Die W-Funktion von $X+Y$ ist die Faltung der W-Funktion von X und der von Y .

Beispiel. (2-maliges Würfeln).

In §3.1 hatten wir die Situation betrachtet $P(X_1=i, X_2=j) = 1/36$ für $1 \leq i \leq 6, 1 \leq j \leq 6$.

Dann gilt gemäß Beispiel 3.2:

$$P(X_1 = i) = 1/6 = P(X_2 = j) \text{ für } 1 \leq i \leq 6, 1 \leq j \leq 6 \text{ und } X_1 \text{ und } X_2 \text{ sind unabhängig.}$$

Dabei können wir X_1 und X_2 auch als \mathbb{N}_0 -wertige Zva auffassen.

Für $i=0$, und $i>6$ haben wir $P(X_m = i) = 0$. Wir erhalten nun

$$\begin{aligned} P(X_1+X_2 = k) &= \sum_{i=0}^k P(X_1 = i) \cdot P(X_2 = k-i) \\ &= \sum_{\substack{1 \leq i \leq 6 \\ 1 \leq k-i \leq 6}} \frac{1}{36} = \frac{1}{36} \cdot \text{card}(\{(i,j); 1 \leq i,j \leq 6, i+j = k\}) \end{aligned}$$

in Übereinstimmung mit §3.1. \square

Beispiel. Sind X und Y unabhängig, hat X eine $b_{n,p}$ -Verteilung und Y eine $b_{m,p}$ -Verteilung, so hat $X + Y$ eine $b_{n+m,p}$ -Verteilung. Zum Beweis kann man nachrechnen:

$$b_{n+m,p}(k) = \sum_{i=0}^k b_{n,p}(i) \cdot b_{m,p}(k-i), \text{ d.h.} \\ \binom{n+m}{k} p^k \cdot (1-p)^{n+m-k} = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} \cdot (1-p)^{m-(k-i)}.$$

Man kann aber auch einen rein w -theoretischen Beweis führen. Dieser benutzt, daß in einer Folge von Bernoulli-Experimenten die Summe der Erfolge bei $n+m$ Versuchen gleich der Summe der Erfolge in der ersten n Versuchen plus der Summe der Erfolge in den anschließenden m Erfolgen ist.

Dazu kann man $n+m$ unabhängige identisch verteilte Zva $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ wählen mit Werten in $\{0,1\}$ und $P(X_\ell = 1) = p$. Setzt man dann

$$S := \sum_{\ell=1}^n X_\ell, \quad S' := \sum_{\ell=1}^m X_{n+\ell}, \text{ so weiß man:}$$

S und S' sind unabhängig nach Satz 3.5. Nach §2.4 weiß man auch:

S hat eine $b_{n,p}$ -Verteilung, S' hat eine $b_{m,p}$ -Verteilung und $S+S'$ hat eine $b_{n+m,p}$ -Verteilung.

Nach dem Korollar ist die Faltung der W -Funktion von S und der von S' die W -Funktion von $S+S'$. Man schreibt auch

$$b_{n+m,p} = b_{n,p} * b_{m,p}. \quad \square$$

Definition 5.7. Ist $P'(k)$ eine W -Funktion auf \mathbb{N}_0 mit

$$P'(k) = e^{-\lambda} \frac{\lambda^k}{k!} =: p(k|\lambda), \quad k = 0, 1, 2, \dots,$$

so heißt P' die W -Funktion einer **Poisson-Verteilung** mit Parameter $\lambda > 0$ ($P(\lambda)$ -Verteilung).

Also hat eine Zva X auf (Ω, P) eine **Poisson-Verteilung** mit Parameter $\lambda > 0$ ($P(\lambda)$ -Verteilung), wenn gilt:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} = p(k|\lambda), \quad k = 0, 1, 2, \dots.$$

Lemma 5.8. Sind die Zva X, Y unabhängig und haben eine $P(\lambda)$ - bzw. $P(\mu)$ -Verteilung, so hat $X+Y$ eine $P(\lambda+\mu)$ -Verteilung.

Beweis. Nach Satz 5.6 gilt:

$$P(X+Y = k) = \sum_{i=0}^k P(X = i) \cdot P(Y = k-i) = \sum_{i=0}^k p(i|\lambda) \cdot p(k-i|\mu) \\ = \sum_{i=0}^k e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} = e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^k k! \frac{\lambda^i}{i!} \frac{\mu^{k-i}}{(k-i)!} \\ = e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} \lambda^i \mu^{k-i} = e^{-(\lambda+\mu)} \frac{1}{k!} (\lambda+\mu)^k \text{ nach der Binomischen Formel. } \quad \square$$

Satz 5.10 (Poisson–Approximation der Binomialverteilung).

Ist $p(n)$ eine Folge mit $0 \leq p(n) \leq 1$ und $n \cdot p(n) \rightarrow \lambda > 0$, so gilt:

$$b_{n,p(n)}(k) = \binom{n}{k} p(n)^k \cdot (1-p(n))^{n-k} \rightarrow p(k|\lambda) = e^{-\lambda} \frac{\lambda^k}{k!} \text{ für } n \rightarrow \infty.$$

Beweis. Wir halten k fest und setzen $\lambda_n = n \cdot p(n)$. Dann ist

$$b_{n,p(n)}(k) = \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \lambda_n^k \cdot \left(1 - \frac{\lambda_n}{n}\right)^n \cdot \left(1 - \frac{\lambda_n}{n}\right)^{-k}$$

Wir wissen: $\frac{\lambda_n}{n} \rightarrow 0$ wegen $\lambda_n \rightarrow \lambda$, $\frac{n-i+1}{n} \rightarrow 1$, $\lambda_n^k \rightarrow \lambda^k$, $\left(1 - \frac{\lambda_n}{n}\right)^n \rightarrow e^{-\lambda}$ wegen $\lambda_n \rightarrow \lambda$.

[Zu $\varepsilon > 0$ existiert nämlich ein n_0 mit $\left(1 - \frac{\lambda+\varepsilon}{n}\right)^n \leq \left(1 - \frac{\lambda_n}{n}\right)^n \leq \left(1 - \frac{\lambda-\varepsilon}{n}\right)^n$ für $n \geq n_0$, wobei bekanntlich $\left(1 - \frac{\lambda \pm \varepsilon}{n}\right)^n \rightarrow e^{-(\lambda \pm \varepsilon)}$.] Damit folgt die Behauptung. \square

Die Bedingung $n \cdot p(n) \rightarrow \lambda$ impliziert, daß $p(n)$ sehr klein wird. Man spricht auch von dem **Gesetz der seltenen Ereignisse**. Für kleine p hat man also

$$b_{n,p}(k) = \binom{n}{k} p^k \cdot (1-p)^{n-k} \approx p(k|np) = e^{-np} \frac{(np)^k}{k!}.$$

Beispiel. (1) In einem Hörsaal seien $n=91$ Studierende. Die W., heute Geburtstag zu haben ist (etwa) $p = 1/365$ und sehr klein. Die Zahl X derer, die heute Geburtstag haben, können wir also als Zva mit einer $P(\lambda)$ -Verteilung modellieren mit $\lambda = 91/365 \approx 0,25$.

Benutzen wir wieder die Bezeichnungen aus Beispiel (1) auf 2.4, so wählen wir X'_1, \dots, X'_n unabhängig und identisch verteilt mit Werten in $\{0,1\}$ und $P(X'_n = 1) = p$, $X = \sum_{m=1}^n X'_m$. Dann gilt für die W., daß mindestens zwei Studierende heute Geburtstag haben:

$$P(X \geq 2) \approx 1 - (e^{-np} + e^{-np} np) = 1 - e^{-np}(1 + np). \quad \square$$

(3) Bei der Produktion von Blitzlampen ist ein kleiner Anteil $p = 0,015$ defekt. Wie muß man n wählen, damit man bei n Exemplaren mit $W. \geq 0,8$ mindestens 100 intakte Exemplare erhält ?

Unter Unabhängigkeitsannahmen wollen wir die Anzahl X der defekten Exemplare als Zva mit einer $b_{n,p}$ -Verteilung modellieren. Wir wollen also n so wählen, daß gilt mit $\lambda_n = np$:

$$P(n - X \geq 100) = P(X \leq n-100) = \sum_{k=0}^{n-100} b_{n,p}(k) \approx e^{-\lambda_n} \sum_{k=0}^{n-100} \frac{\lambda_n^k}{k!} \geq 0,8 \text{ für } n \geq 102.$$

Eigentlich haben wir X als Zva mit einer Hypergeometrischen Verteilung modelliert (vgl. § 1.4). Für eine große Zahl N von Blitzlampen ist allerdings gerechtfertigt, auch mit einer Binomialverteilung zu arbeiten (vgl. Übungen). Anschaulich spielt es bei großem N keine Rolle mehr, ob man zurücklegt oder nicht. \square

3.3 Erwartungswerte

Es liege die Situation von §1.1 vor, in der $P(\omega)$, $\omega \in \Omega$, über eine statistische Schätzung auf Grund einer Stichprobe $x_1, \dots, x_n \in \Omega$ gewonnen wurde gemäß $P(\omega) = \frac{1}{n} \text{card}(\{i; 1 \leq i \leq n, x_i = \omega\})$, z.B. $\Omega = \{0,1\}$.

Betrachten wir jetzt eine Funktion $X(\omega)$ und fragen nach dem arithmetischen Mittel $\frac{1}{n} \sum_{i=1}^n X(x_i)$ der Werte von X in der Stichprobe, so können wir schreiben:

$$\frac{1}{n} \sum_{i=1}^n X(x_i) = \frac{1}{n} \sum_{\omega \in \Omega} \text{card}(\{i; 1 \leq i \leq n, x_i = \omega\}) X(\omega) = \sum_{\omega \in \Omega} P(\omega) X(\omega).$$

Diesen Ausdruck werden wir als Vorbild nehmen, um einer Zva X einen mittleren Wert zuzuordnen.. Wir gewichten dabei die Werte $X(\omega)$ mit den W. $P(\omega)$.

Definition 3.6. Ist X eine reelle Zva auf dem (diskreten) W -raum (Ω, P) , so sagt man, daß die **Erwartung von X existiert**, wenn $\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\omega)$ konvergiert. Dann definieren wir:

$$EX = E(X) = E[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) \quad \text{als den Erwartungswert von } X.$$

Ist Ω endlich, so existiert die Erwartung von X immer.

Unter der Bedingung, daß $\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\omega)$ konvergiert, ist der Wert $\sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$ unabhängig von der Reihenfolge, in der wir aufsummieren.

[Wenn Ω abzählbar ist, so können wir ja $\Omega = \{\omega_1, \omega_2, \dots\}$ schreiben und somit auch

$$\sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) = \sum_{i=1}^{\infty} X(\omega_i) \cdot P(\omega_i).$$

Dieser Wert ist dann unabhängig davon, wie wir Ω abzählen.]

Beispiel. Der Erwartungswert ist eine Verallgemeinerung des arithmetischen Mittels. Ist Ω endlich und P die Gleichverteilung auf Ω , so erhält man gerade das arithmetischen Mittel

$$EX = \frac{1}{\text{card}(\Omega)} \sum_{\omega \in \Omega} X(\omega). \quad \square$$

Nimmt X Werte in $\mathfrak{X} \subset \mathbb{R}$ (überabzählbar) an, so können wir \mathfrak{X} als abzählbar annehmen. Wegen

$\Omega = \bigcup_{x \in \mathfrak{X}} \{X = x\}$ erhalten wir durch eine Umordnung der Summanden (nach dem sogenannten großen Umordnungssatz):

$$EX = \sum_{x \in \mathfrak{X}} \sum_{\omega \in \{X=x\}} P(\omega) \cdot X(\omega).$$

Auf Grund von $P(X=x) = \sum_{\omega \in \{X=x\}} P(\omega)$ gilt somit die wichtige Formel:

$$(3.4) \quad EX = \sum_{x \in \mathfrak{X}} x P(X=x)$$

[= $\sum_i x_i P(X=x_i)$, wenn $\mathfrak{X} = \{x_1, x_2, \dots\}$ eine Abzählung von \mathfrak{X} ist].

Dabei ist $\sum_{x \in \mathfrak{X}} x \cdot P(X=x)$ genau dann absolut konvergent, wenn $\sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$ absolut konvergent ist. An dieser Formel (3.4) sehen wir, daß $EX = \sum_{x \in \mathfrak{X}} x \cdot P'(x)$ nur von der W-Funktion $P'(x) = P(X=x)$ von X abhängt und nicht direkt von P . Man sieht auch, daß die Erwartung von X immer existiert, wenn \mathfrak{X} endlich ist.

Beispiel. (zweimaliges Werfen eines Würfels)

$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$; $X : \Omega \rightarrow \mathfrak{X} = \{2, \dots, 12\}$ Summe der beiden Augenzahlen.

$$EX = \sum_{\omega \in \Omega} \frac{1}{36} X(\omega) = \sum_{x \in \{2, \dots, 12\}} P(X=x) \cdot x.$$

6	·	·	·	·	·	·
5	·	·	·	·	·	·
4	·	·	·	·	·	·
3	·	·	·	·	·	·
2	·	·	·	·	·	·
1	·	·	·	·	·	·
	1	2	3	4	5	6

 \xrightarrow{X}

·	·	·	·	·	·	·	·
	2	4					12

Die Formel in Def. 3.6 ist gut für theoretische Überlegungen, wie der folgende Satz zeigt. Die Formel (3.4) ist gut für Rechnungen.

Satz 3.7. Seien X, Y Zva auf (Ω, P) mit existierender Erwartung; dann gilt:

- (i) Für $\alpha, \beta \in \mathbb{R}$ existiert die Erwartung von $\alpha X + \beta$, und es ist: $E(\alpha X + \beta) = \alpha \cdot E(X) + \beta$.
- (ii) Die Erwartung von $X + Y$ existiert, und es ist $E(X + Y) = EX + EY$.
- (iii) Sind X, Y unabhängig, so existiert die Erwartung von $X \cdot Y$, und es ist $E(X \cdot Y) = EX \cdot EY$.
- (iv) Gilt $X(\omega) \leq Y(\omega)$, $\omega \in \Omega$, so ist $EX \leq EY$.
- (v) Gilt $X(\omega) \geq 0$, $\omega \in \Omega$, so ist $EX \geq 0$.
- (vi) Gilt $|X'(\omega)| \leq X(\omega)$, $\omega \in \Omega$, für eine Zva X' , so existiert die Erwartung von X' und man hat:
 $|E(X')| \leq E(|X'|) \leq EX$ (**Majorantenkriterium**).

Die Eigenschaften (i) und (ii) sind gerade die **Linearität** und (iv) die **Isotonie des Erwartungswertes**.

Beweis. (i) $\sum_{\omega \in \Omega} (\alpha X(\omega) + \beta) \cdot P(\omega) = \alpha \cdot \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) + \beta \cdot \sum_{\omega \in \Omega} P(\omega)$.

(ii) $\sum_{\omega \in \Omega} [X(\omega) + Y(\omega)] \cdot P(\omega) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) + \sum_{\omega \in \Omega} Y(\omega) \cdot P(\omega)$.

Eigenschaft (iii) wird später bewiesen. (iv) und (v) sind klar.

Die Eigenschaft (vi) folgt aus dem Majorantenkriterium für Reihen (und $\pm X' \leq |X'|$). \square

Beispiele.

(1) Hat X eine $b_{n,p}$ -Verteilung, so ist

$$EX = \sum_{k=0}^n k P(X=k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = np.$$

Ist z.B. $p = 1/3$, tritt also etwa in einem Drittel der Versuche Erfolg ein, so ist die mittlere Anzahl der Erfolge bei n Teilversuchen gerade $n/3$.

Für $n=1$ ist dies klar. Für $n>1$ können wir eine Rechnung umgehen, in dem wir unabhängige identisch verteilte Zva X_1, \dots, X_n auf einem Ω -raum (Ω, P) betrachten mit Werten in $\{0,1\}$ und $P(X_m = 1) = p$. Dann gilt für $S := X_1 + \dots + X_n$ einerseits "S hat dieselbe Verteilung wie X und damit den gleichen Erwartungswert" und andererseits " $ES = EX_1 + \dots + EX_n = n \cdot EX_1 = np$ ".

(2) Ist X hypergeometrisch verteilt, gilt also $P(X=s) = \binom{S}{s} \binom{N-S}{n-s} / \binom{N}{n}$, so hat X den gleichen Erwartungswert wie ein $b_{n,p}$ -verteilte Zva mit $p=S/N$ (vgl. Übung), also $EX = n \cdot \frac{S}{N}$.

(3) Hat X eine geometrische Verteilung auf $\{0,1,2,\dots\}$, gilt also $P(X=k) = p(1-p)^k$, so ist

$$EX = \sum_{k=0}^{\infty} k p(1-p)^k = \frac{1-p}{p}.$$

Der Beweis folgt aus $\frac{d}{dq} \sum_{k=0}^{\infty} q^k = \sum_{k=0}^{\infty} k q^{k-1} = \frac{d}{dq} \frac{1}{1-q} = \frac{1}{(1-q)^2} \cdot \square$

Ist X Zva mit Werten in \mathfrak{X} und $f: \mathfrak{X} \rightarrow \mathbb{R}$ eine reelle Funktion, so ist $f \circ X = f(X)$ eine reelle Zva.

Satz 3.8 (a) Ist X Zva auf (Ω, P) mit Werten in \mathfrak{X} und $f: \mathfrak{X} \rightarrow \mathbb{R}$ gegeben, so gilt:

Die Erwartung von $f(X)$ existiert genau dann, wenn $\sum_{x \in \mathfrak{X}} |f(x)| \cdot P(X=x)$ konvergiert. Dann gilt:

$$E(f(X)) = \sum_{x \in \mathfrak{X}} f(x) \cdot P(X=x).$$

(b) Sind X, Y Zva auf (Ω, P) mit Werten in \mathfrak{X} und \mathfrak{Y} und $f: \mathfrak{X} \times \mathfrak{Y} \rightarrow \mathbb{R}$ gegeben, so gilt:

Die Erwartung von $f(X, Y)$ existiert genau dann, wenn

$\sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{Y}} |f(x, y)| \cdot P(X=x, Y=y)$ konvergiert. Dann gilt:

$$E[f(X, Y)] = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{Y}} f(x, y) \cdot P(X=x, Y=y).$$

Beweis. a) Wie beim Beweis von (3.4) haben wir

$$\sum_{\omega \in \Omega} f(X(\omega)) \cdot P(\omega) = \sum_{x \in \mathfrak{X}} \sum_{\omega \in \{X=x\}} f(X(\omega)) \cdot P(\omega) = \sum_{x \in \mathfrak{X}} f(x) P(X=x).$$

Wenn dabei eine der Reihen absolut konvergiert, so auch die anderen beiden.

b) Die Eigenschaft (b) ist keine echte Verallgemeinerung von (a). Setzt man nämlich $\hat{X} := (X, Y)$ und $\hat{\mathfrak{X}} := \mathfrak{X} \times \mathfrak{Y}$, so sind wir wieder in der Situation von (a). \square

Die letzte Eigenschaft folgt aus

$$E(\mathbf{1}_A) = 0 \cdot P(\mathbf{1}_A=0) + 1 \cdot P(\mathbf{1}_A=1) = P(A) \quad \text{gemäß (3.4) bzw.}$$

$$E(\mathbf{1}_A) = \sum_{\omega \in \Omega} \mathbf{1}_A(\omega) \cdot P(\omega) = \sum_{\omega \in A} P(\omega) = P(A) \quad \text{gemäß Def. 3.6.}$$

Aus (3.6) folgt durch Induktion:

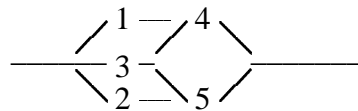
$$(3.11) \quad \mathbf{1}_{A_1 \cap \dots \cap A_n}(\omega) = \mathbf{1}_{A_1}(\omega) \cdot \dots \cdot \mathbf{1}_{A_n}(\omega)$$

Mit der de Morganschen Regel $(A_1 \cup \dots \cup A_n)^c = A_1^c \cap \dots \cap A_n^c$ gilt dann:

$$(3.12) \quad \mathbf{1}_{A_1 \cup \dots \cup A_n}(\omega) = 1 - (1 - \mathbf{1}_{A_1}(\omega)) \cdot \dots \cdot (1 - \mathbf{1}_{A_n}(\omega)).$$

Damit haben wir eine weitere Möglichkeit, die W. von der Vereinigung von nicht notwendigerweise disjunkten Ereignissen zu berechnen.

Beispiel 3.9 aus der *Zuverlässigkeitstheorie*.



Wir betrachten ein System mit 5 Stationen 1, ..., 5. Dazu wählen wir unabhängige Zva X_1, \dots, X_5 mit Werten in $\{0, 1\}$ und mit $P(X_i=1) = p_i$. Das Ereignis $\{X_i=1\}$ bedeutet, daß Station i intakt ist. Das System fällt nicht aus, solange es einen Weg von links nach rechts über intakte Stationen gibt. Das entsprechende Ereignis A ist also:

$$A := \{X_1=1, X_4=1\} \cup \{X_3=1, X_4=1\} \cup \{X_3=1, X_5=1\} \cup \{X_2=1, X_5=1\}.$$

Dann haben wir mit (3.12)

$$\mathbf{1}_A = 1 - (1 - \mathbf{1}_{\{X_1=1, X_4=1\}}) \cdot (1 - \mathbf{1}_{\{X_3=1, X_4=1\}}) \cdot (1 - \mathbf{1}_{\{X_3=1, X_5=1\}}) \cdot (1 - \mathbf{1}_{\{X_2=1, X_5=1\}})$$

Dann ergibt sich durch Ausmultiplizieren und Wegheben einiger Terme:

$$\begin{aligned} \mathbf{1}_A &= \mathbf{1}_{\{X_3=1, X_5=1\}} + \mathbf{1}_{\{X_2=1, X_5=1\}} + \mathbf{1}_{\{X_1=1, X_4=1\}} + \mathbf{1}_{\{X_3=1, X_4=1\}} \\ &\quad - \mathbf{1}_{\{X_2=1, X_3=1, X_5=1\}} - \mathbf{1}_{\{X_3=1, X_4=1, X_5=1\}} - \mathbf{1}_{\{X_1=1, X_3=1, X_4=1\}} \\ &\quad - \mathbf{1}_{\{X_1=1, X_2=1, X_4=1, X_5=1\}} + \mathbf{1}_{\{X_1=1, X_2=1, X_3=1, X_4=1, X_5=1\}}. \end{aligned}$$

Damit erhält man mit der Linearität des Erwartungswerts und der Unabhängigkeit der X_i :

$$\begin{aligned} p &= E(\mathbf{1}_A) = p_3 \cdot p_5 + p_2 \cdot p_5 + p_1 \cdot p_4 + p_3 \cdot p_4 - p_2 \cdot p_3 \cdot p_5 - p_3 \cdot p_4 \cdot p_5 - p_1 \cdot p_3 \cdot p_4 - \\ &\quad - p_1 \cdot p_2 \cdot p_4 \cdot p_5 + p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5 \quad \square \end{aligned}$$

Beispiel. Seien X_1, \dots, X_n unabhängige, identisch verteilte Zva mit Werten in \mathfrak{X} und sei $B \subset \mathfrak{X}$.

Dann ist $\mathbf{1}_B$ eine Funktion f auf \mathfrak{X} , und man kann die Zva $f(X_m) = \mathbf{1}_B(X_m)$ bilden mit Werten in $\{0, 1\}$. Dabei ist $\mathbf{1}_B(X_m(\omega)) = 1$ für $X_m(\omega) \in B$ und sonst $= 0$. Wie in Beispiel (1) aus §2.4 gilt:

$\mathbf{1}_B(X_1), \dots, \mathbf{1}_B(X_n)$ sind unabhängige, identisch verteilte Zva mit Werten in $\{0, 1\}$ und

$$p := P(\mathbf{1}_B(X_m) = 1) = P(X_m \in B). \quad \square$$

3.5 Varianz und Kovarianz

Gesucht ist nun ein Maß für die "Streuung" einer Zva um den Erwartungswert $\mu := EX$. Wir wählen den mittleren quadratischen Abstand von X und μ .

Definition 3.12. Seien X, Y reelle Zva auf dem (diskreten) Ω -raum (Ω, P) , sodaß $E(X^2)$ und $E(Y^2)$ existieren. Dann existieren auch $\mu := EX, \nu := EY$, und es heißen

$$\text{Var}(X) := E[(X - \mu)^2] \quad \text{die Varianz von } X,$$

$$\text{Cov}(X, Y) := E[(X - \mu) \cdot (Y - \nu)] \quad \text{die Kovarianz von } X \text{ und } Y,$$

$$\rho = \rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \quad \text{der Korrelationskoeffizient.}$$

X und Y heißen **unkorreliert**, wenn $\text{Cov}(X, Y) = 0 = \rho(X, Y)$ ist.

Ebenso naheliegend wäre $E[|X - \mu|]$ als Streuungsmaß. Mit der oben gegebenen Definition läßt sich aber besser rechnen, wie Satz 3.13(viii) zeigt.

Bemerkungen: (2) Offenbar gilt nach Satz 3.8, wenn X und Y Werte in \mathfrak{X} bzw. \mathfrak{Y} annehmen:

$$\text{Var}(X) = \sum_{x \in \mathfrak{X}} (x - \mu)^2 P(X=x),$$

$$\text{Cov}(X, Y) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{Y}} (x - \mu) \cdot (y - \nu) P(X=x, Y=y).$$

(0) Zva mit gleichem Erwartungswert können sehr verschiedene Varianzen haben. Gilt für die Gewinne X und X' von Glücksspielen:

$$P(X = 1) = \frac{1}{2}, P(X = -1) = \frac{1}{2},$$

$$P(X' = c) = \frac{1}{c+1}, P(X' = -1) = \frac{c}{c+1},$$

so hat man $EX = 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = 0,$

$$E(X') = c \cdot \frac{1}{c+1} - 1 \cdot \frac{c}{c+1} = 0 \quad (\text{man spricht dann auch von fairen Spielen})$$

und $\text{Var}(X) = 1^2 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2} = 1$

$$\text{Var}(X') = c^2 \cdot \frac{1}{c+1} + (-1)^2 \cdot \frac{c}{c+1} = c \left[\text{also } \text{Var}(X') > \text{Var}(X) \Leftrightarrow c > 1 \right].$$

(1) Die Ausdrücke in Def. 3.12 existieren wegen der folgenden Beziehungen:

$$1 + E(X^2) = E(1 + X^2) \geq E(|X|) \geq |EX| \quad \text{nach dem Majorantenkriterium 3.7(vi).}$$

$$E(X^2) - 2c \cdot EX + c^2 = E(X^2 - 2c \cdot X + c^2) = E[(X - c)^2] \quad \text{für alle } c \in \mathbb{R}, \text{ z.B. } c = \mu.$$

$$E(X^2) + E(Y^2) \geq E(|X| \cdot |Y|) \geq |E(XY)| \quad \text{wegen } (a^2 + b^2) \geq 2|ab| \geq |ab|.$$

In der letzten Beziehung kann man auch X und Y durch $X - \mu$ und $Y - \nu$ ersetzen.

(3) Es ist offenbar: $\text{Var}(X) = \text{Cov}(X, X)$.

(6) Es gilt: " $\text{Var}(X) = 0 \Leftrightarrow \exists c$ mit $P(X = c) = 1$ " (man sagt: X ist fast sicher konstant).

Der letzte Fall ist aber nicht interessant. Zum Beweis:

Aus $P(X = c) = 1$ folgt offenbar $EX = c$ und damit $\text{Var}(X) = 0$.

Gilt andererseits, etwa für $Z = X - \mu$: $E(Z^2) = \sum_z z^2 P(Z=z) = 0$, so folgt:

$P(Z=z) = 0$ für $z \neq 0$ und somit $P(Z=0) = 1$.

Satz 3.13 (Rechenregeln). Sind X, Y reelle Zva, so daß $E(X^2), E(Y^2)$ existieren, und sind $a, b, \in \mathbb{R}$, so gilt:

- (i) $\text{Var}(X) = E(X^2) - (EX)^2$;
- (ii) $\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$, $\left[\text{vgl. } E(aX+b) = aEX + b \right]$;
- (vi) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.
- (vii) Sind X und Y unabhängig, so sind X und Y unkorreliert.
- (viii) (Bienaymé) Sind X_1, \dots, X_n **unabhängige** reelle Zva, $\left[\text{soda\ss } E(X_m^2) \text{ existiert } \forall m, \right]$ so gilt:
 $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$.

Beweis. Sei $\mu = EX, v = EY$.

$$(i) E[(X - \mu)^2] = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu \cdot EX + \mu^2 = E(X^2) - \mu^2.$$

$$(ii) E[(aX+b) - (a\mu+b)]^2 = E(a^2 \cdot (X-\mu)^2) = a^2 \cdot E((X-\mu)^2).$$

(vi) Es ist nach Satz 3.7(ii):

$$\begin{aligned} E[(X+Y - (\mu+v))^2] &= E[(X - \mu + Y - v)^2] \\ &= E[(X - \mu)^2 + (Y - v)^2 + 2 \cdot (X - \mu) \cdot (Y - v)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y). \end{aligned}$$

(vii) Sind X und Y unabhängig, so auch $f(X) := X - \mu$ und $g(Y) := Y - v$ nach Satz 3.5.

Damit gilt nach Satz 3.7(iii): $\text{Cov}(X, Y) = E(X-\mu) \cdot E(Y-v) = 0 \cdot 0 = 0$.

Aussage (viii) folgt nun aus (vi) und (vii) mit Induktion nach n . \square

Beispiel. Hat X eine $P(\lambda)$ -Verteilung (Poisson-Verteilung), so gilt: $EX = \lambda = \text{Var}(X)$.

Es ist $E[f(X)] = \sum_{k=0}^{\infty} f(k) e^{-\lambda} \frac{\lambda^k}{k!}$, also

$$EX = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda,$$

$$E[X \cdot (X-1)] = \sum_{k=0}^{\infty} k \cdot (k-1) \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \lambda^2 \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = e^{-\lambda} \cdot \lambda^2 \cdot e^{\lambda} = \lambda^2.$$

Damit ergibt sich nach Satz 3.13(i):

$$\text{Var}(X) = E(X^2) - (EX)^2 = E[X \cdot (X-1) + X] - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad \square$$

Satz 3.14. Sind X, Y reelle Zva, so daß $E(X^2), E(Y^2)$ existieren, so gilt:

(Cauchy-Schwarzsche Ungleichung) $(E(X \cdot Y))^2 \leq E(X^2) \cdot E(Y^2)$.

Beweis. Sei $\alpha := E(Y^2)$ und $\beta := -E(XY)$.

Fall 1. $\alpha = 0$. Dann ist $P(Y = 0) = 1$ (siehe Bemerkung 6).

Wegen $\{Y=0\} \subset \{XY=0\}$ gilt auch $P(XY=0) = 1$ und damit auch $E(XY) = 0$.

Fall 2. $\alpha > 0$. Dann erhalten wir nach den Rechenregeln:

$$\begin{aligned} \alpha \cdot [E(X^2) \cdot E(Y^2) - (E(XY))^2] &= \alpha \cdot [E(X^2) \cdot E(Y^2) - 2(E(XY))^2 + (E(XY))^2] \\ &= \alpha^2 E(X^2) + 2\alpha\beta E(XY) + \beta^2 E(Y^2) = E(\alpha^2 X^2 + 2\alpha\beta XY + \beta^2 Y^2) = E((\alpha X + \beta Y)^2) \geq 0. \end{aligned}$$

Damit folgt die Ungleichung. \square

Bemerkungen. (1) Ersetzt man X, Y durch $|X|, |Y|$, so erhält man auch:

$$(E(|X \cdot Y|))^2 \leq E(X^2) \cdot E(Y^2).$$

(2) Ersetzt man X, Y durch $X' = X - EX, Y' = Y - EY$, so hat man :

$$(3.15) \quad \begin{aligned} \text{Cov}(X, Y)^2 &\leq \text{Var}(X) \cdot \text{Var}(Y), \quad |\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}, \\ -1 &\leq \rho(X, Y) \leq 1. \end{aligned}$$

Interpretation. Seien $X'(\omega) = X(\omega) - EX, Y'(\omega) = Y(\omega) - EY$ wieder die zentrierten Zva.

Positive [bzw. negative] Kovarianz bedeutet nach Def. 3.6

$$\text{Cov}(X, Y) = E(X' \cdot Y') = \sum_{\omega} X'(\omega) \cdot Y'(\omega) P(\omega) > 0 \text{ [bzw. } < 0 \text{]}.$$

Das kann man so interpretieren, daß X' häufig das gleiche [bzw. das entgegengesetzte] Vorzeichen hat wie Y' . Positive Kovarianz ist gleichbedeutend mit **positiver Korrelation** [$\rho(X, Y) > 0$].

Man kann leicht zeigen, daß die Extremfälle $\rho(X, Y) = \pm 1$ nur dann gelten, wenn

$$P(X = c \cdot Y + d) = 1 \text{ oder } P(Y = c \cdot X + d) = 1 \text{ für gewisse Zahlen } c, d \in \mathbb{R}.$$

Stellen etwa X und Y Schulnoten eines zufällig ausgewählten Schülers dar, so wird man positive Korrelation erwarten. \square

Beispiel. Hat X eine $b_{n,p}$ -Verteilung, so ist $\text{Var}(X) = np(1-p)$.

Für $n=1$ ist dies leicht zu sehen, denn dann ist $\text{Var}(X) = E(X^2) - (EX)^2$ mit $EX = p = E(X^2)$.

Für $n>1$ können wir eine Rechnung umgehen, in dem wir unabhängige identisch verteilte Zva X_1, \dots, X_n auf einem Ω -raum (Ω, P) betrachten mit Werten in $\{0, 1\}$ und $P(X_i = 1) = p$. Dann gilt für $S := X_1 + \dots + X_n$ einerseits: S hat eine $b_{n,p}$ -Verteilung, also die gleiche Verteilung wie X und damit die gleiche Varianz und andererseits nach Satz 3.13(viii):

$$\text{Var}(S) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = np(1-p).$$

Offenbar ist $\text{Var}(X) = np(1-p)$ maximal für $p = \frac{1}{2}$. \square

3.6 Das schwache Gesetz der großen Zahl

Dieser Abschnitt bringt ein mathematisches Gegenstück zu dem empirischen Gesetz der großen Zahl, das man in der realen Welt beobachten kann. Zur Vorbereitung dient:

Satz 3.15. Sei X eine reelle Zva auf dem (diskreten) Ω -raum (Ω, P) , so daß $E(X^2)$ existiert. Dann gilt für $\varepsilon > 0$ die **Tschebyschewsche Ungleichung** mit $\mu := EX$:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot \text{Var}(X).$$

Beweis. Wir zeigen allgemeiner für $Z := X - a$ und jedes $a \in \mathbb{R}$: $P(|Z| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot E(Z^2)$.

Es ist: $E(Z^2) \geq E(Z^2 \cdot \mathbf{1}_{\{|Z| \geq \varepsilon\}}) \geq E(\varepsilon^2 \cdot \mathbf{1}_{\{|Z| \geq \varepsilon\}}) = \varepsilon^2 \cdot E(\mathbf{1}_{\{|Z| \geq \varepsilon\}}) = \varepsilon^2 \cdot P(|Z| \geq \varepsilon)$. \square

Satz 3.16. (*Schwaches Gesetz der großen Zahl*). Seien X_1, \dots, X_n unabhängige, identisch verteilte Zva auf (Ω, P) , sodaß $E(X_m^2)$ existiert und damit $\mu := E(X_m)$, $\sigma^2 := \text{Var}(X_m)$. Dann gilt für $\bar{X} := \frac{1}{n}(X_1 + \dots + X_n)$: $P\left[|\bar{X} - \mu| \geq \varepsilon\right] \leq \frac{1}{n} \cdot \frac{1}{\varepsilon^2} \cdot \sigma^2 \rightarrow 0$ für $n \rightarrow \infty$, $\forall \varepsilon > 0$.

Dieser Satz rechtfertigt ebenfalls, den Erwartungswert μ als mittleren Wert für einen Einzelversuch anzusehen, der durch jedes der X_m beschrieben wird.

Beweis: Wir haben wegen Satz 3.7 und Satz 3.13:

$$E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{1}{n} \cdot \sigma^2;$$

denn $E(\bar{X}) = \frac{1}{n} \cdot n\mu$ und $\text{Var}(\bar{X}) = \frac{1}{n^2} \cdot \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} \cdot [\text{Var}(X_1) + \dots + \text{Var}(X_n)] = \frac{1}{n^2} \cdot n\sigma^2$.

Die Bildung des arithmetischen Mittels erhält also den Erwartungswert und verringert die Varianz. Nun folgt die Behauptung sofort aus der Tschebyschewschen Ungleichung. \square

Beispiel. (*Schwache Form des Borelschen Gesetzes der großen Zahl*).

Seien Z_1, \dots, Z_n unabhängige, identisch verteilte Zva auf (Ω, P) mit Werten in \mathfrak{Z} , die wieder die (unabhängige) Wiederholung von n Einzelerperimenten beschreiben sollen. Für $B \subset \mathfrak{Z}$ sei nun $X_m := \mathbf{1}_B(Z_m)$ wie im Beispiel in §3.4. Dann ist $X_m(\omega) = 1$ für $Z_m(\omega) \in B$ und $= 0$ sonst; also:

$$\frac{1}{n} \cdot \text{card}(\{m; 1 \leq m \leq n, Z_m(\omega) \in B\}) = \frac{1}{n} \cdot \sum_{m=1}^n X_m(\omega) = \bar{X}(\omega).$$

Somit ist \bar{X} die **relative Häufigkeit** davon, daß bei einem Einzelversuch das Ergebnis Z_m in B liegt. Dann sind nach Satz 3.5 X_1, \dots, X_n wieder unabhängige, identisch verteilte Zva auf (Ω, P) , jetzt mit Werten in $\{0, 1\}$ und $P(X_m = 1) = P(Z_m \in B) =: p$, $E(X_m) = p = \mu$, $\text{Var}(X_m) = p(1-p) = \sigma^2$.

Wir haben jetzt nach Satz 3.16 wegen $p(1-p) \leq \frac{1}{4}$:

$$P\left[|\bar{X} - p| \geq \varepsilon\right] \leq \frac{1}{4} \cdot \frac{1}{n} \cdot \frac{1}{\varepsilon^2} \rightarrow 0 \text{ für } n \rightarrow \infty.$$

§ 5 Approximation der Binomialverteilung

5.1 Dichte der Normalverteilung

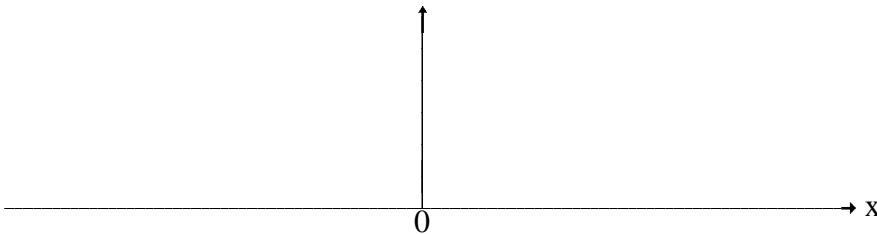
Die folgende sogenannte **Gaußsche Glockenkurve** wird von großer Bedeutung sein:

$$(5.11) \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x^2\right).$$

Sie heißt **Dichte der (Standard-) Normalverteilung** und hat die folgende Eigenschaft:

Lemma 5.3. Es gilt $\int_{-\infty}^{\infty} \varphi(x) dx = 1$.

Der Beweis soll hier nicht geführt werden.



Zeichne den Verlauf von $\varphi(x)$.

5.2 Der Satz von de Moivre–Laplace

Zu φ definieren wir noch die Funktion

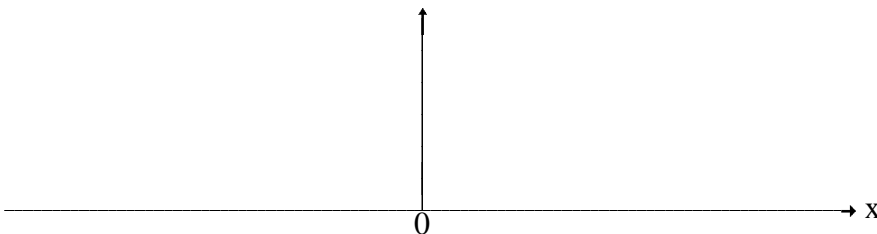
$$\Phi(x) := \int_{-\infty}^x \varphi(t) dt.$$

Sie heißt **Verteilungsfunktion Φ der (Standard-) Normalverteilung**. Offenbar ist

$$\int_a^b \varphi(t) dt = \Phi(b) - \Phi(a) \quad \text{für } a < b.$$

Für Φ existieren Tabellen und Approximationen. Es genügt eine Tabelle, die $\Phi(x)$ für $x \geq 0$ angibt. Wegen der Symmetrie $\varphi(-t) = \varphi(t)$ gilt nämlich

$$(5.13) \quad \Phi(-x) = \int_{-\infty}^{-x} \varphi(t) dt = \int_x^{\infty} \varphi(t) dt = 1 - \Phi(x).$$



Zeichne $\Phi(b) - \Phi(a)$, $\Phi(-x)$, $1 - \Phi(x)$ für $-x < a < b < x$ als Fläche unter φ !

Wir betrachten nun die folgende

Situation: X_1, \dots, X_n seien unabhängige, identisch verteilte reelle Zva, sodaß $E(X_m^2)$ existiert und damit $\mu := EX_m$ und $\sigma^2 := \text{Var}(X_m)$. Es sei $\sigma^2 > 0$. Wir setzen $S_n := X_1 + \dots + X_n$ und $\bar{X}_n = \frac{1}{n} S_n$.

Mit der Forderung $\sigma^2 > 0$ schließen wir gemäß Bemerkung (6) in §3.5 nur den trivialen Fall aus, daß $P(X_m = c) = 1$ für ein $c \in \mathbb{R}$ gilt. Wir wissen, in der obigen Situation gilt (vgl. Sätze 3.13, 3.16):

$$E(S_n) = n \cdot \mu, \text{Var}(S_n) = n \cdot \sigma^2, E(\bar{X}_n) = \mu, \text{Var}(\bar{X}_n) = \frac{1}{n} \sigma^2.$$

Dann ist $\bar{X}_n - \mu$ eine **zentrierte Zva** in dem Sinne, daß gilt: $E(\bar{X}_n - \mu) = 0$. Für die Varianz haben wir $\text{Var}(\bar{X}_n - \mu) = \text{Var}(\bar{X}_n) = \frac{1}{n} \sigma^2$. Durch eine Normierung der Form $c \cdot (\bar{X}_n - \mu)$ kann man noch erreichen, daß $\text{Var}(c \cdot (\bar{X}_n - \mu)) = 1$ gilt. Offenbar muß man $c = \sqrt{n}/\sigma$ wählen in

Hinblick auf Satz 3.13(ii). Wir setzen $S_n^* := \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu)$ und haben also

$$E(S_n^*) = 0, \text{Var}(S_n^*) = 1.$$

Man nennt S_n^* eine **standardisierte Zva**. Es gilt dabei offenbar

$$S_n^* = (S_n - n\mu)/\sigma \cdot \sqrt{n} = (S_n - E(S_n))/\sqrt{\text{Var}(S_n)} = (\bar{X}_n - E(\bar{X}_n))/\sqrt{\text{Var}(\bar{X}_n)}.$$

Für jede Zva Y ist allgemein $Z = (Y - E[Y])/\sqrt{\text{Var}(Y)}$ eine standardisierte Zva mit $EZ = 0$, $\text{Var}(Z) = 1$. Die Funktion $F_n(t) := P(S_n^* \leq t)$ werden wir Verteilungsfunktion von S_n^* nennen. [Manche Autoren nennen auch $P(S_n^* < t)$ Verteilungsfunktion von S_n^* .]

Nun können wir den Hauptsatz formulieren, der die Konvergenz der Verteilungsfunktion von S_n^* (also der standardisierten Summe von unabhängigen, identische verteilten Zva) gegen die Verteilungsfunktion der Normalverteilung beinhaltet:

Satz 5.4. (zentraler Grenzwertsatz). In der obigen Situation gilt:

$$\lim_{n \rightarrow \infty} P(S_n^* \leq t) = \Phi(t) \text{ gleichmäßig in } t \in \mathbb{R};$$

$$\lim_{n \rightarrow \infty} P(S_n^* < t) = \Phi(t) \text{ gleichmäßig in } t \in \mathbb{R}.$$

Das Grenzverhalten hängt also nur über den Erwartungswert μ und die Varianz σ^2 von der Verteilung der X_m ab.

Der Beweis von Satz 5.4 soll hier nicht geführt werden. Wegen

$$P(a \leq S_n^* \leq b) = P(\{S_n^* \leq b\} \setminus \{S_n^* < a\}) = P(S_n^* \leq b) - P(S_n^* < a)$$

gemäß (1.6) erhält man aus Satz 5.4

Korollar. $\lim_{n \rightarrow \infty} P(a \leq S_n^* \leq b) = \Phi(b) - \Phi(a)$ für alle $a < b \in \mathbb{R}$.

Bemerkung: Aussagen über S_n und \bar{X}_n .

$$|P(\alpha \leq S_n \leq \beta) - [\Phi(\frac{\beta - n\mu}{\sigma \cdot \sqrt{n}}) - \Phi(\frac{\alpha - n\mu}{\sigma \cdot \sqrt{n}})]| \rightarrow 0$$

$$|P(\gamma \leq \bar{X}_n - \mu \leq \delta) - [\Phi(\sqrt{n} \cdot \frac{\delta}{\sigma}) - \Phi(\sqrt{n} \cdot \frac{\gamma}{\sigma})]| \rightarrow 0 .$$

Zum Beweis der Limesaussage schreiben wir

$$a_n := \frac{\alpha - n\mu}{\sigma \cdot \sqrt{n}}, \quad b_n := \frac{\beta - n\mu}{\sigma \cdot \sqrt{n}}.$$

$$\begin{aligned} \text{Dann ist: } P(\alpha \leq S_n \leq \beta) &= P(\frac{\sqrt{n}}{\sigma} (\frac{1}{n} \alpha - \mu) \leq S_n^* \leq \frac{\sqrt{n}}{\sigma} (\frac{1}{n} \beta - \mu)) \\ &= P(a_n \leq S_n^* \leq b_n) = P(S_n^* \leq b_n) - P(S_n^* < a_n). \end{aligned}$$

Damit ergibt sich

$$\begin{aligned} &|P(\alpha \leq S_n \leq \beta) - [\Phi(b_n) - \Phi(a_n)]| \\ &= | [P(S_n^* \leq b_n) - \Phi(b_n)] - [P(S_n^* < a_n) - \Phi(a_n)] | \\ &\leq | [P(S_n^* \leq b_n) - \Phi(b_n)] | + | [P(S_n^* < a_n) - \Phi(a_n)] | \\ &\leq \sup_t | [P(S_n^* \leq t) - \Phi(t)] | + \sup_t | [P(S_n^* < t) - \Phi(t)] | \rightarrow 0 \quad \text{für } n \rightarrow \infty \end{aligned}$$

wegen der gleichmäßigen Konvergenz in Satz 5.4.

Für die **praktische Anwendung** erhalten wir somit das Rezept:

$$P(\alpha \leq S_n \leq \beta) \approx \Phi(\frac{\beta - n\mu}{\sigma \cdot \sqrt{n}}) - \Phi(\frac{\alpha - n\mu}{\sigma \cdot \sqrt{n}}).$$

Die Aussage für \bar{X}_n folgt ebenso. \square

Korollar. (Satz von de Moivre-Laplace). Sei $0 < p < 1$.

Ist S_n eine Zva auf (Ω, P) mit einer $b_{n,p}$ -Verteilung, so gilt:

$$\lim_{n \rightarrow \infty} P\left[a \leq \frac{1}{\sqrt{n \cdot p \cdot (1-p)}} (S_n - np) \leq b \right] = \Phi(b) - \Phi(a) \quad \text{für alle } a < b.$$

Zum **Beweis** wählen wir in der obigen Situation X_1, \dots, X_n als unabhängige, identisch verteilte Zva mit Werten in $\{0, 1\}$ und $P(X_m = 1) = p$. Dann gilt $\mu = EX = p$ und $\sigma^2 = \text{Var}(X_m) = p \cdot (1-p) > 0$.

Wir setzen $S'_n := X_1 + \dots + X_n$. und $S_n^* = (S'_n - np) / \sqrt{np(1-p)}$. Dann hat auch S'_n eine $b_{n,p}$ -Verteilung und es gilt also für ein geeignetes Intervall I :

$$P(a \leq (S_n - np) / \sqrt{np(1-p)} \leq b) = P(S_n \in I) = P(S'_n \in I) = P(a \leq (S'_n - np) / \sqrt{np(1-p)} \leq b) \rightarrow \Phi(b) - \Phi(a)$$

in Hinblick auf das vorangehende Korollar. \square

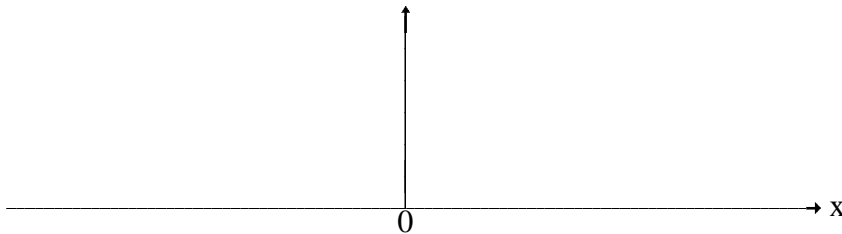
Für die **praktische Anwendung** ergibt sich jetzt in der Situation der Bemerkung:

$$P(i \leq S_n \leq k) = \sum_{j=i}^k \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{i - np}{\sqrt{np(1-p)}}\right).$$

Von *de Moivre* wurde die Approximation der $b_{n,p}$ -Verteilung untersucht für $n = 2m$, $p = \frac{1}{2}$.

Dann ist $np = m$, $np \cdot (1-p) = \frac{1}{2}m$, $\frac{1}{\sqrt{np \cdot (1-p)}} = \sqrt{\frac{2}{m}}$. Wir haben hier schon Symmetrie wegen:

$$b_{2m, \frac{1}{2}}(m-k) = \frac{(2m)!}{(m-k)! \cdot (m+k)!} \left(\frac{1}{2}\right)^{m-k} \left(\frac{1}{2}\right)^{m+k} = b_{2m, \frac{1}{2}}(m+k)$$



Zeichne $P\left[\sqrt{\frac{2}{m}} \cdot (S_n - m) = \pm x\right] > 0$

für $x = 0, \sqrt{\frac{2}{m}}, 2\sqrt{\frac{2}{m}}, \dots, (m-1)\sqrt{\frac{2}{m}}, m\sqrt{\frac{2}{m}} = \sqrt{2m}$.

Ist das Intervall $[a, b]$ noch so klein und liegt noch so "weit draußen", so liegen doch mit wachsendem n immer mehr Punkte der Form $x = k \cdot \sqrt{2/m}$, $-m \leq k \leq m$ in $[a, b]$.

Die W. $P(S_n^* \in \{\pm k \cdot \sqrt{2/m}, 0 \leq k \leq m\}) = 1$ wird im Limes immer mehr über \mathbb{R} "verschmiert". Es kann weiter gezeigt werden, daß gilt:

$$\sqrt{\frac{m}{2}} \cdot b_{n,p}(m+k) \approx \varphi(k \cdot \sqrt{2/m}).$$

5.3 Anwendung: Würfelexperiment.

Wie groß ist näherungsweise die W., bei 600 Würfeln mit einem korrekten Würfel mindestens 90 und höchstens 100 Sechsen zu erhalten.

Es ist $n = 600$, $p = 1/6$, also $np = 100$, $\sqrt{n \cdot p \cdot (1-p)} \approx 9,13$. Daraus folgt für die Anzahl S_n der geworfenen Sechsen mit $k = 100$, $i = 90$:

$$\begin{aligned} P(i \leq S_n \leq k) &\approx \Phi\left[\frac{k - 100}{9,13}\right] - \Phi\left[\frac{i - 100}{9,13}\right] \\ &= \Phi(0) - \Phi\left(-\frac{10}{9,13}\right) = \frac{1}{2} - \Phi(-1,095) \approx 0,36. \end{aligned}$$

Der exakte Wert ist 0,4024. \square

§4 Grundbegriffe der Schätztheorie

Beispiel 4.1a (*Qualitätskontrolle*, vgl. §1.4).

Wir stellen uns eine Lieferung von N gleichartigen Objekten vor, von denen S defekt und damit $N-S$ intakt sind. Als Stichprobe greift man n Objekte heraus und überprüft diese. Dann ist

$$h(x;n,N,S) = \binom{S}{x} \cdot \binom{W}{n-x} / \binom{N}{n}, \quad 0 \leq x \leq n \leq N, \quad n-x = w, \quad N-S = W,$$

die W . dafür, daß s der überprüften Objekte defekt sind.

Nun stellen wir uns aber die Situation vor, daß der Parameter S unbekannt ist, wir aber bereits s kennen. Wir wollen nun von x auf S schließen, also einen **Schätzwert** $\hat{S} = \hat{S}(x)$ für S ermitteln, der x als Information benutzt. Dies ist eine typische Aufgabe der Statistik.

In der W -Theorie geht man von bekannten Parametern aus und ermittelt dann die W ., daß ein bestimmtes Ereignis eintritt. In der Statistik geht man davon aus, daß ein Parameter nicht bekannt ist und ein bestimmtes Ereignis eingetreten ist. Dann versucht man, Rückschlüsse von dem Eintreten des Ereignisses auf den unbekannt Parameter zu ziehen.

Eine plausible Schätzung $\hat{S}(x)$ von S ergibt sich aus der naheliegenden Beziehung:

$$\hat{S}(x) / N \approx x/n, \quad \text{also} \quad \hat{S}(x) \approx \frac{x}{n} \cdot N.$$

Diese plausible Schätzung ergibt sich auch aus einem anderen einleuchtenden Prinzip, dem sogenannten *Maximum-Likelihood-Ansatz*, der den Vorteil hat, stark verallgemeinerungsfähig zu sein. Man überlegt sich, unter welchem Parameter [hier S] das bestimmte Ereignis [hier: es werden x defekte Objekte festgestellt] am wahrscheinlichsten ist. Man sucht also ein $\hat{S}(x)$ mit

$$h(x;n,N,\hat{S}(x)) = \max_S h(x;n,N,S).$$

Man sieht leicht, daß gilt: $\frac{h(x;n,N,S)}{h(x;n,N,S-1)} = \frac{S \cdot (N-S+1-n+x)}{(S-x) \cdot (N-S+1)}$.

Damit ist $h(x;n,N,S) > [=] h(x;n,N,S-1)$ äquivalent zu: $S < [=] \frac{x}{n} \cdot (N+1) =: \vartheta_0$.

Fall 1. $\vartheta_0 \in \mathbb{N}$. Dann liegt das Maximum $\max_S h(x;n,N,S)$ bei $S \in \{\vartheta_0 - 1, \vartheta_0\}$.

Fall 2. $\vartheta_0 \notin \mathbb{N}$. Dann liegt das Maximum bei $S = [\vartheta_0]$.

Also gilt in beiden Fällen für unseren Schätzer $\hat{S}(x)$: $\vartheta_0 - 1 \leq \hat{S}(x) \leq \vartheta_0$, $\hat{S}(x) \in \mathbb{N}_0$.

Damit gilt wieder $\hat{S}:N \approx x:n$ für große N . Man nennt $\hat{S}(s)$ einen **Maximum-Likelihood-Schätzer**.

Wir werden noch weitere Ansätze kennenlernen, mit denen man einen Schätzwert ermittelt.

4.1 Der allgemeine Rahmen von Schätzproblemen

Was ist Statistik ?

In der Umgangssprache versteht man darunter eine **Zusammenfassung und Darstellung von Daten**, beispielsweise in der Meinungsforschung oder in der Medizin. Diese Form der Statistik wird **deskriptive Statistik** genannt.

Die **Mathematische Statistik** geht von einer abstrakten Formulierung eines konkreten Problems in einem mathematischen Modell aus, das auf der W -theorie basiert. Ihr Ziel ist es, in dem Modell Verfahren zu entwickeln, die es in einer konkreten Situation gestatten, aufgrund gegebener Daten sogenannte **statistische Entscheidungen** zu fällen. Man spricht auch von **schließender** oder **analytischer Statistik**. Bei dem Bemühen, die bei den Entscheidungen auftretenden Fehler in einem geeigneten Sinn klein zu halten, werden Optimierungsmethoden verwendet.

Man hat die folgende Ausgangssituation. Ein Experiment liefert ein Ergebnis x , auch **Stichprobe** genannt, aus dem sogenannten **Stichprobenraum** \mathfrak{X} . Der Ausgang des Experiments birgt eine **doppelte Ungewißheit** in sich:

1. Gewisse Parameter sind unbekannt.
2. Selbst bei bekannten Parametern ist das Resultat des Experiments nicht vorhersagbar.

Problematik 2 läßt sich mit Mitteln der W -Theorie erfassen. Sind die Parameter bekannt, so läßt sich das Experiment durch eine W -Maß beschreiben. Dies ist eine **Grundannahme der Statistik**.

Problematik 1 ist charakteristisch für die Statistik. Sie wird dadurch berücksichtigt, daß man eine Familie von W -Maßen auf \mathfrak{X} vorausgesetzt wird. Unter diesen W -Maßen soll eins sein, das das Experiment richtig beschreibt.

Zur formalen Beschreibung eines **Schätzproblems** benötigen wir:

- eine (nichtleere, höchstens abzählbare) Menge \mathfrak{X} , den **Stichprobenraum**,
- eine Familie $\{P_{\vartheta}; \vartheta \in \Theta\}$ von W -maßen auf \mathfrak{X} , (die **Verteilungsannahme**),
 Θ heißt dann der **Parameterraum**,
- ein zu schätzender (1-dim.) Parameter $g(\vartheta)$, $g : \Theta \mapsto \mathbb{R}$ (z.B. $g(\vartheta) = \vartheta$).

Anstelle von Ω benutzt man in der Statistik gern ein anderes Symbol, hier also \mathfrak{X} . Oft ist ϑ ein Vektor; dann kann $g(\vartheta)$ etwa eine Komponente von ϑ sein.

Im Beispiel 4.1 war $\mathfrak{X} := \{0, \dots, n\}$, $\vartheta := S$, $P_{\vartheta}(s) := h(s; n, N, \vartheta) = h(s; n, N, S)$, $\Theta = \{0, 1, \dots, N\}$.

Die Parameter n, N werden als bekannt, also fest gewählt angesehen.

Definition. Eine Abbildung $\hat{g} : \mathfrak{X} \mapsto \mathbb{R}$ [oder $\hat{g} : \mathfrak{X} \mapsto g(\Theta) = \{g(\vartheta); \vartheta \in \Theta\}$] heißt ein **Schätzer** von $g(\vartheta)$.

Ist $g(\vartheta) = \vartheta$, $g(\Theta) = \Theta$, so schreibt man auch $\hat{\vartheta}(x)$ statt $\hat{g}(x)$.

Wir möchten noch einen Namen für einen typischen Fall einführen (vgl. § 2.3). Ist ein Parameter, der einen Versuch beschreibt, unbekannt, so wird man diesen Versuch unter gleichen Bedingungen möglichst oft wiederholen.

Definition. Ein **Einstichprobenproblem** liegt vor, wenn gilt:

$\mathfrak{X} = \mathfrak{X}' \times \dots \times \mathfrak{X}' \ni x = (x_1, \dots, x_n)$ mit den **Stichprobenumfang** n ,

$P_{\vartheta}(x) = P'_{\vartheta}(x_1) \cdot \dots \cdot P'_{\vartheta}(x_n)$ mit W -Funktionen P'_{ϑ} auf \mathfrak{X}' , $\vartheta \in \Theta$ (vgl. (2.7)).

$X_m(x) := x_m$ beschreibe das Ergebnis des m -ten Teilversuchs.

Dann gilt also (vgl. § 2.3):

Unter P_{ϑ} , d.h. bei festem ϑ , sind X_1, \dots, X_n unabhängige, identische verteilte Zva mit der W -funktion P'_{ϑ} . Das letzter bedeutet $P_{\vartheta}(X_m = x') = P'_{\vartheta}(x') \forall x' \in \mathfrak{X}'$.

Von nun an sollen hier nur Einstichprobenprobleme behandelt werden.

4.2 Maximum-Likelihood-Schätzer

Es liege ein Schätzproblem vor mit $g(\vartheta) = \vartheta$. Gesucht ist ein Schätzer $\hat{\vartheta}(x)$.

Definition. $L_x(\vartheta) := P_{\vartheta}(x)$ heißt **Likelihood-Funktion**. Gilt für einen Schätzer $\hat{\vartheta} : \mathfrak{X} \rightarrow \Theta$:

$$(4.1) \quad L_x(\hat{\vartheta}(x)) = \max \{L_x(\vartheta); \vartheta \in \Theta\},$$

so heißt $\hat{\vartheta}$ **Maximum-Likelihood-Schätzer** von ϑ .

Beispiel 4.2 (Schätzung einer Erfolgswahrscheinlichkeit).

Ein Einstichprobenproblem beschreibe n Bernoulli-Versuche mit unbekannter Erfolgsw.

Wir wählen: $\vartheta = p$, $\mathfrak{X} = \{0, 1\}^n$, $P'_{\vartheta}(1) := \vartheta$, $P'_{\vartheta}(0) = 1 - \vartheta$, also $\mathfrak{X}' = \{0, 1\}$ und für $x' \in \mathfrak{X}'$

$$P'_{\vartheta}(x') = \vartheta^{x'} \cdot (1 - \vartheta)^{1 - x'},$$

$$P_{\vartheta}(x) = \prod_{m=1}^n \vartheta^{x_m} \cdot (1 - \vartheta)^{1 - x_m} = L_x(\vartheta) \quad \text{für } x = (x_1, \dots, x_n) \in \mathfrak{X}.$$

Anstelle von $L_x(\vartheta)$ können wir auch $\ln L_x(\vartheta)$ maximieren. Dann gilt:

$$\ln L_x(\vartheta) = \sum_{m=1}^n \left[x_m \cdot \ln \vartheta + (1 - x_m) \cdot \ln (1 - \vartheta) \right] = \left(\sum_1^n x_m \right) \cdot \ln \vartheta + \left(n - \sum_1^n x_m \right) \cdot \ln (1 - \vartheta).$$

Setzt man $\bar{x} := \frac{1}{n} \sum_1^n x_m$ für die mittlere Anzahl der Erfolge, so hat man

$$\ln L_x(\vartheta) = n \cdot \bar{x} \cdot \ln \vartheta + n \cdot (1 - \bar{x}) \cdot \ln (1 - \vartheta) \quad \text{und}$$

$$(4.2) \quad \frac{d}{d\vartheta} \ln L_x(\vartheta) = n\bar{x} \cdot \frac{1}{\vartheta} - (n - n\bar{x}) \cdot \frac{1}{1 - \vartheta},$$

also $\frac{d}{d\vartheta} \ln L_x(\vartheta) = 0 \Leftrightarrow \vartheta = \bar{x}$. In der Tat ist \bar{x} die Maximumstelle von $\ln L_x(\vartheta)$.

$\bar{X} := \frac{1}{n} \sum_1^n X_m$ (also $\bar{X}(x) = \bar{x}$) die mittlere Anzahl der Erfolge.

Dabei gilt $\bar{X} : \mathfrak{X} \rightarrow [0, 1]$ und \bar{X} ist ein Maximum-Likelihood-Schätzer.

Nachdem man in einfachen Fällen überprüft hat, daß sich als Maximum-Likelihood-Schätzer ein sehr plausibler Schätzer ergibt, ist man geneigt, einen Maximum-Likelihood-Schätzer auch in komplizierten Situationen zu ermitteln.

4.3 Erwartungstreue

Es liege ein Schätzproblem zu einem eindimensionalen Parameter $g : \Theta \rightarrow \mathbb{R}$ vor.

Definition. Ein Schätzer \hat{g} von $g(\vartheta)$ heißt **erwartungstreu**, wenn gilt:

$$(4.3) \quad E_{\vartheta}(\hat{g}) := \sum_{x \in \mathcal{X}} \hat{g}(x) P_{\vartheta}(x) = g(\vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Da der "wahre Parameter" unbekannt ist, muß man alle Möglichkeiten in Betracht ziehen.

Wird das Schätzverfahren häufig angewandt, so liegt man also bei einem erwartungstreuen Schätzer zumindest im Mittel richtig, egal welcher Parameter der "wahre" (richtige) ist.

Gegeben sei nun ein **Einstichprobenproblem** (im Sinne von §4.1). Dann sind X_1, \dots, X_n unabhängige, identisch verteilte Zva auf $(\mathcal{X}, P_{\vartheta})$ (für jedes feste ϑ).

Wir interessieren uns nun für den Erwartungswert als wichtigen Parameter

$$\mu(\vartheta) := E_{\vartheta}(X_m) = \sum_{x' \in \mathcal{X}'} x' P_{\vartheta}(X_m = x') = \sum_{x' \in \mathcal{X}'} x' P'_{\vartheta}(x')$$

und setzen voraus, daß $\mu(\vartheta)$ für alle ϑ existiert. [Wir schreiben also hier $\mu(\vartheta)$ statt $g(\vartheta)$.]

Dann ist $\mu(\vartheta)$ also der Erwartungswert des Ergebnisses bei einem Teilversuch. Als naheliegenden Schätzer hat man:

Definition. $\bar{X} := \frac{1}{n} \sum_1^n X_m$, also $\bar{X}(x) = \bar{X}((x_1, \dots, x_n)) := \frac{1}{n} \sum_1^n x_m =: \bar{x}$, heißt das **Stichprobenmittel**.

Satz. Der Schätzer \bar{X} ist erwartungstreu für den Parameter $\mu(\vartheta)$.

Beweis. $E_{\vartheta}(\bar{X}) = \frac{1}{n} \sum_1^n E_{\vartheta}(X_m) = \mu(\vartheta)$ für alle $\vartheta \in \Theta$.

Nun interessieren wir uns für die Varianz als zweiten wichtigen Parameter

$$\sigma^2(\vartheta) := \text{Var}_{\vartheta}(X_m) = \sum_{x' \in \mathcal{X}'} (x' - \mu(\vartheta))^2 P_{\vartheta}(X_m = x') = \sum_{x' \in \mathcal{X}'} (x' - \mu(\vartheta))^2 P'_{\vartheta}(x')$$

und setzen die Existenz von $E_{\vartheta}(X_m^2)$ für alle $\vartheta \in \Theta$ voraus. [Hier spielt also $\sigma^2(\vartheta)$ die Rolle von $g(\vartheta)$.] Einen Schätzer für diesen Parameter liefert:

Definition. Die **Stichprobenstreuung** wird definiert als

$$(4.4) \quad S^2 := \frac{1}{n-1} \sum_{m=1}^n (X_m - \bar{X})^2, \quad \text{d.h. } S^2(x) := \frac{1}{n-1} \sum_{m=1}^n (x_m - \bar{x})^2.$$

Durch den Faktor $\frac{1}{n-1}$ anstelle von $\frac{1}{n}$ erreicht man gerade, daß gilt:

Satz. S^2 ein erwartungstreuer Schätzer ist für $\sigma^2(\vartheta)$.

Beweis: Es ist

$$(4.*) \quad Z_n := \sum_{m=1}^n (X_m - \bar{X})^2 = \sum_{m=1}^n (X_m - a)^2 - n \cdot (\bar{X} - a)^2 \quad \text{für alle } a \in \mathbb{R}$$

(vgl. Übung) sowie $\text{Var}_{\vartheta}(\bar{X}) = E_{\vartheta}((\bar{X} - \mu(\vartheta))^2) = \frac{1}{n} \sigma^2(\vartheta)$ (vgl. Bew. von Satz 3.16).
Damit ergibt sich mit $a = \mu := \mu(\vartheta)$, $\sigma^2 = \sigma^2(\vartheta)$:

$$E_{\vartheta}(Z_n) = E_{\vartheta}[\sum_{m=1}^n (X_m - \mu)^2] - n \cdot E_{\vartheta}[(\bar{X} - \mu)^2] = \sum_{m=1}^n \text{Var}_{\vartheta}(X_m) - n \cdot \text{Var}_{\vartheta}[\bar{X}]$$

$$= n \cdot \sigma^2 - \sigma^2 = (n-1) \cdot \sigma^2. \quad \text{Damit haben wir}$$

$$(4.5) \quad E_{\vartheta}(S^2) = \sigma^2(\vartheta).$$

4.6 Konsistenz

Es liege die gleiche Situation wie §4.3 vor.

Man kann auch das Gesetz der großen Zahl zur Rechtfertigung von $\bar{X} := \frac{1}{n} \sum_1^n X_i =: \bar{X}_n$ als Schätzer für den Erwartungswert $\mu(\vartheta)$ heranziehen. Gemäß Satz 3.16 haben wir nämlich

$$P_{\vartheta}(|\bar{X}_n - \mu(\vartheta)| \geq \varepsilon) \leq \frac{1}{n} \cdot \frac{1}{\varepsilon^2} \cdot \sigma^2(\vartheta).$$

Satz. Die Folge von Schätzern \bar{X}_n ist **konsistent** für $\mu(\vartheta)$, d.h.

$$P_{\vartheta}(|\bar{X}_n - \mu(\vartheta)| \geq \varepsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty \quad \forall \varepsilon > 0, \forall \vartheta \in \Theta,$$

falls $E_{\vartheta}(X_m^2)$ existiert für alle $\vartheta \in \Theta$.

Zum Nachweis der Konsistenz der Folge $S^2 =: S_n^2$ von Schätzern für $\sigma^2(\vartheta)$ ersetzen wir in den obigen Überlegungen X_m durch $f(X_m) = (X_m - \mu(\vartheta))^2$. Dann ist bei festem ϑ $f(X_1), \dots, f(X_n)$ wieder eine Folge von unabhängigen, identisch verteilten Zva, deren Varianz existiert, wenn wir noch die Existenz von $E_{\vartheta}(X_m^4)$ fordern. Somit bekommen wir wie oben mit

$$E_{\vartheta}[(X_m - \mu(\vartheta))^2] = \sigma^2(\vartheta) \quad \text{die Beziehung}$$

$$P_{\vartheta}(|\frac{1}{n} \sum_{m=1}^n (X_m - \mu(\vartheta))^2 - \sigma^2(\vartheta)| > \varepsilon) \rightarrow 0 \quad \text{für } n \rightarrow \infty \quad \forall \varepsilon > 0, \forall \vartheta \in \Theta.$$

Wir ersetzen jetzt in dieser Beziehung $\mu(\vartheta)$ durch einen Schätzer und schreiben mit (4.*) und $\sigma^2 > 0$ unter Benutzung der Dreiecksungleichung:

$$|\frac{1}{n} Z_n - \sigma^2| \leq |\frac{1}{n} \sum_{m=1}^n (X_m - \bar{X})^2 - \sigma^2| \leq |\frac{1}{n} \sum_{m=1}^n (X_m - a)^2 - \sigma^2| + (\bar{X} - a)^2$$

und erhalten :

$$\{|\frac{1}{n} Z_n - \sigma^2| > \varepsilon\} \subset \{|\frac{1}{n} \sum_{m=1}^n (X_m - a)^2 - \sigma^2| > \frac{1}{2}\varepsilon\} \cup \{(\bar{X} - a)^2 > \frac{1}{2}\varepsilon\}$$

und somit für $a = \mu(\vartheta)$

$$P_{\vartheta}(|\frac{1}{n}Z - \sigma^2(\vartheta)| > \varepsilon) \leq P_{\vartheta}(|\frac{1}{n} \sum_{m=1}^n (X_m - \mu(\vartheta))^2 - \sigma^2(\vartheta)| > \frac{1}{2}\varepsilon) + P_{\vartheta}(|\bar{X} - \mu(\vartheta)| > \sqrt{\frac{1}{2}\varepsilon}).$$

Dabei geht der zweite Term wieder gegen Null wegen der Konsistenz von \bar{X}_n ; aber auch der erste Term geht gegen Null, wie oben gezeigt wurde.

Satz. Die Folgen von Schätzern $\hat{\sigma}_n^2 := \frac{1}{n} \cdot Z_n$ bzw. $\hat{\sigma}_n^2 := \frac{1}{n-1} \cdot Z_n = S_n^2$ sind **konsistent** für $\sigma^2(\vartheta)$, d.h.

$$P_{\vartheta}(|\hat{\sigma}_n^2 - \sigma^2(\vartheta)| \geq \varepsilon) \rightarrow 0 \text{ für } n \rightarrow \infty \quad \forall \varepsilon > 0, \forall \vartheta \in \Theta,$$

falls $E_{\vartheta}(X_m^4)$ existiert für alle $\vartheta \in \Theta$.

Dabei unterscheidet sich S_n^2 von $\frac{1}{n} \cdot Z_n$ lediglich um den Faktor $\frac{n}{n-1}$. Dieser Teil des Beweises ist eine Übungsaufgabe.

Wir brauchen auch konsistente Schätzer für die **Standardabweichung** $\sigma(\vartheta) := \sqrt{\sigma^2(\vartheta)}$. Es gilt :

Korollar. Die Folgen von Schätzern (i) $\hat{\sigma}_n := \sqrt{S_n^2}$ bzw. (ii) $\hat{\sigma}_n := \sqrt{\frac{1}{n} \cdot Z_n}$ sind **konsistent** für $\sigma(\vartheta)$, d.h.

$$P_{\vartheta}(|\hat{\sigma}_n - \sigma(\vartheta)| \geq \varepsilon) \rightarrow 0 \text{ für } n \rightarrow \infty \quad \forall \varepsilon > 0, \forall \vartheta \in \Theta,$$

falls $E_{\vartheta}(X_m^4)$ existiert für alle $\vartheta \in \Theta$.

Beweis. Wir zeigen zunächst für $x, y, \varepsilon \geq 0$:

$$(**) \quad " |x-y| \geq \varepsilon " \Rightarrow " |x^2-y^2| \geq \varepsilon^2 ".$$

Wäre nämlich $x < \varepsilon$ und $y < \varepsilon$, so auch $|x-y| < \varepsilon$. Also gilt $x \geq \varepsilon$ oder $y \geq \varepsilon$ und damit $x+y \geq \varepsilon$. Es folgt $(x+y) \cdot |x-y| = |x^2-y^2| \geq \varepsilon^2$. Wir haben mit (**)

$$\{ |\hat{\sigma}_n - \sigma(\vartheta)| \geq \varepsilon \} \subset \{ |\hat{\sigma}_n^2 - \sigma^2(\vartheta)| \geq \varepsilon^2 \}$$

also nach dem obigen Satz:

$$P(|\hat{\sigma}_n - \sigma(\vartheta)| \geq \varepsilon) \leq P(|\hat{\sigma}_n^2 - \sigma^2(\vartheta)| \geq \varepsilon^2) \rightarrow 0 \quad . \quad \square$$

Beispiel.(Schätzung einer Erfolgswahrscheinlichkeit).

Es liege wieder Beispiel 4.2 vor. Wir haben dann: $\vartheta = p$, $\mathcal{X} = \{0,1\}^n$. Unter ϑ sind X_1, \dots, X_n unabhängig und identisch verteilt mit Werten in $\{0,1\}$ und $P_{\vartheta}(X_m = 1) = \vartheta \in (0,1) = \Theta$.

Es ist $\mu(\vartheta) = E_{\vartheta}(X_m) = \vartheta$. Dann ist \bar{X} also nicht nur ein Maximum-Likelihood-Schätzer, sondern auch ein erwartungstreuer Schätzer für $\mu(\vartheta) = \vartheta$.

Ferner bildet $\bar{X}_n = \bar{X}$ nach dem ersten Satz auch eine konsistente Folge von Schätzern für ϑ .

Für die Varianz haben wir $\sigma^2(\vartheta) = \text{Var}_{\vartheta}(X_m) = \vartheta \cdot (1-\vartheta)$.

Wegen $X_m \in \{0,1\}$ ergibt sich nach (4.*) mit $a = 0$:

$$\sum_{m=1}^n (X_m - \bar{X})^2 = \sum_{m=1}^n X_m^2 - n \cdot \bar{X}^2 = n \cdot \bar{X} - n \cdot \bar{X}^2; \text{ also}$$

$$\frac{1}{n} Z_n = \frac{1}{n} \sum_{m=1}^n (X_m - \bar{X})^2 = \bar{X} \cdot (1 - \bar{X}).$$

Damit ist $S_n^2 = \frac{n}{n-1} \frac{1}{n} Z_n = \frac{n}{n-1} \bar{X}_n \cdot (1 - \bar{X}_n)$ ein erwartungstreuer Schätzer für $\sigma^2(\vartheta) = \vartheta \cdot (1-\vartheta)$.

Ferner sind die Folgen von Schätzern

$$\bar{X}_n \cdot (1 - \bar{X}_n) \quad \text{bzw.} \quad S_n^2 = \frac{n}{n-1} \bar{X}_n \cdot (1 - \bar{X}_n) \quad \text{konsistent für } \sigma^2(\vartheta) = \vartheta \cdot (1-\vartheta), \text{ sowie}$$

$$\sqrt{\bar{X}_n \cdot (1 - \bar{X}_n)} \quad \text{bzw.} \quad S_n := \sqrt{S_n^2} = \sqrt{\frac{n}{n-1} \bar{X}_n \cdot (1 - \bar{X}_n)} \quad \text{konsistent für } \sigma(\vartheta) = \sqrt{\vartheta \cdot (1-\vartheta)}.$$

[Dies hätte man auch aus der Konsistenz der Folge \bar{X}_n für ϑ und einem zusätzlichen Stetigkeitsargument herleiten können.]

Bei einem **Zahlenbeispiel** sei $n = 14$, $x = (x_1, \dots, x_{14}) = (1, 0, 1, 1, \dots)$ (7 Einsen, 7 Nullen).

$$\text{Dann ist } \bar{X}(x) = \frac{7}{14} = \frac{1}{2}, \quad \text{und}$$

$$\frac{1}{n} Z_n = \bar{X}(x) \cdot (1 - \bar{X}(x)) = \frac{1}{2} \cdot (1 - \frac{1}{2}) = 0,25 \quad \text{sowie} \quad S^2(x) = \frac{14}{13} \frac{1}{2} \cdot (1 - \frac{1}{2}) = \frac{14}{52} = 0,27. \quad \square$$

Beispiel. (2-parametrische negative Binomialverteilung).

Es liege wieder ein Einstichprobenproblem vor mit: $\mathfrak{X} = \mathbb{N}_0^n$, $\vartheta = (r, p) \in (0, \infty) \times (0, 1) = \Theta$.

Bei festem ϑ sind X_1, \dots, X_n unabhängig und identisch verteilt mit Werten in \mathbb{N}_0 und

$$P_{\vartheta}(X_m = k) = \frac{(r+k-1) \cdot (r+k-2) \cdot \dots \cdot r}{k!} \cdot p^r \cdot (1-p)^k, \quad k \in \mathbb{N}_0.$$

Im Fall $r = 1$ haben wir gerade eine geometrische Verteilung.

Dann sind $g_1(\vartheta) = r$ sowie $g_2(\vartheta) = p$ mögliche Parameter. Weitere wichtige Parameter sind

$$\mu(\vartheta) = E_{\vartheta}(X_m) = \frac{r \cdot (1-p)}{p}, \quad \sigma^2(\vartheta) = \frac{r \cdot (1-p)}{p^2}.$$

[Der Beweis verläuft wie bei der Poissonverteilung oder der hypergeometrischen Verteilung.]

Als Zahlenbeispiel betrachten wir $n = 7$, $x = (x_1, \dots, x_7) = (24, 7, 17, 9, 20, 11, 18)$.

Dann ist $\bar{X}(x) = \bar{x} = 15,14$ Schätzung für $\mu(\vartheta)$,

$$Z(x) := \sum_{m=1}^n (x_m - \bar{x})^2 = \sum_{m=1}^n (x_m - a)^2 - n \cdot (\bar{x} - a)^2 \quad \text{mit } a = 10, \text{ also}$$

$$Z(x) = 14^2 + 3^2 + 7^2 + 1^2 + 10^2 + 1^2 + 8^2 - 7 \cdot (5,14)^2 = 420 - 184,93 = 235,07,$$

$$\frac{1}{n} Z(x) = 33,58 \quad \text{und} \quad S^2(x) = \frac{1}{n-1} Z(x) = 39,178 \quad \text{Schätzungen für } \sigma^2(\vartheta),$$

$$\sqrt{\frac{1}{n} Z(x)} = 5,79 \quad \text{und} \quad S(x) = \sqrt{S^2(x)} = 6,26 \quad \text{Schätzungen für } \sigma(\vartheta).$$

Aus der Beziehung $p = \mu/\sigma^2$ und $r = \mu^2/(\sigma^2 - \mu)$ erhält man naheliegende Schätzer für p und r , die aber nicht erwartungstreu sein müssen. \square

4.7 Konfidenzintervalle

In diesem Abschnitt soll die Angabe eines Schätzers \hat{g} von $g(\vartheta)$ dahingehend präzisiert werden, daß man eine Angabe $g(\vartheta) = \hat{g}(x) \pm \varepsilon$ macht in der Form:

$$P_{\vartheta}(|\hat{g}(x) - g(\vartheta)| \leq \varepsilon) \geq 1 - \alpha \quad \text{für alle } \vartheta \in \Theta.$$

Setzt man $C(x) = [\hat{g}_-(x), \hat{g}_+(x)]$, so bedeutet dies:

$$(4.13) \quad P_{\vartheta}(\{x \in \mathcal{X}, C(x) \ni g(\vartheta)\}) \geq 1 - \alpha \quad \text{für alle } \vartheta \in \Theta.$$

Allgemeiner wird man das folgende Konzept verwenden:

Definition. Sei $g(\vartheta)$ ein reeller Parameter. Ist $\{C(x), x \in \mathcal{X}\}$ eine Familie von Intervallen und gilt (4.13) für ein $0 < \alpha < 1$ (etwa $\alpha = 0,05$ oder $0,01$), so heißt $C(x)$ **Konfidenzintervall** für $g(\vartheta)$ zum **Konfidenzniveau** $1 - \alpha$ (zur Stichprobe x).

Ein Konfidenzintervall überdeckt also den richtigen Parameter $g(\vartheta)$ mit großer W. unter allen ϑ .

Zum Beispiel kann das Konfidenzintervall C die Form haben $C(x) = [\hat{g}_-(x), \hat{g}_+(x)]$. Dabei sollen $\hat{g}_{\pm}(x)$ den Parameter $g(\vartheta)$ über- bzw. unterschätzen. Ist $\hat{g}(x)$ ein Schätzer, so kann etwa $\hat{g}_{\pm}(x) = \hat{g}(x) \pm \varepsilon$ gewählt werden. Wir fragen jetzt nach der Abhängigkeit von ε und α .

Hier werden insbesondere Konfidenzintervalle für die Erfolgs-W. $p = \vartheta$ [vgl. Beispiel 4.2] ermittelt werden, die natürlich möglichst klein sein sollen. So ist $C(x) = [0,1]$ immer ein Konfidenzintervall zum Niveau 1, aber unbrauchbar. Dazu betrachten wir die folgende Situation:

Situation (1): Wir betrachten ein Einstichprobenproblem. Dabei seien X_1, \dots, X_n unabhängige, identisch verteilte $\{0,1\}$ -wertige Zva mit $P_{\vartheta}(X_m = 1) = \vartheta$ unter dem Parameter $\vartheta \in \Theta = [0,1]$.

Dann hat $X_1 + \dots + X_n$ eine $b_{n,\vartheta}$ -Verteilung unter ϑ [vgl. (2.8)].

Wir schreiben wieder $\bar{X} = \bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$.

Bemerkung. In der obigen Situation (1) gilt:

$$C = [\bar{X} - \varepsilon, \bar{X} + \varepsilon], \quad \text{d.h. } C(x) = [\bar{x} - \varepsilon, \bar{x} + \varepsilon]$$

ist ein Konfidenzintervall zum Konfidenzniveau $(1 - \frac{1}{4n \cdot \varepsilon^2})^+$.

Diese Aussage ist noch sehr grob und ergibt sich aus der Borelschen Gesetz der großen Zahl von §3.6 [oder der Konsistenz von \bar{X}_n für ϑ in § 4.6].

Demnach ist für jedes feste ϑ : $P_{\vartheta}(|\bar{X} - \vartheta| \geq \varepsilon) \leq \frac{1}{n \cdot \varepsilon^2} \vartheta \cdot (1 - \vartheta) \leq \frac{1}{4n \cdot \varepsilon^2}$. Es folgt:

$$P_{\vartheta}(\{x \in \mathcal{X}; C(x) \ni \vartheta\}) = P_{\vartheta}(|\bar{X} - \vartheta| \leq \varepsilon) \geq 1 - P_{\vartheta}(|\bar{X} - \vartheta| > \varepsilon) \geq 1 - \frac{1}{4n \cdot \varepsilon^2}. \quad \square$$

Schärfere Aussagen bekommt man **für große n** durch den zentralen Grenzwertsatz. Aus dem Satz von de Moivre–Laplace erhalten wir für festes ϑ und $\sigma(\vartheta) := \sqrt{\vartheta \cdot (1-\vartheta)} \leq \frac{1}{2}$:

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{\vartheta}(-c \leq \frac{\sqrt{n}}{\sigma(\vartheta)} [\bar{X}_n - \vartheta] \leq c) &= \Phi(c) - \Phi(-c) = 2 \cdot \Phi(c) - 1 \\ &= \lim_{n \rightarrow \infty} P_{\vartheta} \left[\left[\bar{X}_n - \frac{c \cdot \sigma(\vartheta)}{\sqrt{n}}, \bar{X}_n + \frac{c \cdot \sigma(\vartheta)}{\sqrt{n}} \right] \ni \vartheta \right] \\ &\leq \lim_{n \rightarrow \infty} P_{\vartheta} \left[\left[\bar{X}_n - \frac{c}{2\sqrt{n}}, \bar{X}_n + \frac{c}{2\sqrt{n}} \right] \ni \vartheta \right]. \end{aligned}$$

Zu gegebenem Konfidenzniveau $1 - \alpha$ wählen wir c so, daß gilt:

$$2 \cdot \Phi(c) - 1 = 1 - \alpha \Leftrightarrow \Phi(c) = 1 - \frac{1}{2}\alpha \Leftrightarrow c = \Phi^{-1}(1 - \frac{1}{2}\alpha).$$

(man sagt: c ist das $\frac{1}{2}\alpha$ -Fraktile der Normalverteilung).

Dieses Ergebnis kann nun so formuliert werden.

Satz. Liegt die Situation (1) vor und ist c das $\frac{1}{2}\alpha$ -Fraktile der Normalverteilung, so ist

$$C_n = \left[\bar{X}_n - \frac{c}{2 \cdot \sqrt{n}}, \bar{X}_n + \frac{c}{2 \cdot \sqrt{n}} \right] \text{ d.h. } C_n(x) = \left[\bar{x}_n - \frac{c}{2 \cdot \sqrt{n}}, \bar{x}_n + \frac{c}{2 \cdot \sqrt{n}} \right]$$

ein Konfidenzintervall für ϑ , das asymptotisch das Konfidenzniveau $1 - \alpha$ hat, d.h.

$$\lim_{n \rightarrow \infty} P_{\vartheta}(C_n \ni \vartheta) = \lim_{n \rightarrow \infty} P_{\vartheta}(\{x \in \mathcal{X}; C_n(x) \ni \vartheta\}) \geq 1 - \alpha \text{ für alle } \vartheta \in \Theta.$$

Anwendung: Wahlvorhersage. Es soll der Prozentsatz der Wähler einer Partei A geschätzt werden. Es werden nacheinander n Wähler zufällig ausgewählt und befragt. Wir nehmen dabei der Einfachheit halber an, daß die gleiche Person mehrmals befragt werden kann (Ziehung mit Zurücklegen). Ist die Gesamtzahl der Wähler groß, so spielt dies keine große Rolle.

Es sei S_n die Anzahl der befragten Personen, die sich für die Partei A entscheiden. Dann benutzt man $\bar{X}_n = S_n/n$ als Schätzer für die W. ϑ , daß eine zufällig ausgewählte Person ein Wähler von Partei A ist. Die Angabe des Schätzers \bar{X}_n wird jetzt ergänzt durch die Angabe eines Konfidenzintervalls. Es soll n so groß sein, daß gilt:

$$P_{\vartheta}(C \ni \vartheta) \approx 1 - \alpha := 0,95 \text{ für das Konfidenzintervall } C = [\bar{X}_n - 0,01, \bar{X}_n + 0,01].$$

Verwendet man die grobe Abschätzung aus der Bemerkung, so muß man n so wählen, daß gilt:

$$\frac{1}{4n \cdot \varepsilon^2} = \frac{10\,000}{4n} = 0,05 \Leftrightarrow n = 50\,000.$$

Nun soll das Ergebnis des obigen Satzes angewendet werden. Dann haben wir

$$1 - \Phi(c) = \frac{1}{2}\alpha \Leftrightarrow \Phi(c) = 1 - \frac{1}{2}\alpha = 0,975.$$

Aus einer Tabelle entnehmen wir $\Phi(1,96) = 0,975$, d.h. $c = 1,96$. Also müssen wir n so wählen, daß gilt: $\frac{1,96}{2 \cdot \sqrt{n}} \approx 0,01$, d.h. $n = \frac{1}{4} \cdot (1,96)^2 \approx 9600$. Also reichen 9600 Befragungen aus.

Bei lediglich 1000 Befragungen erhält man ein Konfidenzintervall der Größe

$$C = [\bar{X}_n - 0,03, \bar{X}_n + 0,03]. \quad \square$$

In der bisherigen Vorgehensweise ist noch unbefriedigend, daß wir $\sigma^2 = \vartheta \cdot (1-\vartheta)$ grob durch $1/4$ abgeschätzt haben. Man bekommt schärfere Aussagen, wenn man die unbekannte Varianz σ^2 durch einen Schätzer ersetzt, statt die grobe Abschätzung $\sigma^2 = \vartheta \cdot (1-\vartheta) \leq 1/4$ zu verwenden. Dazu kann man eine Variante des zentralen Grenzwertsatz verwenden, in der σ durch einen Schätzer ersetzt wird.

Wir behandeln erst den allgemeineren Fall, wo der Erwartungswert μ geschätzt werden soll.

Situation (2): Wir betrachten ein Einstichprobenproblem. Dabei seien X_1, \dots, X_n unabhängige, identisch verteilte \mathcal{X} -wertige Zva mit W-Funktion $P_\vartheta(X_m = x') = P'_\vartheta(x')$ (d.h. P'_ϑ ist die Verteilung von X_m) unter dem Parameter $\vartheta \in \Theta$. Es existiere wieder $E_\vartheta(X_m^4)$ für all $\vartheta \in \Theta$.

Dann bildet $\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{m=1}^n X_m$ eine konsistente Folge von Schätzern für $\mu(\vartheta) = E_\vartheta(X_m)$.

Sei ferner $\hat{\sigma}_n$ eine konsistente Folge von Schätzern für $\sigma(\vartheta) = \sqrt{\text{Var}_\vartheta(X_m)}$, d.h.

$$P_\vartheta(|\hat{\sigma}_n - \sigma(\vartheta)| \geq \varepsilon) \rightarrow 0 \text{ für } n \rightarrow \infty \quad \forall \varepsilon > 0, \forall \vartheta \in \Theta.$$

In dieser Situation gilt nach dem zentralen Grenzwertsatz:

$$\lim_{n \rightarrow \infty} P_\vartheta(a \leq \frac{\sqrt{n}}{\sigma(\vartheta)} (\bar{X}_n - \mu(\vartheta)) \leq b) = \Phi(b) - \Phi(a) \text{ für alle } -\infty < a < b < \infty.$$

Zusatz zum zentralen Grenzwertsatz. Es liege die Situation (2) vor. Dann gilt

[mit $\Phi(-\infty) = 0, \Phi(\infty) = 1$] für alle $\vartheta \in \Theta$:

$$\lim_{n \rightarrow \infty} P_\vartheta(a \leq \frac{\sqrt{n}}{\hat{\sigma}_n} (\bar{X}_n - \mu(\vartheta)) \leq b) = \Phi(b) - \Phi(a) \text{ für alle } -\infty \leq a < b \leq \infty.$$

Der Beweis folgt relativ einfach aus dem zentralen Grenzwertsatz, soll aber hier nicht gebracht werden.

Konfidenzintervalle für den Erwartungswert.

Gegeben sei die Situation (2). Wir interessieren uns für den Erwartungswert als Parameter

$$\mu(\vartheta) := E_{\vartheta}(X_m)$$

Dann bilden

$$(i) \quad \hat{\sigma}_n^2 = S_n^2 := \frac{1}{n-1} \sum_{m=1}^n (X_m - \bar{X}_n)^2$$

$$(ii) \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{m=1}^n (X_m - \bar{X}_n)^2$$

beides konsistente Folgen von Schätzern für $\sigma^2(\vartheta)$ gemäß § 4.6.

Als Korollar dazu haben wir erhalten, daß in beiden Fällen $\sigma_n = \sqrt{\hat{\sigma}_n^2}$ eine konsistente Folge von Schätzern für $\sigma(\vartheta) = \sqrt{\sigma^2(\vartheta)}$ bildet. Nach dem Zusatz zum zentralen Grenzwertsatz gilt nun

$$\lim_{n \rightarrow \infty} P_{\vartheta} \left(-c \leq \frac{\sqrt{n}}{\hat{\sigma}_n} [\bar{X}_n - \mu(\vartheta)] \leq c \right) = \Phi(c) - \Phi(-c) = 2 \cdot \Phi(c) - 1.$$

Zu gegebenem Konfidenzniveau $1 - \alpha$ wählen wir nun wieder c so, daß gilt:

$$2 \cdot \Phi(c) - 1 = 1 - \alpha \Leftrightarrow \Phi(c) = 1 - \frac{1}{2}\alpha \Leftrightarrow c = \Phi^{-1}(1 - \frac{1}{2}\alpha).$$

Dann haben wir $\lim_{n \rightarrow \infty} P_{\vartheta} \left(-c \leq \frac{\sqrt{n}}{\hat{\sigma}_n} [\bar{X}_n - \mu(\vartheta)] \leq c \right) = 1 - \alpha$, d.h.

$$\lim_{n \rightarrow \infty} P_{\vartheta} \left([\bar{X}_n - \varepsilon_n, \bar{X}_n + \varepsilon_n] \ni \mu(\vartheta) \right) = 1 - \alpha \quad \text{mit } \varepsilon_n := c \cdot \hat{\sigma}_n / \sqrt{n}.$$

Als **Resultat** erhalten wir:

Satz. Sei $\hat{\sigma}_n$ wie in (i) oder (ii) und c das $\frac{1}{2}\alpha$ -Fraktil der Normalverteilung.

Dann ist in der Situation (2)

$$C = [\bar{X}_n - c \cdot \hat{\sigma}_n / \sqrt{n}, \bar{X}_n + c \cdot \hat{\sigma}_n / \sqrt{n}]$$

ein Konfidenzintervall für den Parameter $\mu(\vartheta)$, das asymptotisch das Konfidenzniveau $1 - \alpha$ hat, d.h. $\lim_{n \rightarrow \infty} P_{\vartheta}(C_n \ni \mu(\vartheta)) \geq 1 - \alpha \quad \forall \vartheta \in \Theta$.

In beiden Fällen wird also mit wachsendem n das Konfidenzintervall kleiner und mit wachsendem Niveau $1 - \alpha$ größer. Dabei hängt ε_n von $\hat{\sigma}_n$ ab und damit die Intervalllänge des Konfidenzintervalls von der Stichprobe! Vor der Stichprobenentnahme weiß man also nicht, wie klein das Konfidenzintervall ausfallen wird.

Der Satz gestattet natürlich auch eine Anwendung auf Konfidenzintervalle für die Erfolgs-W, also auf die spezielle Situation (1), denn dort ist ja $E_{\vartheta}[X_m] = \mu(\vartheta) = \vartheta$.

§ 6 Tests

Wie bei Schätzproblemen und wie immer in der Statistik werden wir eine Familie von W -Maßen betrachten, unter denen ein W -maß sein soll, das das Experiment richtig beschreibt, das sogenannte 'wahre' W -maß. Zur Beschreibung eines Testproblems benötigen wir wieder:

- eine (nichtleere, höchstens abzählbare) Menge \mathcal{X} , den **Stichprobenraum**,
- eine Familie $\{P_{\vartheta}, \vartheta \in \Theta\}$ von W -maßen auf \mathcal{X} , (die **Verteilungsannahme**),
 Θ ist der **Parameterraum**.

Wir werden uns wieder auf **Einstichprobenprobleme** beschränken. Dann hat man also:

$\mathcal{X} = \mathcal{X}' \times \dots \times \mathcal{X}' \ni x = (x_1, \dots, x_n)$ mit den **Stichprobenumfang** n ;

$X_m(x) := x_m$ beschreibe das Ergebnis des m -ten Teilversuchs.

Unter jedem festem ϑ sind X_1, \dots, X_n unabhängige, identisch verteilte Zva mit der W -Funktion $P_{\vartheta}(X_m = x') =: P_{\vartheta}'(x')$.

Im Gegensatz zu Schätzproblemen ist man jetzt nicht an dem genauen Wert von ϑ oder $g(\vartheta)$ interessiert, sondern nur daran, ob ϑ eine bestimmte Eigenschaft hat oder nicht. Mathematisch wird dies dadurch ausgedrückt, ob für eine bestimmte Teilmenge $H \subset \Theta$ gilt: $\vartheta \in H$ oder $\vartheta \notin H$.

Definition. Ein **Testproblem** ist gegeben durch eine Teilmenge H von Θ , der **Hypothese** $H \subset \Theta$.

Ziel ist es, das **Verwerfen der Hypothese statistisch abzusichern**. Die Annahme der Hypothese wird dabei lediglich als Stimmenthaltung zu interpretieren sein.

Beispiel. (*Testen einer Erfolgswahrscheinlichkeit*).

Beschreibt $\vartheta \in \Theta = [0,1]$ oder $\Theta = (0,1)$ die Erfolgs- $W.$, so ist man interessiert an:

$$H = \{\vartheta \in \Theta; \vartheta \leq \vartheta_0\}.$$

Ist ϑ bzw. ϑ_0 die $W.$, daß ein neu entwickeltes bzw. etabliertes Medikament wirksam ist, so interessiert, ob man die Hypothese $\vartheta \in H$ verwerfen kann, ob man also das neue Medikament im Großen herstellen kann.. \square

Ein Test soll eine Entscheidungsregel sein, die auf Grund einer Stichprobe $x \in \mathcal{X}$ festlegt, ob man die Hypothese verwirft oder nicht.

Definition. Ein **Test** wird beschrieben durch Angabe des **Verwerfungsbereichs** $R \subset \mathcal{X}$, verbunden mit der Vorschrift: Liegt die Stichprobe x in R , so wird die Hypothese verworfen. Liegt die Stichprobe x nicht in R , so wird die Hypothese nicht verworfen.

Innerhalb des gewählten Modells interessiert der folgende Fehler (der sogenannte **Fehler 1. Art**):

Die Hypothese wird verworfen, obwohl sie richtig ist, d.h. obwohl der Parameter ϑ , der das Experiment richtig beschreibt, – der sogenannte 'wahre Parameter' – in H liegt.

Ein typischer Verwerfungsbereich R hat die Gestalt

$$(+)\quad R = \{x \in \mathfrak{X}; T(x) > t\} \text{ oder } R = \{x \in \mathfrak{X}; |T(x)| > t\}$$

mit einer Funktion T auf \mathfrak{X} .

Bemerkung. Die Funktion T auf \mathfrak{X} nennt man (Test-) **Statistik** oder **Prüfgröße**. In der Situation (+) heißt t auch **kritischer Wert**.

Man kann sich eine Statistik T als eine Zusammenfassung der durch x gegebenen Daten vorstellen. Dies steht im Einklang mit der herkömmlichen Vorstellung von einer Statistik.

Beispiel. (Fortsetzung) Es liegt ein Einstichprobenproblem vor; dabei beschreiben die $\{0,1\}$ -wertigen Zva X_1, \dots, X_n die Ergebnisse der Einzelversuche. Dann wird die Entscheidung abhängen von der Anzahl $S := X_1 + \dots + X_n$ der Erfolge.

$T(x) = S(x) = x_1 + \dots + x_n$ für $x = (x_1, \dots, x_n)$ ist dann also eine Statistik im obigen Sinne. Für den Verwerfungsbereich R der Hypothese $H = \{\vartheta \in \Theta; \vartheta \leq \vartheta_0\}$ wird man den Ansatz machen:

$$R_t = \{x; S(x) > t\}$$

mit einem geeignet zu wählenden t . \square

In Anwendungen wird die Hypothese einer "**etablierten Theorie**" entsprechen – im Gegensatz zu "**neuen Effekten**" –, die man nur dann verwirft, wenn eindeutige Hinweise vorliegen. Die Verwerfung der etablierten Meinung hat in der Regel weitreichende Konsequenzen, so daß man den Fehler (1. Art) klein halten möchte.

Beispiel. (Fortsetzung). Man wird ein neues Medikament einem gebräuchlichen nur dann vorziehen und in Serienproduktion herstellen, wenn wirklich die Daten deutlich dafür sprechen.

Der Fehler, ein weniger wirksames Medikament auf den Markt zu werfen, hätte schwerwiegende Folgen. \square

Definition. Es liege ein Test mit Verwerfungsbereich R vor. Dann heißen

$$P_{\vartheta}(R), \vartheta \in H, \text{ Fehlerwahrscheinlichkeiten (1. Art).}$$

Die Tatsache, daß der Fehler 1. Art schwerwiegend ist, trägt man dadurch Rechnung, daß man eine kleine Fehlerschranke $0 < \alpha < 1$ vorgibt und fordert, daß die Fehler-W. (1. Art) durch α beschränkt ist.

Definition. Ein Test mit Verwerfungsbereich R heißt **Test zum Niveau α** (z.N. α), wenn gilt:

$$P_{\vartheta}(R) \leq \alpha \text{ für alle } \vartheta \in H.$$

Bei einem Test z.N. α ist es unwahrscheinlich, daß unter H eine Stichprobe x in R liegt, sodaß es im Fall $x \in R$ naheliegender ist, H zu verwerfen. Im Fall $x \notin R$ trifft man keine echte Entscheidung und übt gewissermaßen Stimmhaltung, denn $P_{\vartheta}(R^c)$ kann auch für $\vartheta \notin H$ groß sein

Man läßt jetzt nur Tests z.N. α zu. Üblich sind Werte $\alpha = 0,05$, $\alpha = 0,01$.

Beispiel (Fortsetzung).

Man spricht auch von einem **einseitigen Testproblem**, weil die Werte von $H = \{\vartheta \in \Theta; \vartheta \leq \vartheta_0\}$ auf der einen Seite von ϑ_0 liegen. Es beschreibt S wieder die Anzahl der Erfolge. Betrachten wir nun Tests mit Verwerfungsbereich R der Form:

$$R_t = \{x; S(x) > t\} \quad \text{mit } t = -1, 0, 1, \dots, n.$$

Dann gilt:

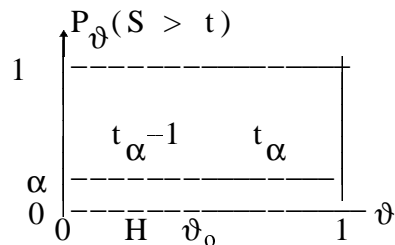
- S hat eine $b_{n, \vartheta}$ -Verteilung unter ϑ ,
- $P_{\vartheta}(R_t) = P_{\vartheta}(S > t) := \sum_{i=t+1}^n \binom{n}{i} \vartheta^i (1-\vartheta)^{n-i}$,
- $P_{\vartheta}(S > t) \geq P_{\vartheta}(S > t+1)$, $\vartheta \in \Theta$,
- $\frac{d}{d\vartheta} P_{\vartheta}(S > t) > 0$, also ist $P_{\vartheta}(S > t)$ eine monotone Funktion von ϑ auf $(0, 1)$ für $0 \leq t \leq n-1$.
- R_t beschreibt einen Test z.N. $\alpha \Leftrightarrow P_{\vartheta}(S > t) \leq \alpha \quad \forall \vartheta \leq \vartheta_0 \Leftrightarrow P_{\vartheta_0}(S > t) \leq \alpha$.

$$\Leftrightarrow t \geq t_{\alpha} \quad \text{mit } t_{\alpha} := \min \{t; P_{\vartheta_0}(S > t) \leq \alpha\}$$

$$\Leftrightarrow t \geq t_{\alpha} \quad \text{mit } P_{\vartheta_0}(S > t_{\alpha}) \leq \alpha < P_{\vartheta_0}(S > t_{\alpha} - 1) = P_{\vartheta_0}(S \geq t_{\alpha}).$$

Dabei heißt $t_{\alpha} =: b_{n, \vartheta_0, \alpha}$ **α -Fraktile** der b_{n, ϑ_0} -Verteilung mit

$$(\#) \quad P_{\vartheta_0}(S > t_{\alpha}) \leq \alpha < P_{\vartheta_0}(S \geq t_{\alpha}).$$



Zeichne den Verlauf von $P_{\vartheta_0}(S > t_{\alpha})$ und $P_{\vartheta_0}(S > t_{\alpha} - 1)$ ein, sodaß (#) gilt.

Wegen der Minimumeigenschaft von t_{α} gilt: $\emptyset = R_n \subset R_t \subset R_{t_{\alpha}}$, wenn R_t einen Test zum Niveau α beschreibt. Der Test mit dem Verwerfungsbereich R_n verwirft die Hypothese nie, also kann auch nie ein Fehler (1. Art) gemacht werden. Solche Tests sind natürlich nicht interessant. Wenn man die Hypothese ablehnen darf, so soll dies auch geschehen, d.h. R_t soll unter den Verwerfungsbereichen von Tests zum Niveau α maximal sein, die vorgegebene Fehlerschranke α soll voll ausgeschöpft werden.

Satz. Es liege die Situation des Beispiels vor mit dem durch $H = \{\vartheta \in \Theta; \vartheta \leq \vartheta_0\}$ gegebenen einseitigen Testproblem. Dann wählt man den Test z.N. α mit dem Verwerfungsbereich

$$R_{t_{\alpha}} = \{x; S(x) > t_{\alpha}\}. \quad \text{Dabei ist } t_{\alpha} \text{ das } \alpha\text{-Fraktile der } b_{n, \vartheta_0}\text{-Verteilung.}$$

Zahlenbeispiel.

Es sei $n = 14$, $\vartheta_0 = \frac{1}{4}$, $\alpha = 0,01$ oder $\alpha = 0,0103$ oder $\alpha = 0,05$.

t	$P_{\vartheta_0}(S > t)$
-1	1
0	0,9822
1	0,8922
2	0,7189
3	0,4789
4	0,2585
5	0,1117
6	0,0383
7	0,0103 d.h. $P_{\vartheta_0}(S > 7) = 0,0103$
8	0,0022
9	0,0003
10	0,0000

In diesem Zahlenbeispiel ist für $\alpha = 0,01$ also $t_\alpha = 8$. Hier wird man allerdings $\alpha = 0,0103$ wählen und erhält: $t_\alpha = 7$. Bei unwesentlicher Vergrößerung von α erhält man einen größeren Verwerfungsbereich. Bei Benutzung des Tests mit Verwerfungsbereich R_7 könnte man sich also erst dann für eine Verwerfung der Hypothese $H: \vartheta \leq \frac{1}{4}$ entscheiden, wenn mehr als 7 Erfolge bei $n = 14$ Einzelexperimenten vorliegen, also mehr als die Hälfte aller Einzelversuche erfolgreich verlaufen. Daß $t_\alpha = 7$ so groß ist, liegt daran, daß n klein ist relativ zum Wert von α . Für $\alpha = 0,05$ erhält man $t_\alpha = 6$. Die Antwort nach Durchführung des Tests lautet etwa:

Bei $x \in R$: Es ist z.N. α **statistisch gesichert**, daß für den wahren Parameter ϑ gilt: $\vartheta > \frac{1}{4}$.

Bei $x \notin R$: Es ist z.N. α **nicht statistisch gesichert**, daß $\vartheta > \frac{1}{4}$ gilt.

Die Antwort darf hier **nicht** lauten: Es ist z.N. α statistisch gesichert, daß $\vartheta \leq \frac{1}{4}$ gilt. Das wäre bei 7 Erfolgen von 14 Einzelversuchen auch merkwürdig.

Große n.

Für große n kann man den Satz von de Moivre–Laplace heranziehen. Schreiben wir $\bar{X}_n = \frac{1}{n} S$ für die mittlere Anzahl der Erfolge, $\sigma_0^2 = \vartheta_0 \cdot (1 - \vartheta_0)$, so erhalten wir für den Verwerfungsbereich

$$R_t = \{x; S(x) > t\} = \left\{x; \frac{\sqrt{n}}{\sigma_0} (\bar{X}_n(x) - \vartheta_0) > c\right\} \quad \text{für } c = \frac{\sqrt{n}}{\sigma_0} (t - \vartheta_0).$$

Nun ist $P_{\vartheta_0} \left(\frac{\sqrt{n}}{\sigma_0} (\bar{X}_n(x) - \vartheta_0) \leq c \right) \approx \Phi(c) = 1 - \alpha$. Wählt man $R^* = \left\{x, \frac{\sqrt{n}}{\sigma_0} (\bar{X}_n(x) - \vartheta_0) > c\right\}$, so gilt $P_{\vartheta_0}(R^*) \leq P_{\vartheta_0}(R^*) \approx 1 - \Phi(c) = \alpha$, $\vartheta \in H$, für große n , wenn c das α -Fraktile der Normalverteilung ist. Also definiert der Verwerfungsbereich R^* einen Test, der asymptotisch ein Test z.N. α ist.

Korollar. In der Situation des Satzes wählt man für große n den Test mit dem Verwerfungsbereich

$$R^* = \{x; \bar{X}_n(x) > \vartheta_0 + c \cdot \sqrt{\frac{1}{n} \cdot \vartheta_0 \cdot (1 - \vartheta_0)}\};$$

dabei ist c das α -Fraktile der Normalverteilung. Für große n ist der Test näherungsweise ein Test z.N. α , der die vorgegebene Fehlerschranke α ausschöpft.

Im Gegensatz zu unserem Zahlenbeispiel mit $n = 14$, braucht für große n die mittlere Anzahl \bar{X}_n der Erfolge nicht viel größer als ϑ_0 zu sein, damit z.N. α statistisch gesichert ist, daß $\vartheta > \vartheta_0$ gilt.

Wir wollen jetzt noch ein so genanntes **zweiseitiges Testproblem** anschauen. Wir wählen dazu

$$H_2 = \{\vartheta_0\}.$$

Im Fall $\vartheta_0 = \frac{1}{2}$, könnte man also testen, ob eine echte (symmetrische) Münze vorliegt oder nicht. Hier wollen wir nur die Situation für große n betrachten. Wie in § 4.7 erhalten wir aus dem Satz

von de Moivre–Laplace mit $\sigma_0 = \sqrt{\vartheta_0 \cdot (1 - \vartheta_0)}$:

$$P_{\vartheta_0}(-c \leq \frac{\sqrt{n}}{\sigma_0} \cdot [\bar{X}_n - \vartheta_0] \leq c) \approx \Phi(c) - \Phi(-c) = 2 \cdot \Phi(c) - 1 = 1 - \alpha \quad (\text{für } n \rightarrow \infty),$$

wenn c das $\frac{1}{2}\alpha$ -Fraktile der Normalverteilung ist. Es folgt

$$P_{\vartheta_0}(\frac{\sqrt{n}}{\sigma_0} \cdot |\bar{X}_n - \vartheta_0| > c) = P_{\vartheta_0}(|\bar{X}_n - \vartheta_0| > c \cdot \sigma_0 / \sqrt{n}) \approx \alpha \quad (\text{für } n \rightarrow \infty).$$

Definieren wir nun den Verwerfungsbereich R_2^* gemäß

$$R_2^* := \{x \in \mathcal{X}; |\bar{X}_n - \vartheta_0| > c \cdot \sqrt{\frac{1}{n} \cdot \vartheta_0 \cdot (1 - \vartheta_0)}\},$$

so haben wir $P_{\vartheta_0}(R_2^*) \approx \alpha \Leftrightarrow P_{\vartheta}(R_2^*) \approx \alpha \quad \forall \vartheta \in H_2$.

Damit definiert R_2^* einen Test, der asymptotisch ein Test z.N. α ist und die vorgegebene Fehlerschranke α voll ausschöpft.

Satz. Es liege die Situation des Beispiels vor mit dem durch $H_2 = \{\vartheta_0\}$ gegebenen zweiseitigen Testproblem. Dann wählt man für großes n den Test mit dem Verwerfungsbereich

$$R_2^* := \{x \in \mathcal{X}; |\bar{X}_n - \vartheta_0| > c \cdot \sqrt{\frac{1}{n} \cdot \vartheta_0 \cdot (1 - \vartheta_0)}\};$$

dabei ist c das $\frac{1}{2}\alpha$ -Fraktile der Normalverteilung. Für große n ist der Test näherungsweise ein Test z.N. α , der die vorgegebene Fehlerschranke α ausschöpft.

Bezeichnet man ϑ_0 als Sollwert, so verwirft man die Hypothese, daß der wahre Parameter ϑ gleich dem Sollwert ist, wenn der Schätzer \bar{X}_n für ϑ stark vom Sollwert abweicht. Was hier stark bedeutet, wird im Satz präzisiert.

Kapitel III Markow–Ketten.

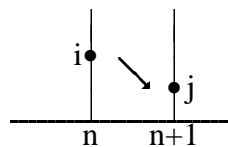
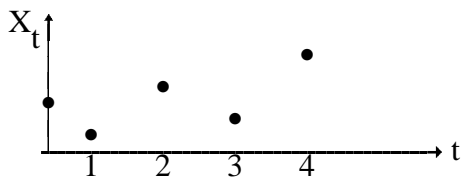
§14 Die Markow–Eigenschaft.

Sei I eine endliche oder abzählbare Menge.

Definition. 14.1. Sei (Ω, P) ein W –Raum. Eine Familie von $Zva \{X_t, t \in T\}$ heißt **stochastischer Prozeß** mit Parameterbereich T und **Zustandsraum** I , falls X_t Zva mit Werten in I ist für alle t .

Dabei wird X_t als Zustand zur Zeit t interpretiert.

Wir betrachten hier den Fall: $T = \{0, \dots, N\}$ mit einem **endlichen Horizont** $N \in \mathbb{N}$.



Definition 14.7a. $\mathbb{P} = (p_{ij})_{i,j \in I}$ heißt **stochastische Matrix** und die p_{ij} **Übergangswahrscheinlichkeiten**, falls gilt:

$$(14.6) \quad p_{ij} \geq 0, i, j \in I, \sum_{j \in I} p_{ij} = 1 \quad (\text{d.h. } j \mapsto p_{ij} \text{ ist eine } W\text{-Funktion auf } I) \quad \forall i \in I.$$

Dabei wird p_{ij} die $W.$ sein, daß das System nach j geht, wenn es zuvor in i gewesen ist.

Proposition. Sei $\{X_t, t \in T\}$ wie in Def. 14.1 und $\mathbb{P} = (p_{ij})$ eine stochastische Matrix.

Dann sind äquivalent:

$$(14.7a) \quad P(X_0 = i_0, \dots, X_n = i_n) = P(X_0 = i_0) \cdot p_{i_0 i_1} \cdot p_{i_1 i_2} \cdot \dots \cdot p_{i_{n-1} i_n}, \quad \forall i_0, \dots, i_n \in I, n \in T;$$

$$(14.7b) \quad P(X_0 = i_0, \dots, X_{n+1} = i_{n+1}) = P(X_0 = i_0, \dots, X_n = i_n) \cdot p_{i_n i_{n+1}}, \quad \forall i_0, \dots, i_{n+1} \in I, n, n+1 \in T;$$

$$(14.7c) \quad P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i_n) = p_{i_n i_{n+1}} \\ \forall i_0, \dots, i_{n+1} \in I, n, n+1 \in T, \text{ mit } P(X_0 = i_0, \dots, X_n = i_n) > 0.$$

Beweis: Die Äquivalenz von (14.7a) und (14.7b) ist offensichtlich. Die Äquivalenz von (14.7c) mit (14.7b) ergibt sich mit Division durch $P(X_0 = i_0, \dots, X_n = i_n)$ in (14.7b). \square

Definition 14.7b. $(X_t, t \in T)$ heißt (zeit–) **homogene Markow–Kette**, falls eine stochastische Matrix $\mathbb{P} = (p_{ij})$ existiert, sodaß eine der drei äquivalenten Bedingungen (14.7a), (14.7b) oder (14.7c) erfüllt ist. Die Eigenschaft (14.7c) heißt **Markow–Eigenschaft** [Unabhängigkeit der Zukunft X_{n+1} von der Vergangenheit (X_0, \dots, X_{n-1}) bei bekannter Gegenwart X_n]. Die Verteilung von X_0 heißt **Startverteilung** und ist gegeben durch die W –funktion $v_i := P(X_0 = i)$.

Bemerkung. Eine homogene Markow-Kette ist eine Verallgemeinerung von identisch verteilten und unabhängigen Zva X_t , $t \in T$, mit Werten in I . Solche erhält man, falls man setzt: $p_{ij} = v_j$. Die Existenz einer Markoff-Kette zeigt man wie im Fall von unabhängigen Zva (vgl. §2.3 Produktexperimente). Man wählt $\Omega = I \times \dots \times I \ni \omega = (i_0, \dots, i_N)$, $P(\omega) = v_{i_0} \cdot p_{i_0 i_1} \cdot \dots \cdot p_{i_{N-1} i_N}$ und X_t wieder als Projektion. \square

Satz 14.5. (Variante der Markow-Eigenschaft): Es liege die Situation aus Def. 14.7b vor. Dann gilt $\forall i, j, j_1, \dots, j_m \in I, E \subset I^n$, $n \leq n+m \in T$, mit $P((X_0, \dots, X_{n-1}) \in E, X_n = i) > 0$:

- a) $P(X_{n+1} = j | (X_0, \dots, X_{n-1}) \in E, X_n = i) = p_{ij}$;
 b) $P(X_{n+1} = j_1, \dots, X_{n+m} = j_m | (X_0, \dots, X_{n-1}) \in E, X_n = i) = p_{ij_1} \cdot \dots \cdot p_{j_{m-1} j_m}$.

Beweis: Übung.

Wichtige Spezialfälle sind der Fall $E = I^n$ [dann kann die Bedingung " $(X_0, \dots, X_{n-1}) \in E$ " gestrichen werden] und der Fall $E = \{(i_0, \dots, i_{n-1})\}$ [dann hat man die Bedingung " $X_0 = i_0, \dots, X_{n-1} = i_{n-1}$ "].

Die **Matrixmultiplikation** von stochastischen Matrizen \mathbb{P}' und \mathbb{P}'' ist wie üblich definiert gemäß

$$\mathbb{P} = \mathbb{P}' \cdot \mathbb{P}'' \Leftrightarrow p_{ik} = \sum_{j \in I} p'_{ij} p''_{jk} .$$

Dann ist \mathbb{P} wieder eine stochastische Matrix. Denn es ist offenbar $p_{ik} \geq 0$, und

$$\sum_{k \in I} p_{ik} = \sum_{k \in I} \sum_{j \in I} p'_{ij} p_{jk} = \sum_{j \in I} p'_{ij} \sum_{k \in I} p_{jk} = \sum_{j \in I} p'_{ij} \cdot 1 = 1 .$$

Bezeichnungen $(p_{ij}^{(0)}) = \mathbb{P}^0$ ist die Einheitsmatrix,
 $(p_{ij}^{(n)}) = \mathbb{P}^n$ ist das n -fache Produkt von \mathbb{P} mit sich selbst.

Wir werden sehen, daß $p_{ij}^{(n)}$ die W. ist, daß das System in n Schritten von i nach j geht.

Lemma. $P(X_n = i, X_{n+m} = k) = P(X_n = i) \cdot p_{ik}^{(m)}$.

Beweis durch vollständige Induktion nach m . Für $m=1$ haben wir mit 14.5a ($E = I^n$):

$$P(X_n = i, X_{n+1} = k) = P(X_n = i) \cdot P(X_{n+1} = k | X_n = i) = P(X_n = i) \cdot p_{ik} .$$

Für den Schritt $m \rightarrow m+1$ schreiben wir mit der Formeln (1.23), (3.0b) von der totalen W.:

$$\begin{aligned} P(X_n = i, X_{n+m+1} = k) &= \sum_j P(X_n = i, X_{n+m} = j, X_{n+m+1} = k) \\ &= \sum_j P(X_n = i, X_{n+m} = j) \cdot P(X_{n+m+1} = k | X_n = i, X_{n+m} = j) \\ &= \sum_j P(X_n = i) \cdot p_{ij}^{(m)} p_{jk} = P(X_n = i) \sum_j p_{ij}^{(m)} p_{jk} = P(X_n = i) p_{ik}^{(m+1)} . \end{aligned}$$

Dabei wurde die Induktionsvoraussetzung und wieder 14.5a benutzt; denn wir haben offenbar $\{X_n = i\} = \{(X_0, \dots, X_{n+m-1}) \in E\}$ für eine geeignete Menge E . \square

14.6 Satz (Chapman–Kolmogorow–Gleichungen). Für $P(X_n = i) > 0$ gilt:

- (a) $P(X_{n+m} = k | X_n = i) = p_{ik}^{(m)}$;
 (b) $P(X_{n+m} = k | X_n = i) = \sum_j P(X_t = j | X_n = i) P(X_{n+m} = k | X_t = j)$, $n < t < n+m$.

Aufgründung der Eigenschaft 14.6a heißen die $p_{ij}^{(m)}$ **m–Schritt–Übergangswahrscheinlichkeiten**.

Beweis: a) Dividiere im Lemma durch $P(X_n = i)$.

b) Wegen $p^m = p^{t-n} \cdot p^{n+m-t}$ gilt auch

$$p_{ik}^{(m)} = \sum_j p_{ij}^{(t-n)} p_{jk}^{(n+m-t)}.$$

Nun folgt die Behauptung aus (a). \square

Korollar 14.10. $P(X_m = k) = \sum_i P(X_0 = i) p_{ik}^{(m)}$.

Beweis. Es ist nach der Formel von der totalen W.: $P(X_m = k) = \sum_i P(X_0 = i, X_m = k)$. Jetzt kann das Lemma eingesetzt werden. \square

Insbesondere gilt: $P(X_1 = k) = \sum_i P(X_0 = i) p_{ik}$.

Die Frage, wann die Verteilungen von X_0 und X_1 übereinstimmen, führt zu folgender Definition.

Definition 14.11. Eine W–Funktion $v_i, i \in I$, beschreibt eine **invariante Verteilung**, falls gilt:

$$(14.18) \quad v_k = \sum_{i \in I} v_i p_{ik}, \text{ d.h. } P(X_0 = i) = v_i \forall i \Rightarrow P(X_1 = k) = v_k, (\forall k \in I).$$

Ist $v_i, i \in I$, die W–Funktion von X_0 , gilt also $P(X_0 = i) = v_i$, so wird die Verteilung von X_0 gemäß (3.0a) beschrieben durch: $P(X_0 \in A) = \sum_{i \in A} v_i$.

Satz 14.12. Ist die Startverteilung eine invariante Verteilung, so sind die $X_n, n \in T$, identisch verteilt. (Man sagt auch: die Markow–Kette ist im Gleichgewicht.)

Beweis: Nach Voraussetzung beschreibt $P(X_0 = i) =: v_i$ eine invariante Verteilung. Der Beweis folgt nun mit Induktion aus (unter Benutzung des Lemmas):

$$P(X_{n+1} = k) = \sum_i P(X_n = i, X_{n+1} = k) = \sum_i P(X_n = i) p_{ik}. \square$$

Bemerkung. (ohne Beweis) Ist I endlich, so existiert immer eine invariante Verteilung. \square

Wir behandeln nun einen typischen Fall einer Markow–Kette.

Satz 14.13. Sind $(Y_n, 1 \leq n \leq N)$ unabhängige und identisch verteilte Zva mit Werten in \mathcal{Y} und existiert eine Funktion $f: I \times \mathcal{Y} \rightarrow I$, sodaß $X_n = f(X_{n-1}, Y_n), 1 \leq n \leq N$, gilt, so ist (X_n) eine Markow–Kette, falls X_0 konstant ist [oder allgemeiner falls auch $\{X_0, Y_n, 1 \leq n \leq N\}$ unabhängig sind]. Die Übergangs–W. sind dabei: $p_{ij} = P(f(i, Y_1) = j)$.

Beweis: Wir zeigen die Eigenschaft 14.7 (a):

$$\begin{aligned}
 P(X_0 = i_0, \dots, X_n = i_n) &= P(X_0 = i_0, f(X_0, Y_1) = i_1, \dots, f(X_{n-1}, Y_n) = i_n) \\
 &= P(X_0 = i_0, f(i_0, Y_1) = i_1, \dots, f(i_{n-1}, Y_n) = i_n) = P(X_0 = i_0) \cdot P(f(i_0, Y_1) = i_1) \cdot \dots \cdot P(f(i_{n-1}, Y_n) = i_n); \\
 &\text{denn es ist } \{X_0 = i_0\} \text{ entweder } \Omega \text{ oder } \emptyset, \text{ und man kann Satz 3.5 anwenden. Nun ist} \\
 P(f(i, Y_m) = j) &= P(f(i, Y_1) = j) = p_{ij}; \quad \square
 \end{aligned}$$

Beispiele (für die Situation von 14.13).

(1) *Summen von unabhängigen, identisch verteilten Zva* erhält man, falls $I = \mathbb{Z} \supset \mathfrak{Y}$ und $f(i, y) = i + y$. Dann hat man

$$(14.19) \quad X_n = i_0 + Y_1 + \dots + Y_n.$$

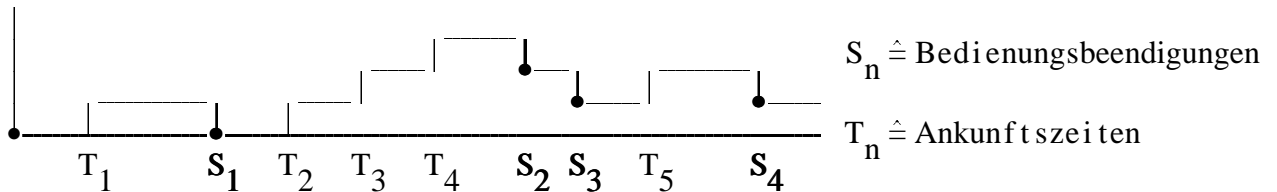
Für den Fall $\mathfrak{Y} = \{-1, 1\}$ heißt $\{X_n\}$ eine (1-dim.) **Irrfahrt**. $\xleftrightarrow[i-1]{i} \xleftrightarrow[i]{i+1} \mathbb{Z}$.

(2) *Ein Warteschlangenmodell*. Es sei $I = \mathbb{N}_0 = \{0, 1, 2, \dots\} \supset \mathfrak{Y}$ und $f(i, y) = (i-1)^+ + y$.

In einem Bedienungssystem mögen zu Beginn $X_0 = i_0$ Kunden auf Bedienung warten. Die Zva Y_n beschreibt die Anzahl der Ankünfte während der Bedienung des n -ten Kunden. Dann soll X_n die Länge der Warteschlange nach der n -ten Bedienung angeben. Dabei haben wir

$$(14.20a) \quad X_{n+1} = (X_n - 1)^+ + Y_{n+1} =: f(X_n, Y_{n+1}), \text{ insbesondere } X_{n+1} \geq X_n - 1.$$

Ist $X_n = 0$, so muß man warten, bis wieder ein Kunde ankommt. Dann ist man in der Situation, daß die Warteschlange zu Beginn der Bedienung gleich eins ist.



Wir schreiben $Y := Y_1$ für ein typisches Y_n . Dann ergibt sich nach Satz 14.12:

$$(14.20b) \quad p_{ij} = P[(i-1)^+ + Y = j]$$

Wir möchten in dieser Situation zunächst eine notwendige Bedingung für die Existenz einer invarianten Verteilung herleiten. Beschreibt $v_i, i \in I$, eine invariante Verteilung, so gilt:

$$(14.21) \quad v_k = \sum_i v_i p_{ik} = \sum_{i \leq k+1} v_i p_{ik} = \sum_{i \leq j} v_i p_{ik} \quad \text{für } k < j.$$

wegen $p_{ik} = 0$ für $k+1 < i$ gemäß (14.20). Wir zeigen nun

$$(14.22) \quad v_j p_{jj-1} = \sum_{i=0}^{j-1} v_i \sum_{k=j}^{\infty} p_{ik} \quad \text{für } j \geq 1.$$

Zum **Beweis** schreiben wir unter Benutzung von (14.21):

$$\begin{aligned}
 \sum_{k < j} v_k &= \sum_{k < j} \left[\sum_{i \leq j} v_i p_{ik} \right] = \sum_{i \leq j} \left[\sum_{k < j} v_i p_{ik} \right] = v_j p_{jj-1} + \sum_{i < j} v_i \sum_{k < j} p_{ik} \quad \text{sowie} \\
 \sum_{i < j} v_i &= \sum_{i < j} v_i \sum_k p_{ik} = \sum_{i < j} v_i \sum_{k < j} p_{ik} + \sum_{i < j} v_i \sum_{k \geq j} p_{ik}.
 \end{aligned}$$

Durch Vergleich dieser Identitäten folgt (14.22). Daraus erhalten wir

wegen $p_{jj-1} = 1 - \sum_{k=j}^{\infty} p_{jk}$:

$$(14.23) \quad v_j = \sum_{i=0}^j v_i \sum_{k=j}^{\infty} p_{ik} \quad \text{für } j \geq 0.$$

$$\begin{aligned} \text{Damit ergibt sich: } 1 &= \sum_{j=0}^{\infty} v_j = \sum_{j=0}^{\infty} \sum_{i=0}^j v_i \sum_{k=j}^{\infty} p_{ik} \\ &= \sum_{i=0}^{\infty} v_i \sum_{k=i}^{\infty} p_{ik} \sum_{j=i}^k 1 = \sum_{i=0}^{\infty} v_i \sum_{k=i}^{\infty} (k-i+1) p_{ik} \\ &= v_0 \sum_{k=0}^{\infty} (k+1) P(Y = k) + \sum_{i=1}^{\infty} v_i \sum_{k=i}^{\infty} (k-i+1) P(i-1 + Y = k) \\ &= v_0 + \sum_{i=0}^{\infty} v_i \sum_{\ell=1}^{\infty} \ell P(Y = \ell) = v_0 + E(Y), \text{ also} \end{aligned}$$

$$(14.24) \quad v_0 = 1 - E(Y) \quad \text{mit } E(Y) = \sum_{\ell=1}^{\infty} \ell P(Y = \ell).$$

Dabei ist $E(Y)$ die mittlere Anzahl der Ankünfte während einer Bedienung. Somit ergibt sich als notwendige Bedingung für die Existenz einer invarianten Verteilung $E(Y_1) \leq 1$. Wir wollen uns noch überlegen, daß sogar $E(Y_1) < 1$ gelten muß. Sei nun $E(Y_1) \leq 1$.

Wäre $p_{jj-1} = P(Y_1 = 0) = 0$, so müßte $P(Y_1 = 1) = 1$ gelten; dieser Fall ist aber leicht zu übersehen und nicht interessant. Sei nun $p_{jj-1} = P(Y_1 = 0) > 0$. Ist dann $E(Y_1) = 1$, so ist $v_0 = 0$. Daraus folgt aber nach (14.23): $v_j = 0 \quad \forall j$; dann beschreibt aber $v_j, j \in I$, keine invariante Verteilung. Also erhalten wir, falls $P(Y_1 = 1) < 1$, als

notwendige Bedingung: $E(Y_1) < 1$.

Es kann sogar gezeigt werden, daß diese Bedingung auch hinreichend ist. Die Angabe von v_0 in (14.24) erlaubt mit Hilfe von (14.22) eine rekursive Berechnung der invarianten Verteilung.

Hat \mathbf{Y} eine **geometrische Verteilung** [und damit alle Y_n], so kann man nicht nur direkt zeigen, daß $E(Y) < 1$ auch hinreichend für die Existenz einer invarianten Verteilung ist, sondern man kann die invariante Verteilung auch angeben, nämlich

$$v_i = (1 - \rho) \cdot \rho^i, \quad i \geq 0, \quad \text{mit } \rho := E(Y) < 1.$$

Dann gilt insbesondere (14.24). Zum Nachweis müssen wir in Hinblick auf (14.21) zeigen:

$$v_k = \sum_{i=1}^{k+1} v_i \cdot P(Y = k-i+1) + v_0 \cdot P(Y = k).$$

Dabei wissen wir nach §3: $P(Y = j) = p \cdot q^j = p \cdot (1-p)^j$; $E(Y) = q/p = \rho$. Nun ist

$$\begin{aligned} \sum_{i=1}^{k+1} v_i \cdot P(Y = k-i+1) + v_0 \cdot P(Y = k) &= \sum_{i=1}^{k+1} (1 - \rho) \cdot \rho^i \cdot p \cdot q^{k-i+1} + (1 - \rho) \cdot p \cdot q^k \\ &= (1 - \rho) \cdot p \cdot q^{k+1} \cdot \sum_{i=1}^{k+1} p^{-i} + (1 - \rho) \cdot p \cdot q^k = (1 - \rho) \cdot q^{k+1} \cdot \sum_{j=0}^k p^{-j} + (1 - \rho) \cdot p \cdot q^k \\ &= (1 - \rho) \cdot q^{k+1} \cdot \frac{p^{-(k+1)} - 1}{p^{-1} - 1} + (1 - \rho) \cdot p \cdot q^k = (1 - \rho) \cdot q^{k+1} \cdot \frac{p}{q} \cdot [p^{-(k+1)} - 1] + (1 - \rho) \cdot p \cdot q^k \\ &= (1 - \rho) \cdot q^k \cdot p^{-k} = (1 - \rho) \cdot \rho^k = v_k. \end{aligned}$$

Damit beschreibt $v_i, i \geq 0$, wirklich eine invariante Verteilung.

Kapitel II Allgemeine Modelle

§ 9 Wahrscheinlichkeitsmaße mit Dichten

Es sollen nun überabzählbare W -Räume Ω behandelt werden. Dabei beschränken wir uns auf den für die Praxis wichtigsten Fall:

$$\Omega \subset \mathbb{R} \text{ oder auch } \Omega \subset \mathbb{R}^n \text{ für ein } n \in \mathbb{N}.$$

Die Vorgehensweise kann weitgehend parallel zu endlichen oder abzählbaren Räumen durchgeführt werden. Dabei gehen wir nach dem Prinzip vor, dass wir Ausdrücke der Form

$$\sum_{\omega \in A} R(\omega) \text{ ersetzen durch Ausdrücke } \int_a^b R(t) dt.$$

So konnte in (1.10) die W . $P(A)$ durch die W -Funktion $P(\omega)$ ausgedrückt werden gemäß

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Die Rolle der W -Funktion wird jetzt eine **Dichte** $f(t)$ übernehmen, derart daß gilt:

$$P([a,b]) = \int_a^b f(t) dt =: \int_{[a,b]} f(t) dt = \int \mathbf{1}_{[a,b]}(t) f(t) dt.$$

Ein Beispiel haben wir schon mit der Dichte ϕ der Normalverteilung kennengelernt. Allerdings werden wir nicht nur Wahrscheinlichkeiten von Intervallen zu betrachten haben, sondern auch solche von Teilmengen/Ereignissen, die sich etwa durch Komplementbildung oder durch Vereinigungen und Durchschnitte von endlich oder abzählbar vielen Intervallen ergeben. Dabei kann man aber nicht wie bisher allen Teilmengen A über die obige Formel eine W . $P(A)$ zuordnen, weil das Integral

$$\int_A f(t) dt = \int \mathbf{1}_A(t) f(t) dt, \quad A \subset \mathbb{R},$$

oder

$$\int \int \dots \int f(t_1, \dots, t_n) dt_1 \dots dt_n = \int \mathbf{1}_A(t_1, \dots, t_n) f(t_1, \dots, t_n) dt_1 \dots dt_n, \quad A \subset \mathbb{R}^n,$$

auch für sehr "gutartige" Funktionen f nicht für alle $A \subset \mathbb{R}$ sinnvoll erklärt werden kann. Das wird dazu führen, daß wir im Falle $\Omega = \mathbb{R}$ oder $\Omega = \mathbb{R}^n$ auch nicht mehr alle $A \subset \Omega$ als Ereignisse interpretieren können.

Das System $\mathfrak{B} = \mathfrak{B}_n$ der **Borelschen Teilmengen (Borel-Mengen)** des \mathbb{R}^n , für das das Integral erklärt werden kann, ist aber sehr groß. Es enthält alle für die Praxis wichtigen Teilmengen.

Das gleiche gilt für das System f der (reellen) Funktion auf \mathbb{R}^n , für die das Integral erklärt werden kann. Dieses ist das System der **Borelschen Funktionen** oder **Borel-Funktionen**. [Wir werden natürlich Integrierbarkeitsvoraussetzungen zu beachten haben, die ja auch schon bei abzählbaren Räumen Ω auftraten. Dort mußte die Existenz gewisser Summen (Reihen) gefordert werden.]

Auf die genaue Definition der Borel-Mengen und der Borel-Funktionen soll hier verzichtet werden. Wir wollen uns vielmehr mit dem Hinweis begnügen, daß man bei Anwendungen ausschließlich mit Borelschen Teilmengen oder Borelschen Funktionen in Berührung kommt. Es erfordert sogar erheblichen mathematischen Aufwand und Sachverstand, eine Menge A oder eine Funktion f zu konstruieren, die nicht Borelsch sind.

Definition 9.4. Sei Ω eine Borelsche Teilmenge des \mathbb{R}^n (z.B. $\Omega = \text{Intervall}$).

Ein **Wahrscheinlichkeitsmaß** $P(A)$ auf Ω ist eine für alle Borelschen Teilmengen A von Ω definierte reellwertige Funktion mit den Eigenschaften:

$$(9.4) \quad P(A) \geq 0 \text{ für alle } A \text{ (Positivität),}$$

$$(9.5) \quad P(\Omega) = 1 \text{ (Normiertheit),}$$

$$(9.6) \quad P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots \text{ für disjunkte } A_1, A_2, \dots \text{ (}\sigma\text{-Additivität).}$$

(Ω, P) heißt **W-Raum** und die Borelschen Teilmengen von Ω heißen **Ereignisse**.

Die früher angegebenen Eigenschaften von W-Maßen (1.4) – (1.9) gelten weiter, da sie aus den obigen drei Eigenschaften abgeleitet wurden.

Definition. Ist P ein W-Maß auf \mathbb{R} , so heißt die folgende Funktion F **Verteilungsfunktion** zu P :

$$(9.7) \quad F(t) := P((-\infty, t]), \quad t \in \mathbb{R}.$$

Wegen $(a, b] = (-\infty, b] \setminus (-\infty, a]$ für $a < b$ gilt dann auf Grund der Subtraktivität (1.6) von P :

$$(9.8) \quad P((a, b]) = F(b) - F(a) \text{ für } a < b.$$

Definition. Eine Borel-Funktion $f \geq 0$ auf \mathbb{R} heißt eine **Dichte auf \mathbb{R}** , falls gilt

$$(9.9) \quad \int_{-\infty}^{\infty} f(t) dt = 1.$$

Diese Eigenschaft entspricht (1.11) für W-funktionen. In der Maßtheorie zeigt man:

Existenz- und Eindeutigkeitsatz für W-Maße: Ist f eine Dichte auf \mathbb{R} , so existiert genau ein W-Maß P auf \mathbb{R} mit

$$(9.10) \quad F(t) = P((-\infty, t]) = \int_{-\infty}^t f(s) ds, \quad t \in \mathbb{R};$$

$$(9.11) \quad P((a, b]) = \int_a^b f(t) dt = P((a, b)) = P([a, b]) \text{ , } -\infty < a < b < \infty.$$

In der Situation (9.10) oder (9.11) heißt f **Dichte von P** .

Bemerkung. Man kann P direkt definieren über $P(A) = \int \mathbf{1}_A(t) \cdot f(t) dt$ für alle Borel-Mengen A . Nichtmathematiker schreiben die Beziehung (9.11) gern mit infinitesimalen Größen in der Form

$$P([t, t+dt]) = f(t) dt. \quad \square$$

Will man $P(A)$ berechnen für $A = [a, b] \cup [c, d]$ mit $a < b < c < d$, so hat man auf Grund der Additivität von P :

$$P(A) = \int_a^b f(t) dt + \int_c^d f(t) dt =: \int_{[a,b] \cup [c,d]} f(t) dt.$$

In der Situation (9.10) ist F offenbar eine stetige und monoton wachsende Funktion. Ist f stetig, so ist F auch eine Stammfunktion zu f .

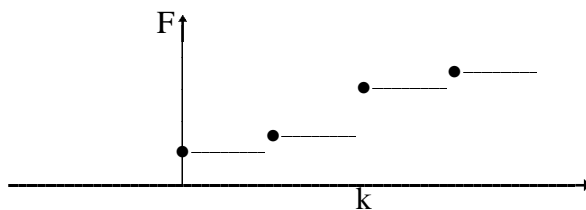
Ist P dagegen ein W -Maß auf $\mathbb{N}_0 = \{0,1,2,\dots\}$, so kann man P auch als W -Maß auf \mathbb{R} auffassen mit

$$P(A) := P(\mathbb{N}_0 \cap A) = \sum_{k \in \mathbb{N}_0 \cap A} P(k).$$

Dabei ist $P(A)$ sogar für alle Teilmengen von \mathbb{R} definiert, und es gilt: $F(t) = 0$ für $t < 0$ und

$$F(t) = \sum_{k=0}^{\lceil t \rceil} P(k).$$

Somit ist F wieder monoton wachsend, aber lediglich rechtseitig stetig. F hat einen Sprung der Höhe $P(k)$ im Punkt k und ist in Intervallen $[k, k+1)$ konstant.



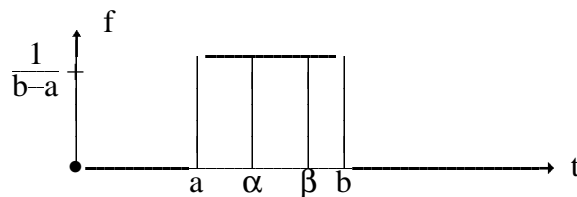
Zusatz zum Existenz- und Eindeutigkeitssatz für W -Maße.

Ist f eine Dichte auf \mathbb{R} und gilt $f(t) = 0$, $t \notin \Omega$, für ein Intervall $\Omega \subset \mathbb{R}$, so gilt für das W -Maß P mit (9.10), (9.11): $P(\Omega) = 1$. Dann kann P auch als **W -Maß auf Ω** aufgefaßt werden.

Spezialfälle.

(a) **Gleichverteilung** P auf einem Intervall $[a, b]$ (oder (a, b) , $(a, b]$). Diese ist das W -Maß P mit der Dichte

$$f(t) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(t).$$



Dann gilt für $a \leq \alpha < \beta \leq b$:

$$P([\alpha, \beta]) = \int_{\alpha}^{\beta} \frac{1}{b-a} dt = \frac{1}{b-a} \cdot (\beta - \alpha).$$

Wir können hier sowohl $\Omega = \mathbb{R}$ als auch $\Omega = [a, b]$ wählen. Dies ist offenbar ein Analogon zur diskreten Gleichverteilung. Dort hatten wir eine konstante W -Funktion, während wir hier eine konstante Dichte auf $[a, b]$ haben. Die Gleichverteilung auf $[0, 1]$ hat die Verteilungsfunktion $F(x) = x$ für $x \in [0, 1]$.

Dreht man ein Glücksrad und zeigt nach der Drehung ein Zeiger auf den Winkel in $\Omega = [0, 2\pi]$, so ist es naheliegend, eine Gleichverteilung auf Ω anzunehmen.

Kommt man zu einem rein zufälligen Zeitpunkt an eine Bushaltestelle, die von einem Bus alle zehn Minuten angefahren wird, so ist es naheliegend, in einem Modell für die Wartezeit mit einer Gleichverteilung auf $[0, 10]$ zu arbeiten.

(b) **Exponentialverteilung** P mit Parameter $\lambda > 0$. Diese ist das W -Maß P mit der Dichte

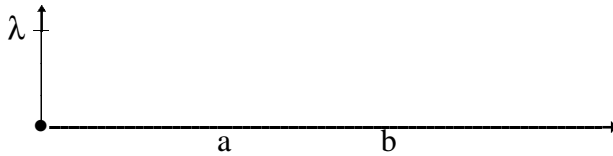
$$f(t) = \lambda \cdot e^{-\lambda t} \cdot \mathbf{1}_{[0, \infty)}(t).$$

Dann gilt für $0 \leq a < b < \infty$:

$$P([a, b]) = \int_a^b \lambda \cdot e^{-\lambda t} dt = e^{-\lambda a} - e^{-\lambda b} \quad \text{für } 0 \leq a < b < \infty,$$

$$F(t) = P((-\infty, t]) = P([0, t]) = 1 - e^{-\lambda t} \quad \text{für } t \geq 0 \quad \text{und } F(t) = 0 \quad \text{für } t \leq 0.$$

Wir können hier sowohl $\Omega = \mathbb{R}$ als auch $\Omega = [0, \infty)$ wählen.



Zeichne $f(t)$ und kennzeichne die Fläche unter f über dem Intervall $[a, b]$.

Eine charakteristische Eigenschaft der Exponentialverteilung P ist die sogenannte **Gedächtnislosigkeit**, d.h. es gilt für die bedingte W .

$$(9.13) \quad P[(t+x, \infty) | (t, \infty)] = P[(x, \infty)] .$$

$$\text{Wegen } P[(t, \infty)] = e^{-\lambda t} \text{ gilt: } P[(t+x, \infty) | (t, \infty)] = P[(t+x, \infty)] / P[(t, \infty)] = e^{-\lambda(t+x)} / e^{-\lambda t} = e^{-\lambda x} .$$

Die Exponentialverteilung ist ein stetiges Analogon zur geometrischen Verteilung. Beschreibt $\Omega = [0, \infty)$ oder $(0, \infty)$ mögliche Zeitpunkte eines Ereignisses (Ausfall einer Maschine, Eintreten eines Schadenfalls, Aussendung eines Teilchens einer radioaktiven Substanz), so wird oft die Exponentialverteilung als W -Maß benutzt.

(c) **Standardisierte Normalverteilung.**

Die standardisierte Normalverteilung ist gegeben durch die Dichte

$$(9.15) \quad \varphi(t) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}t^2) , \quad t \in \mathbb{R}.$$

Wir wählen also $\Omega = \mathbb{R}$.

Die Normalverteilung spielte bereits beim zentralen Grenzwertsatz eine große Rolle. In Hinblick darauf dient sie oft als Approximation für W -Maße bei Phänomenen, die durch viele unabhängige Größen beeinflusst werden.

§10 Zufallsvariable und ihre Momente

10.1 Zufallsvariable

Definition 10.1. Sei $\Omega \subset \mathbb{R}^d$ eine Borel-Menge [z.B. ein (mehrdimensionales) Intervall]. Dann heißt eine Borel-Funktion $X : \Omega \rightarrow \mathbb{R}$ **Zufallsvariable** (Zva) auf (Ω, P) und man schreibt:

$$(10.1) \quad \{X \in A\} = \{\omega \in \Omega, X(\omega) \in A\} \quad \text{für alle Borel-Mengen } A \subset \mathbb{R}.$$

Bemerkung. Die Definition einer Borel-Funktion ist gerade so gemacht, daß $\{X \in A\}$ für alle Borel-Mengen $A \subset \mathbb{R}$ wieder eine Borel-Menge in \mathbb{R}^d ist, also ein Ereignis.

Bemerkung 10.3. Ist X Zva auf Ω und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine Borel-Funktion, so ist $g(X)$ [mit $g(X)(\omega) = g(X(\omega))$] wieder eine Zva.

Bemerkung 10.4. Sind X_1, \dots, X_n Zva auf (Ω, P) , so heißt $X := (X_1, \dots, X_n)$ **Zufallsvektor**.

Ist $g : \mathbb{R}^n \rightarrow \mathbb{R}$ eine Borel-Funktion, so ist $g(X) = g(X_1, \dots, X_n)$ wieder eine Zva.

10.2 Verteilungen von Zufallsvariablen

Definition. Ist $X : \Omega \rightarrow \mathbb{R}$ wie in 10.1 eine Zva auf (Ω, P) mit $\Omega \subset \mathbb{R}^d$, so nennt man das durch

$$(10.3) \quad P'(B) = P(X \in B) := P(\{\omega \in \Omega; X(\omega) \in B\}), \quad B \subset \mathbb{R},$$

auf \mathbb{R} definierte W-Maß P' die **Verteilung** von X .

Wie in §3.1, Bemerkung, sieht man, daß P' wirklich ein W-Maß ist. Die Verteilungsfunktion zu P' ist dabei:

$$F(x) = P(X \in (-\infty, x]) = P(X \leq x).$$

Man nennt F auch **Verteilungsfunktion von X** . Hat die Verteilung P' eine Dichte f , d.h. gilt

$$P(a \leq X \leq b) = P'([a, b]) = \int_a^b f(t) dt,$$

$$P(X \leq x) = P'((-\infty, x]) = \int_{-\infty}^x f(t) dt,$$

so nennt man f auch **Dichte von X** .

Beispiel: X hat eine Gleichverteilung (ist gleichverteilt) auf $[a, b]$, wenn gilt:

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} \frac{1}{b-a} dt = \frac{\beta-\alpha}{b-a} \quad \text{für } a \leq \alpha < \beta \leq b. \quad \square$$

Beispiel: X hat eine **Exponentialverteilung** (mit Parameter λ), wenn P' eine Exponentialverteilung ist, d.h. wenn gilt:

$$P(a \leq X \leq b) = \int_a^b \lambda \cdot e^{-\lambda t} dt = e^{-\lambda a} - e^{-\lambda b},$$

$$P(X \leq x) = \int_{-\infty}^x \lambda \cdot e^{-\lambda t} dt = 1 - e^{-\lambda x},$$

$$P(X > x) = e^{-\lambda x}, \quad x \geq 0.$$

Dann besagt die Gedächtnislosigkeit (9.13):

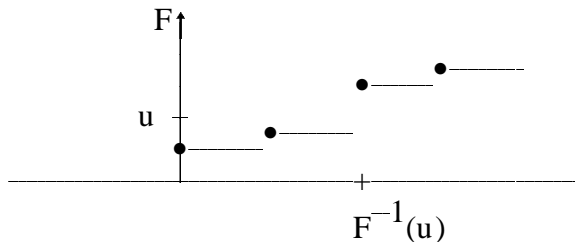
$$P(X > t+x | X > t) = P(X > x), \quad x, t \geq 0. \quad \square$$

Beispiel: X hat eine (standard.) Normalverteilung (ist normalverteilt), wenn gilt:

$$P(a \leq X \leq b) = \int_a^b \varphi(t) dt = \Phi(b) - \Phi(a) \quad \text{für } a < b. \quad \square$$

Beispiel. Die verallgemeinerte Inverse F^{-1} für eine beliebige Verteilungsfunktion F ist:

$$F^{-1}(u) := \inf \{x \in \mathbb{R}; F(x) \geq u\} = \min \{\dots\}, \quad 0 < u < 1.$$



Ist F stetig und streng monoton (wie bei der Normalverteilung), so gibt es genau ein x mit $F(x) = u$, und man hat wie üblich $F^{-1}(u) = x$. Es gilt die folgende

Eigenschaft:

Sei $\Omega = (0,1)$, P die **Gleichverteilung auf $(0,1)$** und F die Verteilungsfunktion eines beliebigen W -Maßes auf \mathbb{R} . Dann ist $X := F^{-1}$ eine Zva auf (Ω, P) mit:

(*) X hat die (vorgegebene) Verteilungsfunktion F .

Kann man also ein Experiment konstruieren, das durch die Gleichverteilung auf $(0,1)$ beschrieben werden kann, so hat man sofort eine Zva mit einer beliebig vorgegebenen Verteilungsfunktion. So können wir z. B. für F die Verteilungsfunktion Φ der Normalverteilung nehmen. Dann ist Φ^{-1} die übliche Inverse, und man weiß dann, daß $X = \Phi^{-1}$ eine Normalverteilung hat.

Auf dem Computer kann man sich für die **Simulation** von stochastischen Vorgängen sogenannte **Zufallszahlen** ausgeben lassen, die man als Ergebnisse von einem Experiment auffassen kann, das sich durch Ω und P wie oben beschreiben läßt.

Zum Beweis von (*) überlegt man sich, daß gilt:

$$(**) \quad F^{-1}(u) \leq x \Leftrightarrow F(x) \geq u, \quad 0 < u < 1.$$

Die Beziehung " \Leftarrow " ist offensichtlich, und für " \Rightarrow " benutzt man, daß F rechtsseitig stetig ist.

Dann folgt für die Verteilungsfunktion von $X = F^{-1}$:

$$P(X \leq x) = P(\{u \in (0,1); F^{-1}(u) \leq x\}) = P(\{u \in (0,1); u \leq F(x)\}) = P((0, F(x)]) = F(x). \quad \square$$

Satz 10.6 Ist $Y = c \cdot X + b$ mit $c \neq 0$, $b \in \mathbb{R}$, und hat die Zva X die Dichte f , so hat Y die Dichte

$$g(y) = \frac{1}{|c|} f\left(\frac{y-b}{c}\right).$$

Beweis: Wir betrachten hier nur den (schwierigeren) Fall $c = -\gamma < 0$. Dann gilt:

$$\begin{aligned} P(\alpha \leq b - \gamma \cdot X \leq \beta) &= P(b - \beta \leq \gamma \cdot X \leq b - \alpha) = P\left(\frac{b-\beta}{\gamma} \leq X \leq \frac{b-\alpha}{\gamma}\right) \\ &= \int_{(b-\beta)/\gamma}^{(b-\alpha)/\gamma} f(x) dx = \int_{\alpha}^{\beta} f\left(-\frac{1}{\gamma}(y-b)\right) \frac{1}{\gamma} dy \quad \text{mit } y = b - \gamma x, \end{aligned}$$

also

$$P(\alpha \leq Y \leq \beta) = \int_{\alpha}^{\beta} g(y) dy. \quad \square$$

10.3 Unabhängigkeit.

Definition. Die Zva X_m , $1 \leq m \leq n$, auf (Ω, P) ($\Omega \subset \mathbb{R}^d$) heißen **unabhängig**, wenn gilt:

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdot \dots \cdot P(X_n \in A_n) \quad \text{für alle Borel-Mengen } A_m.$$

Diese Eigenschaft ist die gleiche wie in (3.3) für diskrete W -Räume. Ist $X = (X_1, \dots, X_n)$ der zugehörige Zufallsvektor, so haben wir $P(X_1 \in A_1, \dots, X_n \in A_n) = P(X \in A_1 \times \dots \times A_n)$.

Wir werden $P(X \in A)$ für $n = 2$ noch für allgemeinere Mengen $A \subset \mathbb{R}^2$ benötigen, nicht nur für kartesische Produkte $A_1 \times A_2$. Bei diskreten W -Räumen hat man die folgende Beziehung für unabhängige Zva X_1, X_2 (vgl. (3.0e)):

$$\begin{aligned} P((X_1, X_2) \in A) &= \sum_{(x_1, x_2) \in A} P((X_1, X_2) = (x_1, x_2)) = \sum_{x_1, x_2} \mathbf{1}_A((x_1, x_2)) \cdot P(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1} \sum_{x_2} \mathbf{1}_A((x_1, x_2)) \cdot P(X_1 = x_1) \cdot P(X_2 = x_2) \\ &= \sum_{x_1} P(X_1 = x_1) \left[\sum_{x_2} \mathbf{1}_A((x_1, x_2)) \cdot P(X_2 = x_2) \right]. \end{aligned}$$

Satz 10.7. Seien X_1, X_2 unabhängige Zva auf (Ω, P) mit Dichten f_1, f_2 . Dann gilt:

$$\begin{aligned} P((X_1, X_2) \in A) &\stackrel{(+)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}_A(x_1, x_2) \cdot f_1(x_1) \cdot f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \left[\int_{-\infty}^{\infty} \mathbf{1}_A(x_1, x_2) \cdot f_2(x_2) dx_2 \right] dx_1 \quad \text{für alle Borel-Mengen } A \subset \mathbb{R}^n. \end{aligned}$$

Bemerkung: Man nennt $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ in der Situation (+) auch die Dichte des Zufallsvektors $X = (X_1, X_2)$. Bei Unabhängigkeit gilt also für die Dichte von (X_1, X_2) :

$$f(x_1, x_2) := f_1(x_1) \cdot f_2(x_2),$$

d.h. die Dichte von (X_1, X_2) ergibt sich aus dem Produkt der Dichten f_m der X_m ($f_m(x_m)$ ist die Dichte von X_m). Im diskreten Fall hatten wir:

$$P((X_1, X_2) = (x_1, x_2)) = P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2);$$

d.h. die W-Funktion von (X_1, X_2) ergibt sich aus dem Produkt der W-Funktionen der X_m . Auch hier entsprechen die W.-Funktionen im diskreten Fall den Dichten im allgemeinen Fall. \square

Beweis von 10.7. Wir betrachten nur den speziellen Fall $A = [a_1, b_1] \times [a_2, b_2]$;

der allgemeinere Fall folgt dann aus Sätzen der Maßtheorie. Es ist dann

$$\begin{aligned} P((X_1, X_2) \in [a_1, b_1] \times [a_2, b_2]) &= P(X_1 \in [a_1, b_1]) \cdot P(X_2 \in [a_2, b_2]) \\ &= \int_{a_1}^{b_1} f_1(x_1) dx_1 \cdot \int_{a_2}^{b_2} f_2(x_2) dx_2 = \int_{a_1}^{b_1} f_1(x_1) \left[\int_{a_2}^{b_2} f_2(x_2) dx_2 \right] dx_1 \\ &= \int_{-\infty}^{\infty} \mathbf{1}_{[a_1, b_1]}(x_1) \cdot f_1(x_1) \int_{-\infty}^{\infty} \mathbf{1}_{[a_2, b_2]}(x_2) \cdot f_2(x_2) dx_2 dx_1 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \left[\int_{-\infty}^{\infty} \mathbf{1}_A(x_1, x_2) \cdot f_2(x_2) dx_2 \right] dx_1 \quad \text{für alle } a_m < b_m. \quad \square \end{aligned}$$

10.3A Faltungen

Sind X_1 und X_2 unabhängige Zva mit Werten in der Menge \mathbb{Z} der ganzen Zahlen [statt in \mathbb{N}_0 wie

in § 5.4], so ist mit $A := \{(i, j); i+j \leq t\}$, $\mathbf{1}_A((i, j)) = \mathbf{1}_{(-\infty, t-i]}(j)$:

$$\begin{aligned} P(X_1 + X_2 \leq t) &= P((X_1, X_2) \in A) = \sum_{(i, j) \in A} P(X_1=i, X_2=j) = \sum_i \sum_j \mathbf{1}_A(i, j) P(X_1=i) \cdot P(X_2=j) \\ &= \sum_i P(X_1=i) \cdot \sum_j \mathbf{1}_{(-\infty, t-i]}(j) \cdot P(X_2=j) = \sum_i P(X_1=i) \cdot \sum_{j \leq t-i} P(X_2=j) \\ &= \sum_i P(X_1=i) \cdot \sum_{k \leq t} P(X_2=k-i) \quad \text{mit } k = i+j \\ &= \sum_{k=-\infty}^t \left[\sum_{i=-\infty}^{\infty} P(X_1=i) P(X_2=k-i) \right]. \end{aligned}$$

Dann folgt übrigens wie in §5.4:

$$P(X_1 + X_2 = k) = \sum_i P(X_2=k-i) \cdot P(X_1=i).$$

Sind nun X_1 und X_2 unabhängige Zva mit Dichten f_1, f_2 , so gehen wir unter Benutzung von Satz

10.7 analog vor mit $A := \{(x_1, x_2) \in \mathbb{R}^2; x_1 + x_2 \leq t\}$, $\mathbf{1}_A((x_1, x_2)) = \mathbf{1}_{(-\infty, t-x_1]}(x_2)$:

$$\begin{aligned} P(X_1 + X_2 \leq t) &= P((X_1, X_2) \in A) = \int f_1(x_1) \left[\int \mathbf{1}_A(x_1, x_2) \cdot f_2(x_2) dx_2 \right] dx_1 \\ &= \int f_1(x_1) \left[\int \mathbf{1}_{(-\infty, t-x_1]}(x_2) \cdot f_2(x_2) dx_2 \right] dx_1 = \int f_1(x_1) \left[\int_{-\infty}^{t-x_1} f_2(x_2) dx_2 \right] dx_1 \\ &= \int f_1(z) \left[\int_{-\infty}^t f_2(y-z) dy \right] dz = \int_{-\infty}^t \left[\int_{-\infty}^{\infty} f_2(y-z) \cdot f_1(z) dz \right] dy \quad \text{mit } y = x_1 + x_2, z = x_1. \end{aligned}$$

Damit erhält man für die Verteilungsfunktion $F(t) = P(X_1 + X_2 \leq t)$ von $X_1 + X_2$:

$$P(X_1 + X_2 \leq t) = F(t) = \int_{-\infty}^t f_1 * f_2(y) dy \quad \text{mit}$$

$$(10.6) \quad f_1 * f_2(y) := \int f_2(y-z) \cdot f_1(z) dz.$$

Definition. $f_1 * f_2$ heißt **Faltung** der Dichten f_1 und f_2 .

Satz 10.8. Sind X_1, X_2 unabhängige Zva mit Dichten f_1, f_2 , so hat X_1+X_2 die Dichte $f_1 \star f_2$.

Die Faltung von zwei Dichten ist wieder eine Dichte (nämlich die von X_1+X_2). Dies kann man wie §5.4 auch direkt zeigen.

Gilt $f_i(z) = 0$ für $z < 0, i=1,2$, wie etwa bei der Exponentialverteilung, so erhält man:

$$(10.7) \quad f_1 \star f_2(y) := \int_0^y f_2(y-z) \cdot f_1(z) dz, \quad y \geq 0.$$

Diese Formel entspricht (5.19). Wir betrachten nun eine **Anwendung**. Dabei sagt man: die Zva X ist normalverteilt/exponentialverteilt, wenn X eine Normalverteilung/ Exponentialverteilung hat.

Satz 10.9. Sind X_1, X_2 unabhängige, identisch verteilte, (standard) normalverteilte Zva, so ist $c_1 \cdot X_1 + c_2 \cdot X_2$ wieder (standard-) normalverteilt, falls $c_1^2 + c_2^2 = 1$.

Beweis: Sei $\varphi(x)$ die Dichte der Normalverteilung, also nach (5.11), (9.15):

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x^2\right).$$

Nach Satz 10.6 hat $c_i \cdot X_i$ die Dichte $g_i(z) := \frac{1}{|c_i|} \cdot \varphi\left(\frac{z}{c_i}\right)$ für $c_i \neq 0$. Ist $c_1 = \pm 1, c_2 = 0$, so folgt die Aussage sofort.

Sei jetzt $c_1 \neq 0 \neq c_2$. Dabei sind $c_i \cdot X_i, i=1,2$ wieder unabhängige Zva. [Satz 3.5 überträgt sich auf diesen Fall.] Also hat $c_1 \cdot X_1 + c_2 \cdot X_2$ die Dichte nach Satz 10.8:

$$f(y) = \int_0^y g_2(y-z) \cdot g_1(z) dz = \frac{1}{2\pi \cdot |c_1 \cdot c_2|} \int \exp\left[-\frac{1}{2} \left[\left(\frac{y-z}{c_2}\right)^2 + \left(\frac{z}{c_1}\right)^2\right]\right] dz.$$

Macht man die Substitution $t := \frac{z}{|c_1 \cdot c_2|} - y \cdot |c_1/c_2|$, so ist $dz = |c_1 \cdot c_2| dt$

und eine kleine Rechnung liefert: $[...] = t^2 + y^2$. Damit erhalten wir

$$\begin{aligned} f(y) &= \frac{1}{2\pi \cdot |c_1 \cdot c_2|} \int \exp\left(-\frac{1}{2} [t^2 + y^2]\right) |c_1 \cdot c_2| dt \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} y^2\right) \cdot \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{1}{2} t^2\right) dt = \varphi(y). \quad \square \end{aligned}$$

10.4 Erwartungswerte.

Bei diskreten W-Räumen hatten wir nach Satz 3.8 die folgende Formel für $E(g(X))$:

$$E(g(X)) = \sum_x g(x) \cdot P(X=x).$$

Diese Formel werden wir als Vorbild nehmen für den Erwartungswert von Zva mit Dichten. In Def. 3.6 war der Erwartungswert zunächst als Summe $E(X) = \sum_{\omega \in \Omega} X(\omega)$ definiert worden. Dies würde einer Integration über Ω entsprechen. Da Ω aber i.a. mehrdimensional ist, ist die andere Formel günstiger.

Definition 10.10. Sei X eine reelle Zva auf (Ω, P) mit der Dichte f . Ferner sei g eine reelle Borel-Funktion auf \mathbb{R} . Dann wird der **Erwartungswert** $E(g(X))$ definiert als

$$(10.8) \quad E(g(X)) := \int_{-\infty}^{\infty} g(x) \cdot f(x) dx =: \int g(x) \cdot f(x) dx,$$

sofern $\int |g(x)| \cdot f(x) dx$ endlich ist. Man sagt dann wieder: $E(g(X))$ **existiert**.

Spezialfälle sind:

$$(10.10) \quad EX = E(X) = \int x \cdot f(x) dx,$$

$$(10.11) \quad E(X^2) = \int x^2 \cdot f(x) dx.$$

Definition. Existiert $E(X^2)$ in der Situation von Def. 10.10, so definiert man die **Varianz** von X als

$$(10.12) \quad \text{Var}(X) := E((X - \mu)^2) = E(X^2) - (EX)^2 \quad \text{mit } \mu = EX.$$

Die letzte Gleichung folgt wie in §3.5: $E((X - \mu)^2) = \int (x - \mu)^2 f(x) dx = \int (x^2 - 2\mu x + \mu^2) \cdot f(x) dx$
 $= \int x^2 \cdot f(x) dx - 2\mu \int x \cdot f(x) dx + \mu^2 \int f(x) dx = E(X^2) - 2\mu \cdot EX + \mu^2 = E(X^2) - \mu^2.$

Beispiel 10.11. (*Normalverteilung*).

Sei X Zva mit einer (Standard-) Normalverteilung und φ die Dichte. Dann gilt

$(-x) \cdot \varphi(-x) = -x \cdot \varphi(x)$. Damit folgt: $\int_{-\infty}^0 x \cdot \varphi(x) dx = - \int_0^{\infty} x \cdot \varphi(x) dx$ und somit

$$EX = \int_{-\infty}^{\infty} x \cdot \varphi(x) dx = 0.$$

Weiter ergibt sich unter Benutzung von partieller Integration

$[uv' = (uv)' - u'v$ mit $u(x) = x$, $v(x) = -\exp(-\frac{1}{2} x^2)$]:

$$E(X^2) = \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-N}^N x^2 \exp(-\frac{1}{2} x^2) dx = \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left[\int_{-N}^N u(x) \cdot v'(x) dx \right]$$

$$= \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left[[u \cdot v]_{-N}^N + \int_{-N}^N u'(x) \cdot v(x) dx \right] = \lim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-N}^N \exp(-\frac{1}{2} x^2) dx = 1$$

wegen $\lim_{N \rightarrow \infty} N \cdot \exp(-\frac{1}{2} N^2) = 0$, also

$$\text{Var}(X) = 1. \quad \square$$

Lemma 10.12. Für $\alpha, \beta \in \mathbb{R}$ gilt:

- (a) $E(\alpha \cdot X + \beta) = \alpha \cdot EX + \beta$, falls EX existiert.
 (b) $\text{Var}(\alpha \cdot X + \beta) = \alpha^2 \cdot \text{Var}(X)$, falls $E(X^2)$ existiert.

Der **Beweis** ist analog zu den Beweisen der Sätze 3.7 und 3.13.

Definition 10.13. Hat die Zva Y eine Standard–Normalverteilung, so heißt die Verteilung von $X = \sigma \cdot Y + \mu$ mit $\sigma \neq 0, \mu \in \mathbb{R}$, **$N(\mu, \sigma^2)$ –Verteilung** oder **Normalverteilung** mit Erwartungswert μ und Varianz σ^2 . Eine Standard–Normalverteilung ist also eine $N(0,1)$ –Verteilung.

Eine Rechtfertigung für diese Bezeichnung liefert der folgende Satz, der auch die Beliebtheit der Normalverteilung erklärt.

Satz 10.14. Hat die Zva X eine $N(\mu, \sigma^2)$ –Verteilung, so gilt:

- (a) $E(X) = \mu, \text{Var}(X) = \sigma^2$;
 (b) X hat die Dichte: $\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(- (x-\mu)^2/2\sigma^2)$.
 (c) $\alpha \cdot X + \beta$ hat eine $N(\alpha \cdot \mu + \beta, \alpha^2 \cdot \sigma^2)$ –Verteilung für $\alpha \neq 0, \beta \in \mathbb{R}$.
 (d) Ist die Zva Z unabhängig von X und hat eine $N(\nu, \tau^2)$ –Verteilung, so hat $X + Z$ eine $N(\mu+\nu, \sigma^2+\tau^2)$ –Verteilung.

Beweis. Sei $X = \sigma \cdot Y + \mu$ und Y habe die Dichte φ .

a) Wir wissen nach dem obigen Beispiel 10.11: $EY = 0, \text{Var}(Y) = 1$. Jetzt kann Lemma 10.12 angewendet werden.

b) Dieser Teil folgt direkt aus Satz 10.6.

c) Es ist $\alpha \cdot X + \beta = (\alpha \cdot \sigma) \cdot Y + (\alpha \cdot \mu + \beta)$. Jetzt kann auf Def. 10.13 zurückgegriffen werden.

d) Seien $Z = \tau \cdot Y' + \nu$ sowie Y und Y' unabhängig und standard–normalverteilt. Dann haben wir die obige Situation, und es gilt mit $c := \sqrt{\sigma^2 + \tau^2}$:

$$X + Z = \sigma \cdot Y + \mu + \tau \cdot Y' + \nu = c \cdot \left(\frac{\sigma}{c} Y + \frac{\tau}{c} Y'\right) + \mu + \nu.$$

Nach Satz 10.9 hat $\frac{\sigma}{c} Y + \frac{\tau}{c} Y'$ wieder eine Standard–Normalverteilung. Damit hat nach Definition 10.13 $X + Z$ eine $N(\mu+\nu, c^2)$ –Verteilung. \square

Beispiel 10.15. (*Gleichverteilung*). Hat die Zva X eine Gleichverteilung auf $[a, b]$, so gilt:

$$EX = \frac{1}{2}(a+b) \text{ (Mittelpunkt von } [a, b]), \text{Var}(X) = \frac{1}{12} \cdot (b-a)^2.$$

$$\text{Es ist } E(X^k) = \int_a^b x^k \cdot \frac{1}{b-a} dx = \frac{1}{k+1} \frac{b^{k+1} - a^{k+1}}{b-a}, \text{ also}$$

$$EX = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{1}{2}(a+b), (EX)^2 = \frac{1}{4}(a^2 + 2ab + b^2),$$

$$E(X^2) = \frac{1}{3} \frac{b^3 - a^3}{b - a} = \frac{1}{3} (b^2 + ab + a^2),$$

$$\text{Var}(X) = E(X^2) - (EX)^2 = \frac{1}{12} (b^2 - 2ab + a^2). \quad \square$$

Beispiel 10.16. (*Exponentialverteilung*).

Hat X eine Exponentialverteilung mit Parameter λ , so gilt:

$$EX = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Der Beweis folgt mit $z = \lambda x$, $\lambda dx = dz$:

$$E(X^k) = \int_0^\infty x^k \lambda e^{-\lambda x} dx = \lambda^{-k} \cdot k! \cdot \int_0^\infty \frac{1}{k!} \lambda^{k+1} x^k e^{-\lambda x} dx = \lambda^{-k} \cdot k!,$$

denn $f(x) = \frac{1}{k!} \lambda^{k+1} x^k e^{-\lambda x}$ ist eine Dichte, und zwar die Dichte einer **Erlang**- oder **Γ -Verteilung**. Also ergibt sich:

$$E(X) = \frac{1}{\lambda}, \quad E(X^2) = \frac{2}{\lambda^2}. \quad \square$$

Wir erweitern die Definition des Erwartungswertes jetzt noch auf Funktionen von mehreren Zva. Dabei beschränken wir uns hier auf zwei unabhängige Zva X_1, X_2 .

Bei diskreten W -Räumen hatten wir nach Satz 3.8 die folgende Formel

$$E(g(X_1, X_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) P((X_1, X_2) = (x_1, x_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) P(X_1 = x_1) \cdot P(X_2 = x_2)$$

im Falle der Unabhängigkeit von X_1, X_2 .

Definition 10.10 (Fortsetzung). Seien X_i unabhängige, reelle Zva auf (Ω, P) mit Dichten f_i , $i = 1, 2$. Ferner sei g eine Borel-Funktion auf \mathbb{R}^2 . Dann wird der **Erwartungswert** $E(g(X_1, X_2))$ definiert als

$$(10.13) \quad E(g(X_1, X_2)) := \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x_1, x_2) \cdot f_1(x_1) \cdot f_2(x_2) dx_2 \right] dx_1,$$

sofern $\int \int |g(x_1, x_2)| \cdot f_1(x_1) \cdot f_2(x_2) dx_2 dx_1$ endlich ist.

Bemerkung. Die Definition des Erwartungswerts in § 3.3 und hier in §10 sind alle konsistent. In der Wahrscheinlichkeitstheorie, die auf der Maßtheorie aufbaut, kann man eine Definition von $E(X)$ geben, sodaß sich die hier gegebenen Definitionen als Folgerungen ergeben. \square

Satz 10.17. Seien X_i unabhängige, reelle Zva auf (Ω, P) mit Dichten f_i , $i = 1, 2$, sodaß deren Erwartungswert existiert. Dann gilt:

(a) $E(X_1 + X_2) = E(X_1) + E(X_2)$

oder allgemeiner für die Zva $Y_i = g_i(X_1, X_2)$ mit Borel-Funktionen g_i auf \mathbb{R}^2 und $c_i \in \mathbb{R}$:

$$E(c_1 Y_1 + c_2 Y_2) = c_1 \cdot E(Y_1) + c_2 \cdot E(Y_2), \text{ falls } E[Y_i] \text{ existiert.}$$

(b) $E(X_1 \cdot X_2) = E(X_1) \cdot E(X_2)$.

Beweis. a) Es ist $E[\sum c_i \cdot g_i(X_1, X_2)] = \iint [\sum c_i \cdot g_i(x_1, x_2)] \cdot f_1(x_1) \cdot f_2(x_2) dx_2 dx_1$
 $= \sum c_i \cdot \iint g_i(x_1, x_2) \cdot f_1(x_1) \cdot f_2(x_2) dx_2 dx_1 = \sum c_i \cdot E[g_i(X_1, X_2)]$.

b) Es ist $E(X_1 \cdot X_2) = \iint x_1 \cdot x_2 \cdot f_1(x_1) \cdot f_2(x_2) dx_2 dx_1$
 $= \int x_1 \cdot f_1(x_1) \cdot \left[\int x_2 \cdot f_2(x_2) dx_2 \right] dx_1 = \int x_1 \cdot f_1(x_1) \cdot dx_1 \cdot \int x_2 \cdot f_2(x_2) dx_2 = EX_1 \cdot EX_2$. \square

Bemerkung. Die Unabhängigkeit von $X_1 + X_2$ ist für $E(X_1 + X_2) = E(X_1) + E(X_2)$ viel zu stark und wurde gemacht, weil hier nur für diesen Fall ein W-Modell für X_1, X_2 angegeben wurde. In der Tat sind ja Y_1, Y_2 i.a. abhängig. Die Linearität des Erwartungswertes wurde hier also gezeigt für Zva, die sich als Funktionen von zwei unabhängigen Zva schreiben lassen. Damit erfaßt man schon viele Fälle aus der Praxis. \square

Korollar 10.18. Seien X_i unabhängige, reelle Zva auf (Ω, P) mit Dichten f_i , sodaß $E(X_i^2)$, $i = 1, 2$ existiert. Dann gilt:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2).$$

Beweis. Sei $\mu_i := EX_i$, dann sind die Voraussetzungen auch für $\tilde{X}_i := X_i - \mu_i$ erfüllt mit

$$E(\tilde{X}_i) = 0, E(\tilde{X}_i^2) = \text{Var}(X_i), E((\tilde{X}_1 + \tilde{X}_2)^2) = \text{Var}(X_1 + X_2).$$

Dann ist

$$\begin{aligned} \text{Var}(X_1 + X_2) &= E[(\tilde{X}_1 + \tilde{X}_2)^2] = E[\tilde{X}_1^2 + 2 \cdot \tilde{X}_1 \cdot \tilde{X}_2 + \tilde{X}_2^2] \\ &= E(\tilde{X}_1^2) + 2 \cdot E(\tilde{X}_1 \cdot \tilde{X}_2) + E(\tilde{X}_2^2) = E(\tilde{X}_1^2) + 2 \cdot E(\tilde{X}_1) \cdot E(\tilde{X}_2) + E(\tilde{X}_2^2) \\ &= E(\tilde{X}_1^2) + E(\tilde{X}_2^2) = \text{Var}(X_1) + \text{Var}(X_2). \end{aligned}$$

Dabei wurde Satz 10.17a benutzt mit $g_i(x_1, x_2) = x_i^2$, $i=1, 2$, $g_3(x_1, x_2) = 2x_1 x_2$. \square

Beispiel. Seien X, Z unabhängige, reelle Zva auf (Ω, P) .

Ist $X \sim N(\mu, \sigma^2)$ -verteilt und $Z \sim N(\nu, \tau^2)$ -verteilt, so stehen Satz 10.17a und Korollar 10.18 in Übereinstimmung mit Satz 10.14a,d, d.h.

$$\mu + \nu = E(X+Z), \quad \sigma^2 + \tau^2 = \text{Var}(X+Z), \quad X+Z \text{ hat eine } N(\mu+\nu, \sigma^2+\tau^2)\text{-Verteilung.}$$

§11 Grenzwertsätze.

Das schwache Gesetz der großen Zahl (vgl. §3.6) und der zentrale Grenzwertsatz (vgl. §5) gelten in gleicher Form auch für Zva mit Dichten.

Startet man bei zentralen Grenzwertsatz bereits mit normalverteilten Zva, so kann man die Grenzwertaussage durch eine Gleichheit schon bei endlichem n ersetzen.

Sind also X_1, \dots, X_n unabhängige, identische verteilte Zva, sodaß die Verteilung von X_m eine $N(\mu, \sigma^2)$ -Verteilung ist, so hat nach Satz 10.14d (+ Induktion nach n) auch S_n eine Normalverteilung. S_n^* ist die zugehörige standardisierte Zva, entsteht also aus S_n durch Anwendung einer lineare Transformation. Nach Satz 10.14c hat auch S_n^* eine Normalverteilung. Die lineare Transformation ist gerade so gewählt worden, daß gilt:

$$E(S_n^*) = 0, \quad \text{Var}(S_n^*) = 1.$$

Also hat S_n^* eine $N(0, 1)$ -Verteilung und es gilt:

$$(11.0) \quad P(S_n^* \leq t) = \Phi(t) \quad \forall t \in \mathbb{R}.$$

Diese Überlegung führt dazu, daß man in der Statistik unter Normalverteilungsannahmen asymptotische Aussagen ersetzen kann durch Aussagen für beliebige (auch kleine) Stichprobenumfänge n .

§12 / 13 Statistik unter Normalverteilungsannahmen.

Die Normalverteilung wurde als erste zur Beschreibung von Abweichungen von einem Sollwert benutzt, obgleich die Symmetrie und der gesamte Raum \mathbb{R} in praktischen Problemen oft nicht gerechtfertigt sind. Die Normalverteilung hieß ursprünglich **Fehlergesetz** und entstand aus der Arbeit der Astronomen und Mathematiker des 18. und 19. Jahrhundert bei dem Bemühen, Beobachtungsfehler zu berücksichtigen.

Wir werden wieder **Einstichprobenprobleme** betrachten, bei denen n Einzelexperimente unter gleichen Bedingungen durchgeführt werden. Dabei beschreibt X_m das Ergebnis des m -ten Einzelexperiment. Die Einzelexperiment werden stets so ausgeführt, daß sie sich nicht gegenseitig beeinflussen und somit durch unabhängige Zva beschrieben werden können. Die Unsicherheit über das W -Maß, das den Versuch am besten beschreibt, wird wieder durch den Parameter $\vartheta \in \Theta$ beschreiben. Genauer haben wir die folgende Situation. Gegeben ist:

- ein Familie $\{f_\vartheta; \vartheta \in \Theta\}$ von Dichten auf \mathbb{R} (die **Verteilungsannahme**),
 Θ ist wieder unser Parameterraum.

Unser **Stichprobenraum** ist jetzt der $\mathfrak{X} = \mathbb{R}^n$ [, also $\mathfrak{X}' = \mathbb{R}$]. Es sei $X_m : \mathbb{R}^n \rightarrow \mathbb{R}$ jeweils wieder die Projektion auf die m -te Koordinate. Unter $\vartheta \in \Theta$ sind dann X_1, \dots, X_n unabhängige, identische verteilte Zva mit der Dichte f_ϑ [, also ist P'_ϑ das durch f_ϑ definierte Maß auf $\mathfrak{X}' = \mathbb{R}$].

Beispiel. Eine Maschine füllt Zucker in Tüten ab. Das tatsächliche Tütengewicht schwankt um den Mittelwert ϑ mit einer Varianz $\sigma_0^2 > 0$, wobei ϑ der Justierung und σ_0^2 einer bekannten Maschinenkonstante entspricht. Zur Kontrolle der Justierung wird eine Stichprobe $x = (x_1, \dots, x_n)$ vom Umfang n entnommen. Hierbei sei x_m das Gewicht der m -ten Tüte. Hier sei etwa Θ ein Intervall in $(0, \infty)$.

Als **Verteilungsannahme** $\{f_{\vartheta}; \vartheta \in \Theta\}$ betrachten wir von nun an

$$f_{\vartheta}(t) = \varphi_{\vartheta, \sigma_0^2}(t) = \frac{1}{\sigma_0} \cdot \varphi\left(\frac{t-\vartheta}{\sigma_0}\right) = \frac{1}{\sqrt{2\pi} \sigma_0} \exp\left(-\frac{(t-\vartheta)^2}{2\sigma_0^2}\right),$$

also gemäß Satz 10.14 Dichten von Normalverteilungen mit unbekanntem Erwartungswert ϑ und bekannter Varianz σ_0^2 . Es soll nun getestet werden, ob der tatsächliche Erwartungswert ϑ mit einem Sollwert μ (z.B. $\mu = 500$ g) übereinstimmt.

Gewisse Abweichungen vom Sollwert μ muß man immer in Kauf nehmen; so beschreibt man ja auch die gemessenen Werte bei den Einzelerperimenten durch Zufallsvariable, nämlich durch X_m . Dazu betrachten wir die

$$\text{Hypothese } H := \{\vartheta \in \Theta; \vartheta = \mu\} = \{\mu\}.$$

Das Verwerfen von H soll wieder statistisch abgesichert werden. Im Beispiel betrachten wir es also als schwerwiegend, daß man sich gegen H entscheidet, obwohl H richtig ist, d.h. daß man die Maschine anhält und neu justiert, obwohl dies garnicht nötig ist. Dies wäre der **Fehler (1. Art)**.

Als naheliegende Klasse von Tests betrachtet man jetzt solche mit den **Verwerfungsbereichen**

$$R_c := \{|\bar{X} - \mu| > c\}, \quad c > 0,$$

wobei $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ wieder das **Stichprobenmittel** ist. Der Test mit dem Verwerfungsbereich R_c hat dann die Fehler-W.

$$P_{\mu}(|\bar{X} - \mu| > c).$$

Wir lassen nur **Tests zum Niveau α** zu, für die also gilt:

$$P_{\mu}(|\bar{X} - \mu| > c) \leq \alpha.$$

Wir wollen nun $P_{\mu}(|\bar{X} - \mu| > c)$ berechnen. Unter $\vartheta = \mu$ sind X_1, \dots, X_n unabhängig und $N(\mu, \sigma_0^2)$ -verteilt. Damit ist nach Satz 10.14d: $X_1 + \dots + X_n$ $N(n\mu, n\sigma_0^2)$ -verteilt.

Nach Satz 10.14c ist \bar{X} nun $N(\mu, \frac{1}{n}\sigma_0^2)$ -verteilt (vgl. auch das schwache Gesetz der großen Zahl)

und $\sqrt{n} \frac{\bar{X} - \mu}{\sigma_0}$ schließlich $N(0,1)$ -verteilt (vgl. (11.0)). Damit ergibt sich:

$$P_{\mu}(|\bar{X} - \mu| > c) = P_{\mu} \left[\sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} < -\sqrt{n} \frac{c}{\sigma_0} \right] + P_{\mu} \left[\sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} > \sqrt{n} \frac{c}{\sigma_0} \right] = 2 \cdot (1 - c')$$

$$\text{mit } c' := \sqrt{n} \frac{c}{\sigma_0}.$$

Die Forderung, daß R_c einen Test z.N. α bestimmt, ist gleichbedeutend mit

$$2 \cdot (1 - \Phi(c')) \leq \alpha \Leftrightarrow c' \geq c^*.$$

Dabei wählen wir wieder

$$c^* \text{ als } \frac{1}{2}\alpha\text{-Fraktile der } N(0,1)\text{-Verteilung, d.h. } 1 - \Phi(c^*) = \int_{c^*}^{\infty} \varphi(t) dt = \frac{1}{2}\alpha.$$

Will man die Fehlerschranke α voll ausschöpfen, so muß man c und damit c' möglichst klein machen. Der zugehörige Verwerfungsbereich ist dann mit $[c^* =: \sqrt{n} \frac{c}{\sigma_0}]$:

$$R^* := \{ |\bar{X} - \mu| > \sigma_0 \cdot c^*/\sqrt{n} \} = \{ |T| > c^* \} \quad \text{mit } T := \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0}.$$

Satz. Es liege die obige Situation vor mit dem durch $H = \{\mu\}$ gegebenen zweiseitigen Testproblem. Dann wählt man den sogenannten **zweiseitiger Gaußtest** mit dem Verwerfungsbereich R^* . Dabei ist c^* das $\frac{1}{2}\alpha$ -Fraktile der $N(0,1)$ -Verteilung.

Der zweiseitige Gaußtest ergab sich in § 6 auch für große n . Hier wurde aber der zentrale Grenzwertsatz nicht benutzt, so daß der zweiseitige Gaußtest in der gegebenen Situation auch für kleine n ein Test z.N. α ist.

Als **Zahlenbeispiel** für obiges Beispiel betrachten wir: $n = 7$, $\sigma_0 = 6$, $\mu = 500$, $\alpha = 0,05$ und die Stichprobe $x = (x_1, \dots, x_7) = (514, 497, 507, 499, 510, 501, 508)$.

$$\text{Dann ist } \bar{X}(x) = \bar{x} = 505,14, \quad T(x) = \sqrt{7} \cdot \frac{5,14}{6} = 2,3.$$

Aus einer Tabelle entnimmt man für das $\frac{1}{2}\alpha$ -Fraktile c^* der $N(0,1)$ -Verteilung mit $\frac{1}{2}\alpha = 0,025$:
 $c^* = 1,960$.

Also liegt dann x im Verwerfungsbereich. Hierbei ist allerdings α relativ hoch. Daran liegt es, daß man trotz geringen Stichprobenumfangs zu einer Ablehnung der Hypothese kommt.

Die Antwort lautet also:

Z.N. $\alpha = 0,05$ ist **statistisch gesichert**, daß der Erwartungswert ϑ vom Sollwert $\mu = 500$ abweicht.

Oder: Die Abweichung des Erwartungswert ϑ vom Sollwert $\mu = 500$ ist **signifikant** z.N. $\alpha = 0,1$.

Wählt man dagegen $\alpha = 0,01$ und damit $\frac{1}{2}\alpha = 0,005$, so ergibt sich $c^* = 2,576 > T(x)$.

Will man also 99% Sicherheit, so kann man die Hypothese noch nicht ablehnen. Man übt gewissermaßen Stimmenthaltung.

Bemerkung. Ist die Varianz σ^2 in obigem Testproblem unbekannt, ist $\vartheta = (\mu, \sigma^2)$, so ersetzt man σ in der Statistik T durch einen Schätzer, nämlich durch die Wurzel S aus der Stichprobenstreuung

S^2 (vgl. (4.4)), also $T = \sqrt{n} \frac{\bar{X} - \mu}{S}$. Das Gleiche hatten wir im Zusatz zum zentralen Grenzwertsatz in §4.7 gemacht. Dann muß man (zumindest für kleine n) das $\frac{1}{2}\alpha$ -Fraktile der Normalverteilung ersetzen durch das

$\frac{1}{2}\alpha$ -Fraktile der sogenannten **Studentschen t_{n-1} -Verteilung**. \square

Konfidenzintervalle.

Es soll nun noch eine Anwendung auf **Konfidenzintervalle für den Erwartungswert einer Normalverteilung** erfolgen.

Seien X_1, \dots, X_n wieder wie oben unabhängige Zva mit einer $N(\vartheta, \sigma_0^2)$ -Verteilung, $\vartheta \in \Theta$, Θ ein Intervall in \mathbb{R} . Dazu betrachten wir den Tests mit dem Verwerfungsbereich

$$R^*(\mu) := \left\{ \sqrt{n} \frac{|\bar{X} - \mu|}{\sigma_0} > c^* \right\}, \quad c^* \text{ das } \frac{1}{2}\alpha\text{-Fraktile der } N(0,1)\text{-Verteilung,}$$

also den zweiseitigen Gaußtest. Dieser ist nach obigen Überlegungen ein Test z.N. α für das Testen von $H(\mu) : \vartheta = \mu$. Wir verfolgen das folgende Prinzip:

Das **Prinzip** (Korrespondenzprinzip) bei der Konstruktion eines Konfidenzintervalls $C(x)$ aus einer Familie von Tests besteht bei Vorliegen der Stichprobe x darin:

Nimm einen Parameter μ in $C(x)$ auf, wenn die Hypothese $H = \{\mu\}$ nicht verworfen wird.

Dabei ergibt sich

$$C(x) = \left\{ \mu; \sqrt{n} \frac{|\bar{X} - \mu|}{\sigma_0} \leq c^* \right\} = \left\{ \mu; \bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon \right\} = [\bar{X} - \varepsilon, \bar{X} + \varepsilon].$$

Dabei gilt: $1 - P_\mu \left[\sqrt{n} \frac{|\bar{X} - \mu|}{\sigma_0} > c^* \right] = P_\mu \left[\sqrt{n} \frac{|\bar{X} - \mu|}{\sigma_0} \leq c^* \right] \geq 1 - \alpha$ (sogar $= 1 - \alpha$) $\forall \mu$

$\Leftrightarrow P_\mu [\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon] \geq 1 - \alpha$ (sogar $= 1 - \alpha$) $\forall \mu$ mit $\varepsilon := \sigma_0 \cdot c^* / \sqrt{n}$.

Somit ist $[\bar{X} - \varepsilon, \bar{X} + \varepsilon]$ ein Konfidenzintervall mit

$$P_\mu [[\bar{X} - \varepsilon, \bar{X} + \varepsilon] \ni \mu] \geq 1 - \alpha \quad (\text{sogar} = 1 - \alpha) \quad \forall \mu \in \Theta.$$

Egal welches $\vartheta = \mu$ der richtige Parameter ist, das Konfidenzintervall $[\bar{X} - \varepsilon, \bar{X} + \varepsilon]$ überdeckt diesen richtigen Parameter μ mit der W. $1 - \alpha$. Eine andere Schreibweise wäre

$$\vartheta = \mu = \bar{X} \pm \varepsilon.$$

Man erkennt hier die Abhängigkeit der Abweichung ε von α , σ_0 und n .

§8 Entropie und Kodierung

Die Kodierung einer Nachricht kann der Verschlüsselung oder der besseren Übertragung dienen. Wir werden hier mehr auf den letzteren Aspekt eingehen. Dabei interessieren Kodierungen von Wörtern x , die mit möglichst geringen Codewortlängen $n(x)$ auskommen.

Sei \mathfrak{X} eine (nicht leere) endliche Menge von **Wörtern** und $A = \{0,1\}$ das **Alphabet** der Kodierung.

Allgemeiner läßt man auch beliebige endliche Mengen für das Alphabet zu; wir wollen uns aber auf diesen wichtigen Fall beschränken.

Ein **Codewort** ist (nicht leere) endliche 01-Folge.

Ein **Code** ist eine eindeutige Abbildung c von \mathfrak{X} in die Menge der Codewörter; dabei ist die Anzahl $n(x)$ von Nullen und Einsen des Codewortes $c(x)$ die **Codewortlänge**.

Besteht z.B. \mathfrak{X} aus 4 Wörtern, ist also o.E. $\mathfrak{X} = \{1,2,3,4\}$, so ist die folgende Abbildung ein Code.

$$(8.2) \quad \begin{aligned} 1 &\mapsto c(1) = 0 \\ 2 &\mapsto c(2) = 10 \\ 3 &\mapsto c(3) = 110 \\ 4 &\mapsto c(4) = 111 \end{aligned}$$

Der Code (8.2) ist übrigens nicht die Dualdarstellung der Zahlen 1,2,3,4. Eine weiteren Code erhält man, wenn man $x \in \mathfrak{X}$ die Dualdarstellung von $x-1$ zuordnet. Dann erhält man:

$$(8.2)^{\circ} \quad \begin{aligned} 1 &\mapsto c(1) = 00 \\ 2 &\mapsto c(2) = 01 \\ 3 &\mapsto c(3) = 10 \\ 4 &\mapsto c(4) = 11. \end{aligned}$$

Besteht eine **Nachricht** aus einer Folge x_1, x_2, \dots, x_k von Wörtern, so kann man sie kodieren, indem man einfach die Codewörter $c(x_1), c(x_2), \dots, c(x_k)$ hintereinanderschreibt. Zur Folge 2,1,4 würde nach dem Code (8.2) dann die Folge 100111 gehören.

Die Codes (8.2) und (8.2)^o haben eine angenehme Eigenschaft: Kein Codewort ist Anfangsstück eines anderen Codewortes. Solche Codes heißen **Präfixcodes**. Hat ein Code diese Eigenschaft, so läßt sich aus der kodierten Nachricht die Nachricht eindeutig zurückgewinnen: Ist die 01-Folge a_1, \dots, a_2 die kodierte Nachricht, so ist genau eine der Teilfolgen a_1, \dots, a_1 ein Codewort $c(x_1)$. Man notiert x_1 und sucht in der verkürzten Folge a_{i+1}, \dots, a_k mit dem gleichen Verfahren das Codewort $c(x_2)$, usw.

Beispiel. Wir betrachten ein weiteres Beispiel und nehmen an, daß ein Wort x aus einer Folge von insgesamt $n = 1000$ Nullen und Einsen [allgemeiner $n \leq 1024$] besteht. [Hier sollen also auch die Worte bereits durch Nullen und Einsen beschrieben werden.] Kommen dann in dem Wort wesentlich weniger Einsen als Nullen vor, so kann man eine Kodierung wählen, indem man nur die Stellen $k_1 < k_2 < \dots < k_r$ angibt, an denen eine Eins steht. Diese wiederum kann man durch die Dualdarstellung a_0, \dots, a_9 beschreiben gemäß:

$$k_i = \sum_{j=0}^9 a_j(k_i) \cdot 2^j + 1.$$

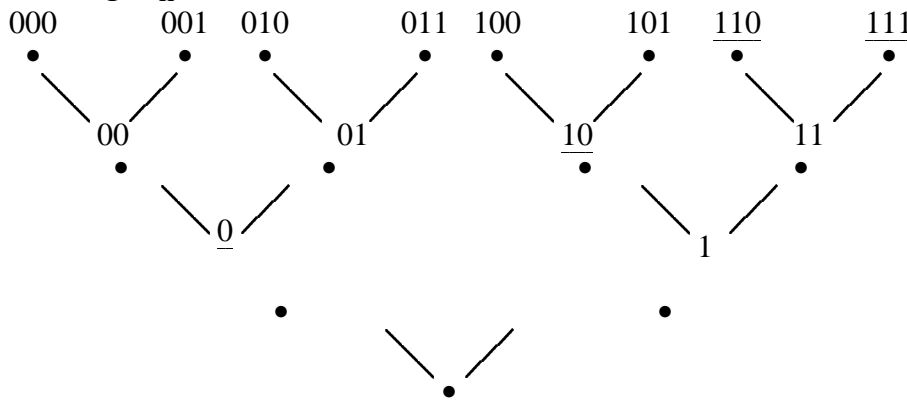
Dann erhält man das Codewort

$$(8.1) \quad c(x) = a_0(k_1) \dots a_0(k_1) a_0(k_2) \dots a_0(k_2) \dots a_0(k_r) \dots a_0(k_r).$$

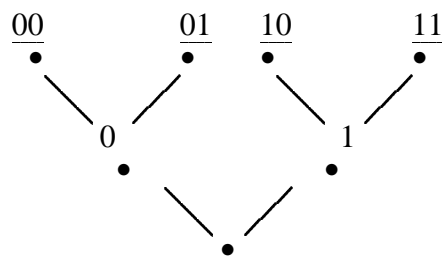
Die Länge dieses Codewortes hängt dann von der Lage der Einsen ab. Um einen Präfixcode zu bekommen, kann man das Ende des Codewortes dadurch signalisieren, daß man z.B zehn Einsen an die Folge (8.1) anhängt. Die Folge von zehn Einsen kommt ja unter den $a(k)$ nicht vor.

Fassen wir dabei das Auftreten von Nullen und Einsen als eine Folge von 1000 Bernoulli-Experimenten auf mit Erfolgs-W. $p = 0,01$ (vgl. §2.4), so ist die mittlere Anzahl der Erfolge gerade $np = 1000 \cdot 0,01 = 10$ (vgl. §3.3). (Die Unabhängigkeit der Bernoulli-Experimente wird dabei noch nicht benötigt.) Damit ist die erwartete Länge des Codewortes 100. Zählt man die 10 Einsen hinzu, so kommt man auf 110 Symbole, um ein Wort der Länge 1000 zu übermitteln. \square

Einen Präfixcode, dessen maximale Codewortlänge N ist, kann man konstruieren, indem man die Folgen $a_1 \dots a_n$ mit $n \leq N$ in Form eines Baumes hinschreibt.



Codewortbaum für (8.2).



Codewortbaum für (8.2)^o.

Die Codes (8.2) und (8.2)^o bestehen aus den unterstrichenen 01-Folgen. Die Präfix-Eigenschaft bedeutet gerade, daß, wenn ein Verzweigungspunkt als Codewort gewählt wird, an den von diesem Punkt ausgehenden Ästen keine weiteren Codewörter mehr vorkommen.

Wir fragen jetzt: Wie groß müssen die Codewortlängen $n(x)$ sein, damit man dazu einen Präfixcode erhalten kann?

Satz 8.1 (Ungleichung von Kraft).

Einen Präfixcode mit Wortlängen $n(x)$, $x \in \mathfrak{X}$, gibt es genau dann, wenn gilt:

$$(8.3) \quad \sum_{x \in \mathfrak{X}} \frac{1}{2^{n(x)}} \leq 1.$$

In den Beispielen (8.2), (8.2)^o haben wir:

$$\sum_{x \in \mathfrak{X}} \frac{1}{2^{n(x)}} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = 1 \quad \text{und} \quad \sum_{x \in \mathfrak{X}} \frac{1}{2^{n(x)}} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1.$$

Die Schranke (8.3) wurde also voll ausgeschöpft; man kann keine Länge $n(x)$ verkleinern und die anderen unverändert lassen.

Beweis von Satz 8.1: Wir schreiben $\mathfrak{X} = \{x_1, \dots, x_M\}$ und $n(x_m) = n_m$. Dann haben wir

$$\sum_{x \in \mathfrak{X}} \frac{1}{2^{n(x)}} = \sum_{m=1}^M \frac{1}{2^{n_m}}.$$

Zuerst zeigen wir, daß (8.3) hinreichend ist. Dabei können wir $n_1 \leq n_2 \leq \dots \leq n_M$ annehmen. Dann ist $n^* := n_M$ die maximale Wortlänge. Wir sagen, eine Folge mit n Elementen liegt im Baum auf Höhe n .

Sei $c(x_1)$ die Folge, die aus n_1 Nullen besteht. Diese liegt auf Höhe n_1 ganz links. Durch Wahl dieses Codewortes sind für jedes $h \geq n_1$ gerade 2^{h-n_1} der auf Höhe h liegenden Punkte von der Wahl als Codewort ausgeschlossen, und zwar diejenigen 2^{h-n_1} Punkte, die am weitesten links liegen. Als $c(x_2)$ wählt man jetzt den am weitesten links liegenden noch verfügbaren Punkt der Höhe n_2 . Nun sind für $h \geq n_2$ die $2^{h-n_1} + 2^{h-n_2}$ am weitesten links liegenden Punkte der Höhe h für die weitere Wahl von Codewörtern ausgeschlossen. So macht man weiter.

Ist $m < M$ und sind m Codeworte auf die angegebene Weise bestimmt, so sind für $h \leq n_m$ (z.B. $h = n_{m+1}$) die am weitesten links liegenden $2^{h-n_1} + 2^{h-n_2} + \dots + 2^{h-n_m}$ Punkte der Höhe h "verboten".

Wegen (8.3) gilt dabei: $2^{h-n_1} + 2^{h-n_2} + \dots + 2^{h-n_m} = 2^h \cdot (2^{-n_1} + 2^{-n_2} + \dots + 2^{-n_m}) < 2^h$. Es muß also mindestens einer noch verfügbar sein. Man wählt dann auf Höhe n_{m+1} den am weitesten links liegenden noch verfügbaren Punkt. Die zu diesem Punkt gehörende Folge $a_1 \dots a_{n_{m+1}}$ wird als Codewort $c(x_{m+1})$ gewählt.

Die andere Richtung des Beweises ist noch einfacher. Ein Codewort der Länge n_m gehört zu einem Punkt der Höhe n_m . An den von diesem Punkt ausgehenden Ästen gibt es $2^{n^*-n_m}$ Punkte zur maximalen Höhe n^* . Da es sich um einen Präfixcode handelt, müssen die auf diese Weise verschiedenen Codewörtern zugeordneten Punkte maximaler Höhe verschieden sein. Da es 2^{n^*} Punkte maximaler Höhe gibt, gilt $\sum_{m=1}^M 2^{n^*-n_m} \leq 2^{n^*}$ und damit (8.3). \square

Wir wollen nun die **mittlere Codewortlänge** eines Präfixcodes abschätzen. Dazu versehen wir jedes Wort $x \in \mathfrak{X}$ mit der **relativen Häufigkeit** $P(x)$, mit der es gesendet wird. Dabei ist $P(x)$ eine W-Funktion auf \mathfrak{X} . Die mittlere Codewortlänge ist dann

$$E[n] := \sum_{x \in \mathfrak{X}} n(x) \cdot P(x),$$

die auch als Erwartungswert interpretiert werden kann. Es wird sich eine überraschende Beziehung zu einem wichtigen Begriff der W-Theorie ergeben.

Definition 8.3. Ist P ein W -Maß auf \mathfrak{X} , so heißt der folgende Ausdruck die **Entropie** von P :

$$H(P) := - \sum_{x \in \mathfrak{X}} P(x) \log_2 P(x) = \sum_{x \in \mathfrak{X}} P(x) \log_2 \frac{1}{P(x)}.$$

Dabei wird $0 \cdot \log_2 0 := 0 =: 0 \cdot \log_2 \frac{1}{0}$ gesetzt. In der Informationstheorie wird gewöhnlich der Logarithmus zur Basis 2 zugrundegelegt. Dies ist nur eine Normierungsfrage wegen $\log_b t = \log_2 2 \cdot \log_2 t$ (etwa für $b = e$ oder $b = 10$). Die Entropie kann als Maß für die Unbestimmtheit eines Experiments angesehen werden. Sie maximal für die Gleichverteilung auf \mathfrak{X} . Andererseits ist sie null, wenn $P(x_0) = 1$ gilt für ein $x_0 \in \mathfrak{X}$.

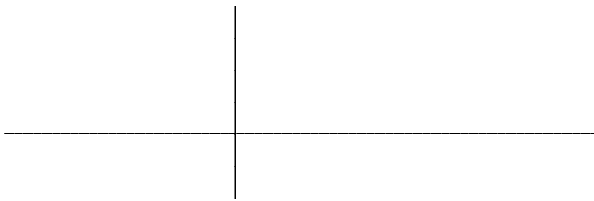
Als Vorbereitung kommt erst ein Lemma, das $H(P)$ als Lösung einer Optimierungsaufgabe zeigt.

Lemma 8.2. Sind P und Q W -Maße auf \mathfrak{X} , so gilt:

$$- \sum_{x \in \mathfrak{X}} P(x) \log P(x) \leq - \sum_{x \in \mathfrak{X}} P(x) \log Q(x)$$

und Gleichheit genau dann, wenn $P = Q$.

Beweis. Die Ungleichheit ist offenbar unabhängig von der Basis des Logarithmus. Wir wählen deshalb den natürlichen Logarithmus zur Basis $b = e$. Dann gilt bekanntlich wegen der strikten Konkavität von $\log t$: $\log(1+t) \leq t$ für $t > -1$ und $\log(1+t) = t$ nur für $t=0$.



Zeichne die Funktionen $f_1(t) = t$ und $f_2(t) = \log(1+t)$.

Wir schreiben \sum^* ... für die Summe $\sum_{x \in \mathfrak{X}, P(x) > 0} \dots$. Dann gilt:

$$(8.4) \quad \sum^* P(x) \log \frac{Q(x)}{P(x)} = \sum^* P(x) \log \left[1 + \frac{Q(x) - P(x)}{P(x)} \right] \leq \sum^* (Q(x) - P(x)) \leq 0$$

denn $\sum^* Q(x) \leq 1 = \sum^* P(x)$. Ist nun $Q(x) - P(x) \neq 0$ für ein x , so haben wir für ein x und damit für die gesamte Summe eine strikte Ungleichheit. \square

Satz 8.4 (Quellen-Kodierungssatz). Ist c ein beliebiger Präfixcode mit Codewortlängen $n(x)$, $x \in \mathfrak{X}$, so gilt: $E[n] \geq H(P)$. Es gibt einen Präfixcode c mit $E[n] \leq H(P) + 1$.

Beweis: Wir schreiben hier $\log := \log_2$. Ist c ein Präfixcode, so folgt aus der Ungleichung (8.3):

$$\sigma := \sum_{x \in \mathfrak{X}} \frac{1}{2^{n(x)}} \leq 1. \text{ Wir setzen: } Q(x) := \frac{1}{\sigma \cdot 2^{n(x)}}.$$

Dann definiert Q ein W -Maß auf \mathfrak{X} . Nach Lemma 8.2 gilt wegen $\sigma \leq 1$ $\left[\Leftrightarrow \log_2 \sigma \leq 0 \right]$:

$$H(P) \leq - \sum_{x \in \mathfrak{X}} P(x) \log_2 Q(x) = \sum_{x \in \mathfrak{X}} P(x) \cdot [n(x) + \log_2 \sigma] \leq E[n].$$

Dabei gilt übrigens:

$$H(P) = E[n] \Leftrightarrow " \sigma = 1 \text{ und } P = Q " \Leftrightarrow P(x) = \frac{1}{2^{n(x)}} \forall x \Leftrightarrow n(x) = - \log_2 P(x).$$

In der letzten Äquivalenzbeziehung ist " \Rightarrow " offensichtlich und " \Leftarrow " folgt aus $\sum P(x) = 1$.

Dabei bedeutet $\sigma = 1$, daß die Schranke 1 in (8.3) voll ausgeschöpft wird. Um die letzte Bedingung möglichst gut zu erfüllen, wählen wir $n(x)$, sodaß gilt:

$$n(x) \geq -\log_2 P(x) > n(x) - 1 \Leftrightarrow \frac{1}{2^{n(x)}} \leq P(x) < \frac{1}{2^{n(x)-1}}.$$

Es folgt dann: $\sum \frac{1}{2^{n(x)}} \leq \sum P(x) = 1$; also gibt es nach Satz 8.1 einen Präfixcode mit den Wortlängen $n(x)$. Dieser hat die Eigenschaft:

$$E[n] = \sum n(x) \cdot P(x) < \sum P(x) \cdot [1 - \log_2 P(x)] = 1 + H(P). \quad \square$$

Der Code (8.2) ist optimal im Sinne von $E[n] = H(P)$, wenn wir $P(1) = \frac{1}{2}$, $P(2) = \frac{1}{4}$ und $P(3) = \frac{1}{8} = P(4)$ vorliegen haben. Der Code (8.2)^o ist optimal für die Gleichverteilung P auf $\mathfrak{X} = \{1,2,3,4\}$.

Beispiel. Wir wollen noch einmal auf das obige Beispiel eingehen und nehmen an:

$$\begin{aligned} \mathfrak{X} &= \{0,1\}^{1000} \ni x = (x_1, \dots, x_{1000}), \\ P(x) &= \prod_{i=1}^{1000} p^{x_i} \cdot (1-p)^{1-x_i} \quad \text{mit } p = 0,01 \end{aligned}$$

(vgl. Beispiel 4.2). Ist $X_i : \mathfrak{X} \rightarrow \{0,1\}$ wieder die Projektion auf die i -te Koordinate, so sind die Zva X_1, \dots, X_n unabhängig, identisch verteilt mit einer $b(1,p)$ -Verteilung. Dann ist

$$\begin{aligned} H(P) &= - \sum_{x \in \mathfrak{X}} P(x) \sum_{i=1}^{1000} \log_2 [p^{x_i} \cdot (1-p)^{1-x_i}] \\ &= - \sum_{x \in \mathfrak{X}} P(x) \sum_{i=1}^{1000} [x_i \cdot \log_2 p + (1-x_i) \cdot \log_2 (1-p)] \\ &= - \sum_{i=1}^{1000} \sum_{x \in \mathfrak{X}} P(x) [x_i \cdot \log_2 p + (1-x_i) \cdot \log_2 (1-p)] \\ &= - \sum_{i=1}^{1000} E[X_i \cdot \log_2 p + (1-X_i) \cdot \log_2 (1-p)] \\ &= - 1000 \cdot E[X_1 \cdot \log_2 p + (1-X_1) \cdot \log_2 (1-p)] = 1000 \cdot H(P'). \end{aligned}$$

Dabei ist $H(P')$ die Entropie eines einzelnen Bernoulli-Experiments wegen

$$- E[X_1 \cdot \log_2 p + (1-X_1) \cdot \log_2 (1-p)] = - [p \cdot \log_2 p + (1-p) \cdot \log_2 (1-p)] = H(P').$$

Daß die Entropie von n unabhängigen Einzelexperimenten gleich der Summe der Entropien dieser Einzelexperimente ist, gilt auch allgemein.

In unserem Fall ist

$$\begin{aligned} H(P') &= - \left[\frac{1}{100} \cdot \log_2 \frac{1}{100} + \frac{99}{100} \cdot \log_2 \frac{99}{100} \right] \approx 0,08, \\ H(P) &= 1000 \cdot H(P') \approx 80. \end{aligned}$$

Für den oben angegebenen Code war die erwartete Codewortlänge $E[n] = 110$. Damit liegt man noch nahe an dem Optimum von $H(P) = 80$. \square