

Kapitel 3

Ein Statistisches Intermezzo

“Strange events permit themselves the luxury of occurring.” (Charlie Chan)

“Conditioning is the soul of statistics.” (Joe Blitzstein)

Unsere Umwelt produziert am laufenden Band Ergebnisse wie Wolken, Aktienkurse, Herzinfarkte oder Schmetterlinge. Wir interessieren uns zum Beispiel dafür, ob wir aus einer Wolkenstimmung auf baldigen Regen schließen können, ob es einen Zusammenhang zwischen Börsencrashes und Herzinfarkten gibt, oder ob sich die berühmten “Schmetterlinge im Bauch” auf den Prüfungserfolg auswirken.

Um solch breit gefassten Fragestellungen empirisch untersuchen zu können benötigen wir ein abstraktes Modell, d.h. auch ein mathematisches Instrumentarium, welches es uns gestattet, Phänomene wie oben geschildert mathematisch zu beschreiben.

Im einführenden Abschnitt haben wir bereits die Idee der wiederholten Stichprobenziehungen aus einer gegebenen Grundgesamtheit erläutert, um den stochastischen Charakter der Stichprobenregressionsfunktion zu betonen.

Diese Vorstellung wiederholter Ziehungen von Stichproben aus einer fix gegebenen Grundgesamtheit ist zwar intuitiv einleuchtend, aber für typisch ökonomische Anwendungen weniger geeignet. Ökonominen wollen das ökonomische System verstehen, den Prozess, der die beobachteten Daten generiert.

Grob vereinfacht können wir uns die Welt, oder auch die ‘Wirtschaft’, als einen riesigen datengenerierenden Prozess vorstellen, die laufend Ergebnisse wie Aktienkurse und Herzinfarkte produziert. Wir wollen hier Methoden entwickeln, die uns später helfen sollen einige Teilaspekte dieses äußerst komplexen Gebildes zu analysieren. Dabei geht es vor allem darum, wie wir aus den durch vielen Zufallsstörungen überlagerten Beobachtungen auf tiefer liegende Gesetzmäßigkeiten schließen können, die dem datengenerierenden Prozess hoffentlich zugrunde liegen, und wie wir diese Gesetzmäßigkeiten aus den beobachteten Daten schätzen können.

Das erste Problem besteht darin, dass uns die Natur ihre Ergebnisse nicht unmittelbar als fix und fertige Zahlen liefert, sondern z.B. in Form von Wolken oder Schmetterlingen. Um diese in ein mathematisches Gerüst zu bringen benötigen wir ein sehr allgemeines Konzept, nämlich Mengen. Mit Mengen kann man zwar fast

beliebige Ergebnisse beschreiben, aber sie haben einen entscheidenden Nachteil, der Umgang mit ihnen ist umständlich, man kann nicht einfach mit ihnen ‘rechnen’. Das Konzept der Zufallsvariablen wird es uns ermöglichen, ganz allgemeine Zufallseignisse in die Zahlenmenge abzubilden. Der allgemein Beweis, dass dies generell möglich ist, wurde von Stochastikern wie z.B. Andrey Nikolaevich Kolmogorov (1903 – 1987) in den dreißiger Jahren des letzten Jahrhunderts erbracht. Tatsächlich sind Zufallsvariablen ziemlich komplexe mathematische Gebilde, was uns hier aber nicht weiter zu kümmern braucht, der Umgang mit ihnen ist denkbar einfach.

Wir werden im folgenden Abschnitt zuerst das Konzept der Zufallsvariablen ein bisschen ausführlicher erläutern, uns dann mit deren Verteilungen und Momenten (z.B. Erwartungswerte und Varianzen) beschäftigen, und schließlich zeigen, dass die bekannte PRF (*‘population regression function’*) einfach eine ‘Bedingte Erwartungswertfunktion’ (CEF, *‘Conditional Expectation Function’*) ist.

Abschließend werden wir uns mit Stichproben und deren Eigenschaften beschäftigen, denn schließlich verwenden wir diese um aus Stichprobenbeobachtungen auf die PRF zu schließen.

Zu Ihrer Beruhigung, wir werden auch in diesem Abschnitt nicht wirklich in die Tiefe gehen, sondern der Intuition wieder den Vorrang gegenüber mathematischer Strenge einräumen. Manche Konzepte werden trotzdem zumindest anfänglich etwas abstrakt anmuten, aber diese Abstraktion hat einen hohen Ertrag, sie erlaubt es uns ein generelles Modell zu entwickeln, auf dessen Grundlage wir spätere Anwendungen aufbauen können.

3.1 Rechnen mit dem Zufall: Zufallsexperimente und Zufallsvariablen

Der logische Ausgangspunkt für die folgenden Überlegungen ist das Gedankenmodell eines Zufallsexperiments. Ein **Zufallsexperiment** (*‘random experiment’*) in unserem Sinne ist ein spezieller ‘Datenerzeugender Prozess’ (DGP), der die folgenden drei Bedingungen erfüllt:

1. alle möglichen Versuchsausgänge, d.h. die Menge aller möglichen *Elementarereignisse* (Ergebnisse) des Experiments sind a priori bekannt;
2. das Ergebnis einer einzelnen Durchführung des Experiments kann nicht mit Sicherheit vorhergesagt werden, aber es gibt eine Regelmäßigkeit bei wiederholten Durchführungen; und
3. das Experiment kann unter identischen Bedingungen beliebig oft wiederholt werden.

Klassische Zufallsexperimente sind zum Beispiel das Werfen einer Münze, das Ziehen einer Karte aus einem Stapel, Roulette oder Black Jack. Man beachte, dass es sich dabei nicht um ein Experiment im üblichen Sinne handeln muss, wir denken dabei bloss an ein Phänomen, dessen einzelne Ausgänge im Einzelfall nicht mit Sicherheit vorhergesagt werden können, obwohl bei wiederholten Ausführungen ein Muster

erkennbar ist. Man beachte auch, dass das Resultat eines Zufallsexperiments in vielen Fällen nicht eine Zahl ist, deshalb betrachten wir die einzelnen möglichen Ausgänge ganz allgemein als Elemente einer Menge.

Die Menge aller möglichen Ausgänge eines Zufallsexperiments wird *Ergebnismenge* oder Menge aller möglichen *Elementarereignisse* (*'outcomes set'*) genannt, und wird häufig mit dem Symbol Ω bezeichnet. Beispiele für Elementarereignisse sind das Geschlecht der nächsten Person, die zur Tür hereinkommt, welche Partei die nächste Wahl gewinnt, die Inflationsrate im nächsten Monat, kurzum, alle Ereignisse, die als Ausgänge eines Zufallsexperimentes interpretiert werden können. Für das Werfen einer Münze besteht $\Omega = \{\text{Wappen}, \text{Zahl}\}$ aus den Elementarereignissen $\{\text{Wappen}\}$ und $\{\text{Zahl}\}$.

Wenn wir eine Karte aus einem gemischten Stapel ziehen und uns für die Farbe der Karte interessieren ist $\Omega = \{\heartsuit, \clubsuit, \diamondsuit, \spadesuit\}$, und $\heartsuit \in \Omega$ bedeutet \heartsuit ist ein Element von Ω .

Die Anzahl der möglichen Ergebnisse eines Zufallsexperiments kann eine endlich große Zahl sein, wie in den oben aufgezählten Beispielen, aber die Anzahl der Elemente von Ω kann auch unendlich groß sein. In diesem Fall kann man weiter unterscheiden, ob Ω *abzählbar* oder *überabzählbar* viele Ergebnisse enthält.

Im Fall einer unendlich großen, aber abzählbaren Menge von Ergebnissen kann jedem Elementarereignis eine natürliche Zahl \mathbb{N} zugeordnet werden; ein Beispiel wäre die Anzahl der Würfe die benötigt wird, bis die erste Sechs gewürfelt wird.

In den späteren Anwendungen werden wir uns hauptsächlich für Zufallsexperimente interessieren, deren Menge von Elementarereignissen Ω eine überabzählbare Anzahl von Elementen enthält, zum Beispiel das Einkommen einer zufällig ausgewählten Person, welches jeden beliebigen Wert innerhalb eines Intervalls annehmen kann. Für die Abbildung solcher Mengen wird in der Regel die Menge der reellen Zahlen \mathbb{R} (bzw. ein Intervall daraus) benötigt.

Wir werden uns in diesem Abschnitt hauptsächlich mit endlichen Ergebnismengen beschäftigen, ganz einfach weil dies einfacher ist. Für überabzählbar große Ergebnismengen wird ein mathematisches Instrumentarium benötigt, welches wir hier nicht voraussetzen wollen. Mathematiker haben aber gezeigt, dass die Intuition für Zufallsexperimente mit einer endlichen Anzahl von möglichen Versuchsausgängen zum größten Teil auch für Zufallsexperimente gilt, deren Ergebnismenge Ω eine überabzählbare Anzahl von Elementen enthält.

Häufig sind wir nicht an einem einzelnen Elementarereignis interessiert, sondern an "*interessierenden Ereignissen*", zum Beispiel könnten wir uns beim Roulette für die Menge der geraden Zahlen größer 15 interessieren, oder beim Pokern dafür, ein 'Full House' zu ziehen.

Ereignisse (*'events'*) setzen sich aus einem oder mehreren Elementarereignissen zusammen. Formal wird ein Ereignis A als eine Teilmenge der Ergebnismenge Ω definiert, d.h. $A \subset \Omega$.¹ Beispielsweise setzt sich beim Würfeln das Ereignis "Werfen einer geraden Augenzahl" $A = \{2, 4, 6\}$ aus den Elementarereignissen $\{2\}$, $\{4\}$ und $\{6\}$ zusammen.

¹ $A \subset \Omega$ wenn jedes Element von A auch ein Element von Ω ist, bzw. etwas abstrakter $A \subset \Omega$ wenn für jedes $a \in A$ impliziert $a \in \Omega$.

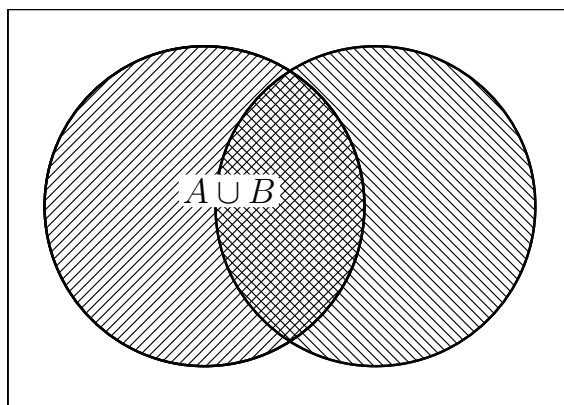


Abbildung 3.1: Vereinigung zweier Ereignisse.

$$A \cup B := \{x: x \in A \text{ oder } x \in B\}$$

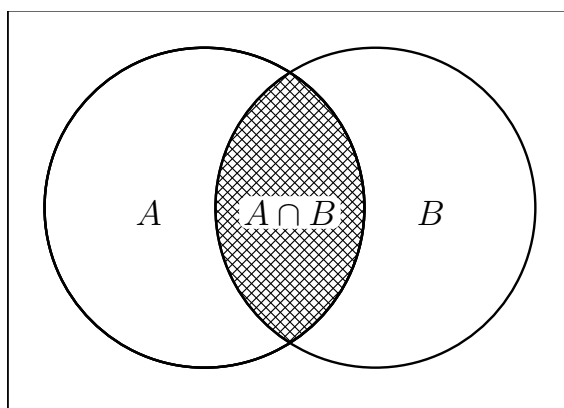


Abbildung 3.2: Durchschnitt zweier Ereignisse

$$A \cap B := \{x: x \in A \text{ und } x \in B\}$$

Wir sagen ein Ereignis A tritt ein, wenn bei der Durchführung des Zufallsexperiments genau eines der in A enthaltenen Elementarereignisse eintritt. Zum Beispiel tritt das Ereignis “Werfen einer geraden Augenzahl” genau dann ein, wenn eines der Elementarereignisse $\{2\}$, $\{4\}$ oder $\{6\}$ gewürfelt wird.

Nun gehen wir einen Schritt weiter und betrachten zwei Ereignisse, z.B. Ereignis A das Würfeln einer geraden Augenzahl und Ereignis B , das Würfeln Augenzahl > 3 . Wenn wir zwei beliebige Ereignisse A und B betrachten können wir die Vereinigung A und B ($A \cup B$) oder den Durchschnitt $A \cap B$ definieren.

Die *Vereinigung zweier Ereignisse* A und B ($A \cup B$) ist die Menge aller Elementarereignisse, die zu A *oder* B gehören (vgl. Abbildung 3.1).

Der *Durchschnitt zweier Ereignisse* A und B ($A \cap B$) ist die Menge aller Elementarereignisse, die zu A *und* B gehören (d.h. wenn A und B gemeinsam eintreten; vgl. Abbildung 3.2).

Ein unmögliches Ereignis wird durch die leere Menge \emptyset dargestellt. Zwei Ereignisse schließen sich gegenseitig aus, wenn $A \cap B = \emptyset$.

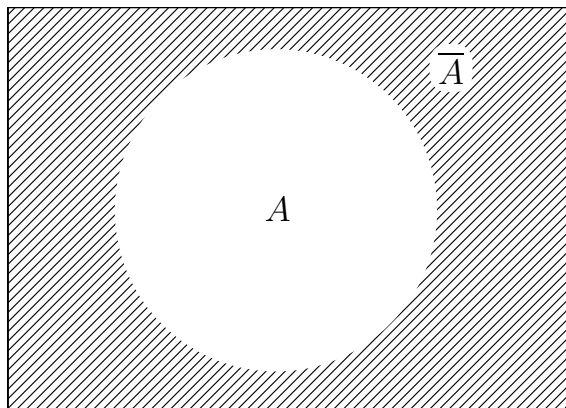


Abbildung 3.3: Komplementäre Menge, $\bar{A} := \{x : x \in \Omega \text{ und } x \notin A\}$

Die *komplementäre Menge* zu A relativ zu einer Universalmenge Ω sind alle Elemente von Ω , die *nicht* in A enthalten sind (vgl. Abbildung 3.3).

Beispiel: Wenn beim Würfeln $A = \{\text{Werfen einer geraden Augenzahl}\} = \{2, 4, 6\}$ und $B = \{\text{Werfen einer Augenzahl} \leq 3\} = \{1, 2, 3\}$, dann ist $A \cap B = \{2\}$, $A \cup B = \{1, 2, 3, 4, 6\}$, $\bar{A} = \{1, 3, 5\}$ und $\bar{B} = \{4, 5, 6\}$, $\bar{A} \cap \bar{B} = \{5\}$, $\bar{A} \cup \bar{B} = \{1, 3, 4, 5, 6\}$.

Mit Hilfe der Definition eines Ereignisses und der Mengenoperationen ist es möglich einen Ereignisraum² (*event space, sample space*) zu definieren.

Ein *Ereignisraum* \mathcal{A} enthält alle interessierenden Ereignisse und hat darüber hinaus eine mathematische Struktur. Wenn uns z.B. die Ereignisse A und B interessieren, enthält \mathcal{A} zusätzlich zu den Ereignissen A und B die leere Menge \emptyset , die Ergebnismenge Ω sowie alle weiteren mit diesen Mengen über Mengenoperationen verknüpfte Mengen, wie z.B. \bar{A} , \bar{B} , $A \cup B$, $A \cap B$ etc. Dies ist aus mathematischen Gründen erforderlich, da dies später die Definition von Zufallsvariablen erlaubt, aber für das Folgende von geringer Bedeutung. In der Sprache der Mathematik bildet \mathcal{A} eine sogenannte σ -Algebra, ein System von Mengen mit einer speziellen mathematischen Struktur. Diesen Elementen von \mathcal{A} können Wahrscheinlichkeiten zugeordnet werden.

3.2 Wahrscheinlichkeit (probability)

Unter Wahrscheinlichkeit versteht man ganz allgemein ein Maß zur Quantifizierung der Sicherheit bzw. Unsicherheit eines Zufallsexperiments. Konkret geht es darum, den Elementen der Ereignismenge \mathcal{A} die dazugehörigen Wahrscheinlichkeiten zuzuordnen.

Eine der ältesten Definitionen von Wahrscheinlichkeit geht auf den Mathematiker Pierre-Simon Marquis de Laplace (1749-1827) zurück und wird manchmal auch Lotterie-Definition oder ‘naive Wahrscheinlichkeitsdefinition’ genannt

$$P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl aller gleichmöglichen Fälle}}$$

²Unter einem Raum versteht man in der Mathematik ganz allgemein eine Menge mathematischer Objekte mit einer zusätzlichen mathematischen Struktur.

wobei zwei Ereignisse als gleichmöglich bezeichnet werden, wenn man das Eintreten aller Ereignisse für ‘gleich wahrscheinlich’ hält. So ist z.B. beim Würfeln die Wahrscheinlichkeit für das Ereignis A “Werfen einer geraden Augenzahl”

$$\Pr(A) = \frac{3}{6} = 0.5$$

Diese Wahrscheinlichkeitsdefinition ist allerdings nur für Zufallsexperimente mit *gleichwahrscheinlichen* Elementarereignissen anwendbar. Wenn Sie sich z.B. fragen, mit welcher Wahrscheinlichkeit Sie die nächste Prüfung bestehen, so gibt es einen günstigen Fall, Sie bestehen die Prüfung, und zwei mögliche Fälle, Sie bestehen die Prüfung oder Sie bestehen sie nicht. Daraus den Schluss zu ziehen, dass Sie die nächste Prüfung mit 50 Prozent Wahrscheinlichkeit bestehen werden, könnte sich als gefährlich erweisen. Außerdem wäre nach dieser Logik die Wahrscheinlichkeit, die nächste Prüfung mit einem ‘sehr gut’ zu bestehen, ebenfalls 50 Prozent, was offensichtlich unsinnig ist. Trotzdem leistet diese naive Wahrscheinlichkeitsdefinition für einfache Beispiele mit gleichwahrscheinlichen Ereignissen manchmal nützliche Dienste, z.B. wenn es um einfache Stichprobenziehungen geht. Für allgemeinere Anwendungen ist sie allerdings ungeeignet, dafür benötigen wir die weiter unten diskutierte axiomatische Definition von Wahrscheinlichkeit.

Frequentistische Wahrscheinlichkeitsdefinition Wenn ein Zufallsexperiment unter identischen Bedingungen beliebig oft wiederholt werden kann und wir die relative Häufigkeit eines Ereignisses A nach n Durchführungen des Experiments mit n_A/n bezeichnen, dann versteht man unter der frequentistischen Definition den Grenzwert dieser relativen Häufigkeit, wenn die Anzahl der Experimente gegen Unendlich geht

$$\Pr(A) = \lim_{n \rightarrow \infty} \left(\frac{n_A}{n} \right)$$

Dieser Wahrscheinlichkeitsbegriff ist in der Ökonometrie immer noch am gebräuchlichsten und liegt auch diesem Skript zugrunde.

Neben dem frequentistischen Wahrscheinlichkeitsbegriff gewinnen gerade in der jüngeren Literatur zunehmend *subjektive Wahrscheinlichkeitsbegriffe* an Bedeutung, wobei insbesondere die *Bayes'sche Sicht* von Bedeutung ist: “*probability is viewed as representing a degree of reasonable belief with the limiting values of zero being complete disbelief or disproof and of one being complete belief or proof.*” (Zellner 1984, 6), mehr dazu erfahren Sie in den Fortgeschrittenenveranstaltungen.

Für die jetzt folgenden Zwecke benötigen wir allerdings keine inhaltliche Interpretation von Wahrscheinlichkeit, für eine rein mathematische Behandlung reicht die **Axiomatische Wahrscheinlichkeitsdefinition** aus, die wesentlich auf *A.N. Kolmogorov* (1903 - 1987) zurückgeht. Dabei wird nicht versucht das ‘Wesen’ von Wahrscheinlichkeit zu ergründen, sondern es werden lediglich die erforderlichen mathematische Eigenschaften definiert.

Sie umfasst die folgenden drei Axiome:

1. $P(\Omega) = 1$

Da die Ergebnismenge Ω alle Elementarereignisse eines Zufallsexperiments enthält ist Ω ein sicheres Ereignis;

2. $P(A) \geq 0$ für alle Ereignisse $A \in \mathcal{A}$

Die Wahrscheinlichkeit $P(A)$ des Ereignisses A ist eine reelle, nichtnegative Zahl; gemeinsam mit 1. folgt $0 \leq P(A) \leq 1$.

3. Sei $\{A_n\}_{n=1}^{\infty}$ eine Folge sich *gegenseitig ausschließender* Ereignisse in \mathcal{A} , dann gilt für die Vereinigung $A = \bigcup_{n=1}^{\infty} A_n$

$$P(A) = \sum_{n=1}^{\infty} P(A_n)$$

Für eine endliche Menge mit n sich wechselseitig ausschließenden Ereignissen A_1, A_2, \dots, A_n bedeutet dies, dass die Wahrscheinlichkeit dafür, dass eines dieser Ereignisse eintritt (A_1 oder A_2 oder \dots A_n) gleich der Summe der Einzelwahrscheinlichkeiten ist: $P(A_1) + P(A_2) + \dots + P(A_n)$.

Wenn sich zwei Ereignisse A und B *nicht* ausschließen gilt der *Additionssatz*

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Beweis: Dies folgt unmittelbar aus den Axiomen. Dazu beachten wir, dass sich A und die Komplementärmenge \bar{A} gegenseitig ausschließen. Deshalb kann das Ereignis $A \cup B$ auch geschrieben werden als

$$A \cup B = A \cup (\bar{A} \cap B)$$

(vgl. Abbildung 3.1), und da sich die Ereignisse ausschließen sind die Wahrscheinlichkeiten

$$P(A \cup B) = P(A) + P(\bar{A} \cap B) \tag{3.1}$$

Ebenso kann B als Vereinigungsmenge zweier sich gegenseitig ausschließender Ereignisse angeschrieben werden

$$B = (A \cap B) \cup (\bar{A} \cap B)$$

Weil sich die Ereignisse ausschließen sind die Wahrscheinlichkeiten

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

Wenn wir dies umschreiben zu $P(\bar{A} \cap B) = P(B) - P(A \cap B)$ und in Gleichung (3.1) einsetzen erhalten wir das gewünschte Ergebnis

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

■

Man beachte, dass $P(\cdot)$ sogenannte Mengenfunktionen sind, die Elementen der Ergebnismenge \mathcal{A} reelle Zahlen zwischen Null und Eins zuordnen; $P(\cdot) : \mathcal{A} \mapsto [0, 1]$.

Dies sind keine üblichen Funktionen $f: \mathbb{R} \mapsto \mathbb{R}$ die wir aus der Schule kennen und mit denen man ‘rechnen’ kann! Wir werden gleich sehen, dass erst Zufallsvariablen dieses Problem lösen werden, erst diese gestatten die Definition von Wahrscheinlichkeits- und Dichtefunktionen $\text{Pr}: \mathbb{R} \mapsto [0, 1]$, die uns die Anwendung des üblichen mathematischen Instrumentariums ermöglichen, d.h., das ‘Rechnen mit dem Zufall’. Aber zuerst noch zu einem weiteren wichtigen Konzept der Statistik, welches für alles Folgende von zentraler Bedeutung ist.

Bedingte Wahrscheinlichkeiten: Häufig hängt die Wahrscheinlichkeit des Eintretens eines Ereignisses A vom Eintritt eines anderen Ereignisses B ab. Die Wahrscheinlichkeit für das Eintreten von A unter der Bedingung, dass Ereignis B vorher eingetreten ist oder gleichzeitig eintritt, wird *bedingte Wahrscheinlichkeit* $P(A|B)$ genannt. Sie ist für $P(B) > 0$ definiert als

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Die Logik wird unmittelbar aus Abbildung 3.2 ersichtlich: wenn $A \cap B \neq \emptyset$ erlaubt uns das Wissen, dass Ereignis B bereits eingetreten ist, eine genauere Einschätzung der Eintrittswahrscheinlichkeit von A .

Beispiel: Betrachten wir einen fairen Würfel und die Ereignisse

$A = \{1, 2, 3\}$ (würfeln einer Zahl kleiner 4), und

$B = \{2, 4, 6\}$ (würfeln einer geraden Zahl).

Angenommen es wurde einmal gewürfelt und wir wissen nur, dass eine gerade Zahl gewürfelt wurde, wie groß ist dann die Wahrscheinlichkeit, dass diese Zahl kleiner als 4 ist?

Da $A \cap B = \{2\}$ ist $P(A \cap B) = 1/6$; $P(B) = 3/6$, deshalb ist $P(A|B) = (1/6)/(3/6) = 1/3$.

Aus der Definition der bedingten Wahrscheinlichkeit folgt unmittelbar der *Multiplikationssatz*

$$P(A \cap B) = P(B) \cdot P(A|B)$$

der die Berechnung der Wahrscheinlichkeit für das Eintreten von A und B (d.h. $P(A \cap B)$) ermöglicht.

Damit ist es auch möglich **stochastische Unabhängigkeit** zu definieren:

Zwei Ereignisse A und B mit $P(A), P(B) > 0$ heißen *stochastisch unabhängig*, wenn die Wahrscheinlichkeit des Eintretens von Ereignis A nicht vom Eintreten oder Nichteintreten des Ereignisses B abhängt, d.h. wenn $P(A|B) = P(A)$.

Falls zwei Ereignisse stochastisch unabhängig sind ist $P(A \cap B) = P(A) \cdot P(B)$, deshalb ist die bedingte Wahrscheinlichkeit $P(A|B) = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$, das heißt, gleich der unbedingten Wahrscheinlichkeit. Die Kenntnis, dass B bereits eingetreten ist, erlaubt bei stochastischer Unabhängigkeit keine genaueren Aussagen über die Eintrittswahrscheinlichkeit von A .

Wahrscheinlichkeitsraum: Nun haben wir alle Ingredienzien beisammen, die wir für eine mathematische Beschreibung eines Zufallsexperiments benötigen, das Triple

$$[\Omega, \mathcal{A}, P(\cdot)]$$

bildet einen sogenannten *Wahrscheinlichkeitsraum* (*‘probability space’*).

Unter einem Wahrscheinlichkeitsraum kann man die mathematische Beschreibung des zugrundeliegenden Zufallsexperiments verstehen. Damit werden zwar die relevanten Aspekte des zugrunde liegenden Zufallsexperiments formal beschrieben, aber wir können immer noch nicht unmittelbar damit ‘rechnen’, da er nur auf Mengen definiert ist!

3.3 Zufallsvariablen

Sehr vereinfacht gesprochen ist eine Zufallsvariable (*‘random variable’*) eine Funktion, die den Ergebnissen eines Zufallsexperimentes (d.h. den Elementarereignissen oder Ereignissen) reelle Zahlen zuordnet. Diese Zuordnung geschieht derart, dass den Zahlen wieder die korrekten Wahrscheinlichkeiten des zugrunde liegenden Zufallsexperimentes zugeordnet werden können. In einem gewissen Sinne kann man also sagen, dass Zufallsvariablen eine Abbildung der relevanten Aspekte des dahinter liegenden Zufallsexperiments in die reellen Zahlen sind, und uns deshalb ermöglichen, mit den Resultaten von Zufallsexperimenten zu ‘rechnen’.

In einem gewissen Sinne leisten Zufallsvariablen in der Statistik etwas ähnliches wie Nutzenfunktionen in der Mikroökonomik. Auch Nutzenfunktionen können als Abbildung von Mengenkonzepten in die reellen Zahlen verstanden werden, eine auf Güterbündel definierte Präferenzordnung wird in die reellen Zahlen abgebildet, womit ein Rechnen mit ‘Nutzen’ ermöglicht wird.

In der Statistik hat es sich eingebürgert Zufallsvariablen mit Großbuchstaben zu bezeichnen (z.B. X), während man für die Realisationen von Zufallsvariablen die entsprechenden Kleinbuchstaben verwendet (z.B. x). Die Wahrscheinlichkeit, dass eine Zufallsvariable X die Realisation x annimmt, wird geschrieben als $\Pr(X = x)$.

Wie schon ausgeführt hat sich diese Schreibweise der Ökonometrie nicht durchgesetzt, aber wir werden sie in diesem Abschnitt von der Statistik übernehmen, weil sie für diese Zwecke ganz einfach praktisch ist.

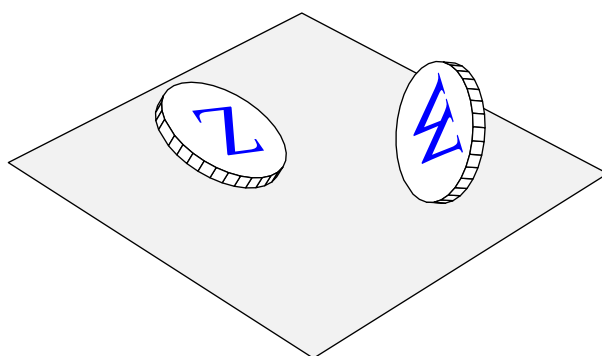
Stark vereinfacht wird das Konzept einer Zufallsvariable in Abbildung 3.4 dargestellt. Das zugrunde liegende Zufallsexperiment sei das Werfen zweier Münzen, und wir interessieren uns z.B. für die ‘Anzahl der Wappen’.

Wie Abbildung 3.4 zeigt kann diese Zufallsvariable X als Funktion aufgefasst werden, die jedem Elementarereignis eine reelle Zahl zuordnet. Der *Definitionsbereich* ist der Ereignisraum Ω des zugrundeliegenden Zufallsexperiments, und der *Wertebereich* ist die Menge der reellen Zahlen.

Achtung: Zufallsvariablen

1. beziehen sich immer auf die relevanten Ereignisse des zugrundeliegenden Zufallsexperiments,

Zufallsexperiment:



Zufallsvariable: Abbildung in die reellen Zahlen

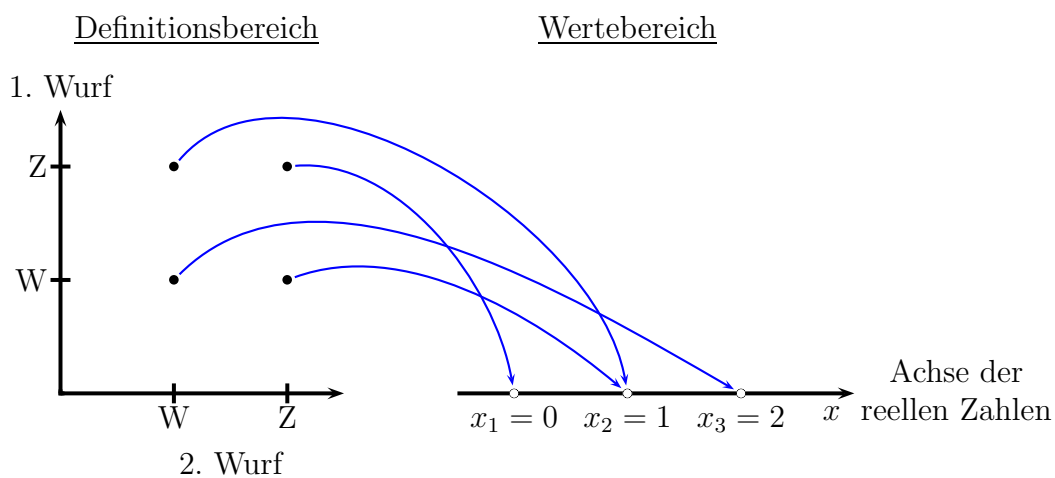


Abbildung 3.4: Definitions- und Wertebereich der Zufallsvariable X : “Anzahl Wappen” beim zweimaligen Werfen einer Münze (nach Bley Müller et al., 2002, 39f)

2. sie beschreiben *alle möglichen* Ausgänge des zugrunde liegenden Zufallsexperiments,
3. die Abbildung der Ereignisse in die reellen Zahlen \mathbb{R} erfolgt derart, dass diesen Zahlen wieder die korrekten Wahrscheinlichkeiten aus dem Zufallsexperiment zugeordnet werden können. Während die Zuordnung von (Teil-)Mengen zu Wahrscheinlichkeiten nur mit Mengenfunktionen $P(\cdot) : \mathcal{A} \mapsto [0, 1]$ möglich ist, können Zufallsvariablen mit Hilfe reeller Funktionen³ Wahrscheinlichkeiten zugeordnet werden. Um diesen Unterschied zu betonen verwenden für diese Wahrscheinlichkeiten das Symbol \Pr , d.h. für diskrete Zufallsvariablen $\Pr(X) : \mathbb{R} \mapsto [0, 1]$, bzw. für stetige Zufallsvariablen $f(X) : \mathbb{R} \mapsto [0, 1]$.

Man unterscheidet zwischen

- *diskreten Zufallsvariablen*: die Ereignismenge \mathcal{A} enthält eine *abzählbare* Anzahl von Elementen; und
- *stetigen Zufallsvariablen*: die Ereignismenge Menge \mathcal{A} enthält *überabzählbar* viele Elemente.

Die vorhergehenden Ausführungen sind eine sehr starke Vereinfachung, aber sie reichen für das prinzipielle Verständnis. Im folgenden Unterabschnitt führen wir die Überlegungen für Interessierte ein bisschen weiter, um zumindest eine erste intuitive Idee von dem Instrumentarium zu vermitteln, das für die Definition stetiger Zufallsvariablen erforderlich ist.

3.3.1 Wahrscheinlichkeitsraum und Zufallsvariablen*

Wir haben schon früher erwähnt, dass Zufallsvariablen ziemlich komplexe mathematische Gebilde sind. Eine wirkliche Einführung in das Konzept der Zufallsvariablen würde den Rahmen dieser Einführung bei weitem sprengen, aber da dieses Konzept für alles Folgende von derartiger Bedeutung ist wollen wir hier zumindest einige zentrale Begriffe kurz vorstellen. Die eilige Leserin kann diesen Abschnitt getrost überspringen ...

Ausgangspunkt der folgenden Überlegungen ist ein Zufallsexperiment, welches in einen Wahrscheinlichkeitsraum $[\Omega, \mathcal{A}, P(\cdot)]$ abgebildet werden kann. Ω die wieder die Ergebnismenge, \mathcal{A} eine Ereignismenge und $P(\cdot)$ eine Mengenfunktion.

Die Ereignismenge \mathcal{A} ist abgeschlossen bezüglich der Komplementbildung, der Vereinigungs- und Durchschnittsbildung. Das bedeutet, wenn eine dieser Mengenoperationen auf irgendein Element von \mathcal{A} angewandt wird, ist das Ergebnis wieder ein Element von \mathcal{A} .

Eine mögliche Ereignismenge ist immer die Potenzmenge, d.h. die Menge aller Teilmengen von Ω . Für einen einfachen Münzwurf mit den Elementarereignissen ‘Wappen’ (W) und ‘Zahl’ (Z) ist die Ereignismenge $\mathcal{A}_1 = \{\emptyset, \{K\}, \{W\}, \Omega\}$.

³Reelle Funktionen sind Abbildungen, in denen sowohl die Definitionsmenge als auch die Wertemenge Teilmengen von \mathbb{R} sind.

Für einen zweifachen Münzwurf mit $\Omega = \{(ZZ), (ZW), (WZ), (WW)\}$ ist die Ereignismenge schon deutlich komplexer, da sie neben den Elementarereignissen, \emptyset und $\Omega = \{(ZZ), (ZW), (WZ), (WW)\}$ auch alle Durchschnitte, Vereinigungen und Komplemente davon enthält

$$\begin{aligned} \mathcal{A} = & \{\{\emptyset\}, \{(ZZ)\}, \{(ZW)\}, \{(WZ)\}, \{(WW)\}, \\ & \{(ZZ), (ZW)\}, \{(ZZ), (WZ)\}, \{(ZZ), (WW)\}, \{(ZW), (WZ)\}, \\ & \{(ZW), (WW)\}, \{(WZ), (WW)\}, \\ & \{(ZZ), (ZW), (WZ)\}, \{(ZZ), (ZW), (WW)\}, \\ & \{(ZZ), (WZ), (WW)\}, \{(ZW), (WZ), (WW)\}, \{\Omega\}\} \end{aligned}$$

Diese Potenzmenge enthält insgesamt bereits 16 Elemente, für praktische Anwendungen ist der Weg über die Potenzmengen häufig nicht gangbar. Glücklicherweise benötigt man selten die wirklichen Potenzmengen, meist reichen deutlich einfachere Ereignismengen.

Wenn wir uns z.B. beim zweimaligen Münzwurf für das Ereignis A “mindestens ein Wappen” interessieren ist $A = \{(WW), (WZ), (ZW)\}$ und der Ereignisraum $\mathcal{A}_W = \{\emptyset, A, \bar{A}, \Omega\} = \{\emptyset, \{(WW), (WZ), (ZW)\}, \{(ZZ)\}, \{(WW), (WZ), (ZW), (ZZ)\}\}$. Die Ereignismenge \mathcal{A} umfasst also alle interessierenden Ereignisse, und darüber hinaus neben \emptyset und Ω auch die über Mengenoperationen damit verknüpften Mengen.

Im mathematischen Sinne bildet die Ereignismenge \mathcal{A} eine σ -Algebra, sie besitzt eine bestimmte mathematische Struktur und erfüllt folgende Bedingungen: (1) $\Omega \in \mathcal{A}$, (2) wenn $A \in \mathcal{A}$ muss $\bar{A} \in \mathcal{A}$, und (3) wenn $A_i \in \mathcal{A}$ für $i = 1, 2, \dots, n, \dots$ dann $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

$P(\cdot)$ ist schließlich eine Mengen-Funktion vom Ereignisraum \mathcal{A} in die reellen Zahlen zwischen Null und Eins, $P(\cdot) : \mathcal{A} \rightarrow [0, 1]$, die bestimmte Axiome erfüllt.

Abbildung 3.5 zeigt diesen Wahrscheinlichkeitsraum für ein sehr einfaches Zufallsexperiment mit nur vier diskreten Elementarereignissen.

Für solche einfachen Zufallsexperimente scheint dies ein bisschen viel Aufwand, aber der Vorteil dieser Herangehensweise liegt darin, dass dies auch für Mengen mit überabzählbar vielen Elementen verallgemeinert werden kann, und somit die Definition stetiger Zufallsvariablen ermöglicht.

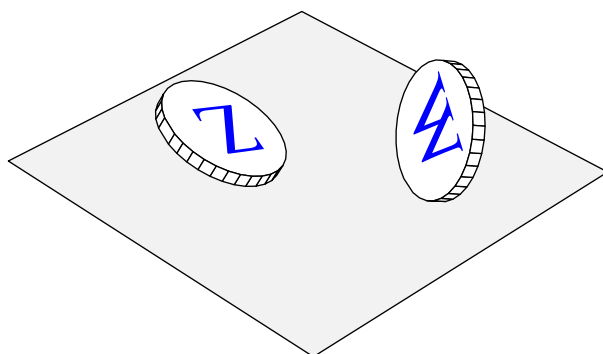
Eine der großen Einsichten von A.N. Kolmogorov bestand darin, dass für dieses Problem eine damals noch relativ neues Teilgebiet der Mathematik anwendbar ist, die Maßtheorie, welche ursprünglich für ganz andere Zwecke entwickelt wurde (es ging v.a. um die Verallgemeinerung von elementargeometrischen Begriffen wie Streckenlänge, Flächeninhalt und Volumen, die es ermöglichte auch komplizierteren Mengen ein Maß zuzuordnen).

Im mathematischen Sinne ist eine Zufallsvariable eine *messbare* Funktion von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \Pr(\cdot))$ in einen Messraum. Messbarkeit bedeutet dabei, dass das Urbild einer Menge wieder in einem bestimmten Mengensystem liegt, in unserem Fall eine Teilmenge der Ereignisalgebra \mathcal{A} ist.

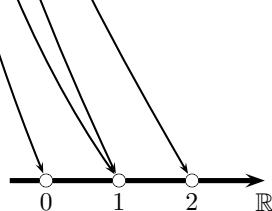
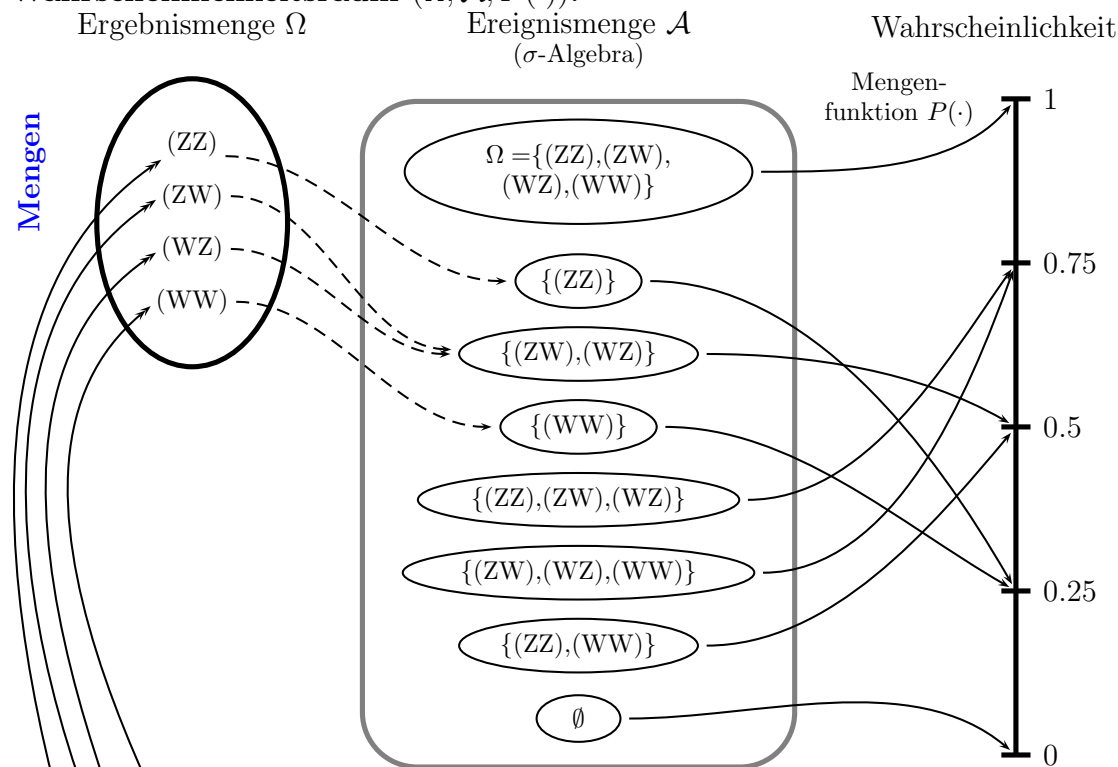
Damit kann eine stetige Zufallsvariable als eine Funktion $X(\cdot) \rightarrow \mathbb{R}$ definiert werden, die (für stetige Ereignisse) folgende Bedingung erfüllt

$$\{\omega : X(\omega) \leq x\} := X^{-1}((-\infty, x]) \in \mathcal{A} \quad \text{für alle } x \in \mathbb{R}$$

Zufallsexperiment:

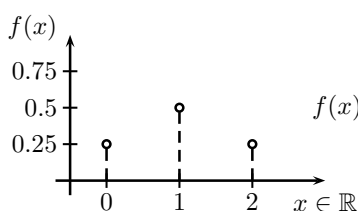


Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P(\cdot))$:



Zufallsvariable:

$X(\cdot): \Omega \mapsto \mathbb{R}_X$, so dass $\{\omega: X(\omega) = x\} := X^{-1}(x) \in \mathcal{A}$ für alle $x \in \mathbb{R}$



Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} 0.25 & \text{für } x = 0 \\ 0.5 & \text{für } x = 1 \\ 0.25 & \text{für } x = 2 \\ 0 & \text{sonst} \end{cases}$$

Abbildung 3.5: Zufallsexperiment Wurf mit zwei Münzen; Wahrscheinlichkeitsraum und Zufallsvariable für das interessierende Ereignis $X =$ Anzahl der Wappen (W). Die Abbildung in die reellen Zahlen erfolgt derart, dass das Urbild ein Element von \mathcal{A} ist.

Zu Ihrer Beruhigung, für das Verständnis des Folgenden benötigen Sie dies nicht wirklich. Die mathematische Theorie hinter den Zufallsvariablen garantiert uns aber, dass wir den folgenden Ausführungen vertrauen können.

Der Begriff *Zufallsvariable* ist eigentlich irreführend, denn im mathematischen Sinne handelt es sich dabei um keine Variable, sondern um eine Funktion. Darüber hinaus spielt der ‘Zufall’ für die mathematische Definition keine Rolle, es geht lediglich um die Abbildung der interessierenden Ereignisse in die reellen Zahlen.

Aber den Ausprägungen diskreter Zufallsvariablen, bzw. den Intervallen über Ausprägungen stetiger Zufallsvariablen, können Wahrscheinlichkeiten zugeordnet werden; für eine Zufallsvariable X existiert für jede reelle Zahl c eine Wahrscheinlichkeit, dass X einen Wert annimmt, der kleiner oder gleich c ist, oder in anderen Worten, für $c \in \mathbb{R}$ existiert immer eine Wahrscheinlichkeit $\Pr(X \leq c)$ (diese Wahrscheinlichkeit kann aber auch Null oder Eins sein).

Dies führt uns zu den nächsten wichtigen Konzepten, zu den Wahrscheinlichkeits-, Dichte- und Verteilungsfunktionen.

Aber vorher fassen wir nochmals zusammen: eine Zufallsvariable bildet alle möglichen Ausgänge des zugrunde liegenden Zufallsexperiments in die Menge der reellen Zahlen \mathbb{R} derart ab, dass die Wahrscheinlichkeiten des zugrunde liegenden Zufallsexperiments korrekt ‘übertragen’ werden können. Deshalb müssen wir uns im Folgenden nicht mit den Ergebnissen des Zufallsexperiments abmühen, die beliebige Mengen sein können, sondern wir können mit deren Abbildung in den reellen Zahlen – d.h. den Zufallsvariablen – rechnen!

3.4 Wahrscheinlichkeits- und Verteilungsfunktionen

Jedem Wert einer diskreten Zufallsvariable sind ein oder mehrere Elemente aus dem Ereignisraum des Zufallsexperiments zugeordnet. Da jedem möglichen Ereignis eines Zufallsexperiments eine Wahrscheinlichkeit zugeordnet ist, kann auch jedem diskreten Wert einer Zufallsvariable eine Wahrscheinlichkeit zugeordnet werden.⁴

Für stetige Zufallsvariablen ist die Mathematik etwas komplexer, aber im Prinzip funktioniert es ähnlich, nur werden statt einzelner Werte jeweils Intervalle betrachtet.

3.4.1 Wahrscheinlichkeitsfunktionen und Verteilungsfunktionen diskreter Zufallsvariablen

Eine **Wahrscheinlichkeitsfunktion** (*‘probability mass function’, pmf*) ordnet jeder der abzählbar vielen Ausprägungen einer diskreten Zufallsvariable die dazugehörige Wahrscheinlichkeit zu.

⁴Für ein einfaches Beispiel siehe Abbildung 3.5, Seite 13.

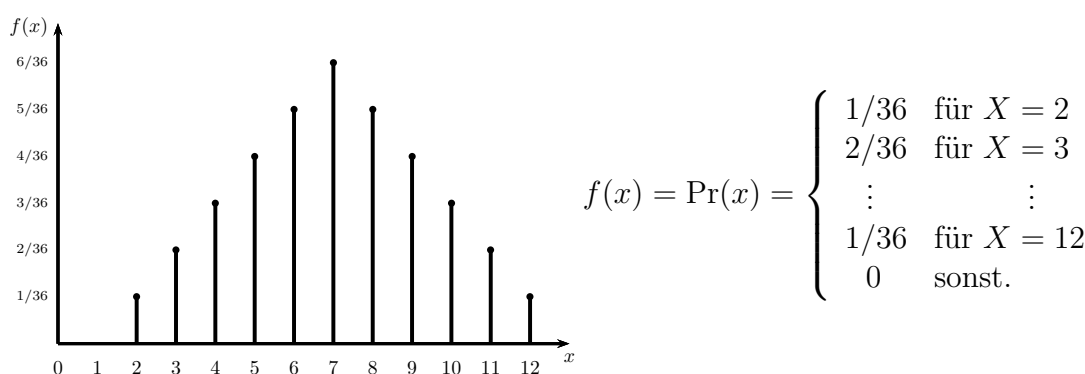
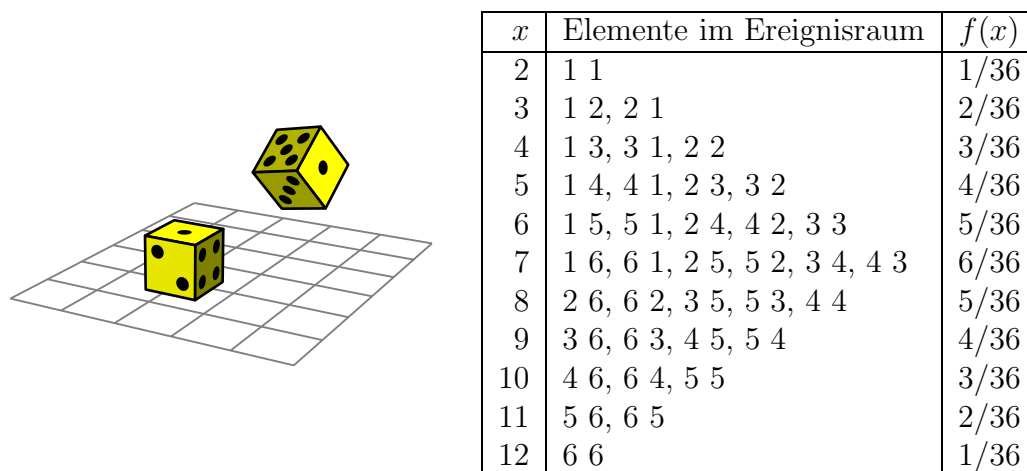


Abbildung 3.6: Beispiel 2: Wahrscheinlichkeitsfunktion der Zufallsvariablen X : “Augensumme bei einem Wurf mit zwei Würfeln”.

Wenn wir die unterschiedlichen Ausprägungen einer diskreten Zufallsvariablen X mit x_1, x_2, \dots bezeichnen gibt die Wahrscheinlichkeitsfunktion $f(x_i)$ also die Wahrscheinlichkeiten ihres Auftretens an

$$f(x_i) = \Pr(X = x_i) \quad \text{für } i = 1, 2, \dots$$

Im Unterschied zur Mengenfunktion $P(\cdot)$ des Wahrscheinlichkeitsraums ist $f(x_i) = \Pr(X = x_i)$ eine reelle Funktion mit der man wie üblich ‘rechnen’ kann.

Jede Wahrscheinlichkeitsfunktion muss die folgenden beiden Eigenschaften erfüllen:

$$f(x_i) \geq 0 \quad \text{für } i = 1, 2, \dots, m$$

$$\sum_{i=1}^m f(x_i) = 1$$

wobei $i = 1, \dots, m$ über alle möglichen Ausprägungen der Zufallsvariable läuft.

Wie kommen wir zu den Wahrscheinlichkeiten? Im wesentlichen gibt es drei Möglichkeiten:

1. in sehr einfachen Fällen können wir die Wahrscheinlichkeiten unmittelbar angeben, wenn wir das zugrunde liegende Zufallsexperiment kennen. Abbildungen 3.5 oder 3.6 sind zwei Beispiele dafür. Dieser Fall ist selten, die Prozesse, die uns interessieren, sind meist deutlich komplexer.

2. In manchen Fällen können wir zwar nicht unmittelbar die Wahrscheinlichkeiten angeben, aber aus theoretischen Überlegungen und praktischen Erfahrungen können wir vermuten, welche theoretische Verteilung sich zur Beschreibung eignet. Interessiert uns für das Zufallsexperiment ‘zweifacher Münzwurf’ eine andere Zufallsvariable Y “*mindestens ein Wappen wird geworfen*” sind nur zwei Ausgänge möglich, nämlich $X = 0$ oder $X = 1$. Die Wahrscheinlichkeitsfunktion wird deshalb durch eine Bernoulli-Verteilung $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$ beschrieben, wobei $0 \leq \theta \leq 1$ ein Parameter der Verteilung ist.

Für viele stetige Zufallsvariablen wissen wir, dass sie in der Natur annähernd normalverteilt sind, z.B. die Körpergröße.

3. In vielen Fällen ist es gar nicht erforderlich eine spezifische Verteilung anzunehmen. Die meisten Schätzfunktionen (wie z.B. Regressionskoeffizienten) können als bestimmte Funktionen von Momenten einer Verteilung geschrieben werden, und für solche Funktionen gelten häufig zentrale Grenzwertsätze. Deshalb konvergiert die (entsprechend skalierte) Verteilung dieser Schätzfunktionen mit zunehmender Stichprobengröße gegen die Normalverteilung, unabhängig davon, wie die ursprünglichen Zufallsvariablen verteilt sind, sofern einige wenig strenge Annahmen erfüllt sind.

Verteilungsfunktion Eine Verteilungsfunktion $F(x)$ (*cumulative distribution function*) gibt die Wahrscheinlichkeit dafür an, dass eine Zufallsvariable X höchstens den Wert x annimmt. Wenn die Ausprägungen x_i (mit $i = 1, 2, \dots, k, \dots$) aufsteigend nach ihrem Wert geordnet sind gilt

$$F(x_k) = \Pr(X \leq x_k) = f(x_1) + f(x_2) + \dots + f(x_k) = \sum_{i=1}^k f(x_i)$$

Abbildung 3.7 zeigt die Verteilungsfunktion für die Zufallsvariable X : “Augensumme bei einem Wurf mit zwei Würfeln” von unserem obigen Beispiel.

Übung: Wie lautet die Wahrscheinlichkeits- und Verteilungsfunktion für das *Produkt* der Augenzahlen bei zwei Würfeln mit einem Würfel?

3.4.2 Dichtefunktionen und Verteilungsfunktionen stetiger Zufallsvariablen

Eine Dichtefunktion (*‘density functions for continuous random variables’*) ist das Analogon zur Wahrscheinlichkeitsfunktion für stetige Zufallsvariablen. Ein wesentlicher Unterschied besteht zu diskreten Wahrscheinlichkeitsfunktionen besteht darin, dass für Dichtefunktion die *Wahrscheinlichkeit als Fläche unter der Dichtefunktion* definiert ist.

Wenn $f(x)$ eine Dichtefunktion ist, dann ist die Wahrscheinlichkeit dafür, dass die stetige Zufallsvariable X einen Wert in einem beliebigen Intervall $[a, b]$ (mit $a < b$ und $a, b \in \mathbb{R}$) annimmt, gleich

$$\Pr(a < X < b) = \int_a^b f(x)dx$$

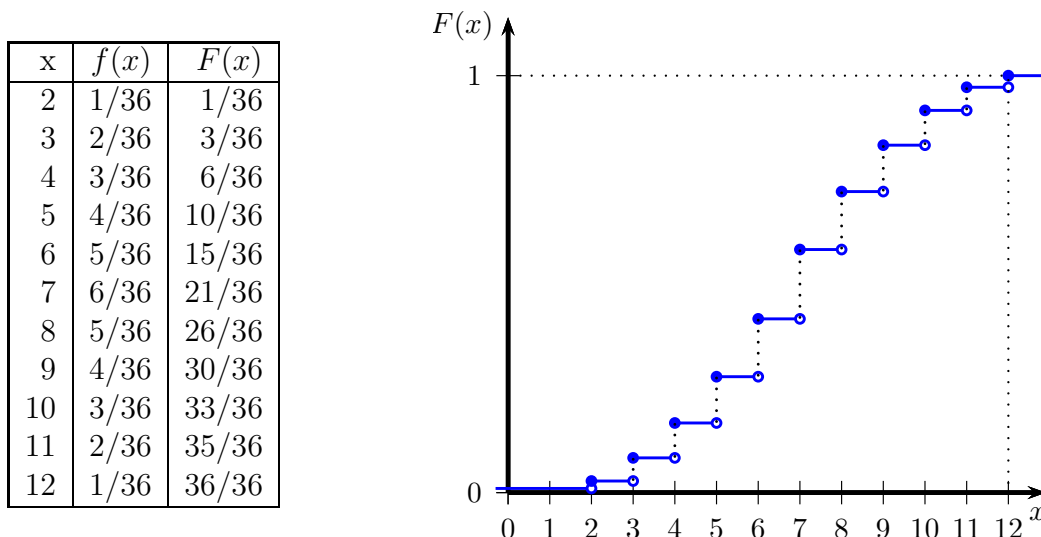


Abbildung 3.7: Verteilungsfunktion der Zufallsvariablen X : “Augensumme bei einem Wurf mit zwei Würfeln”.

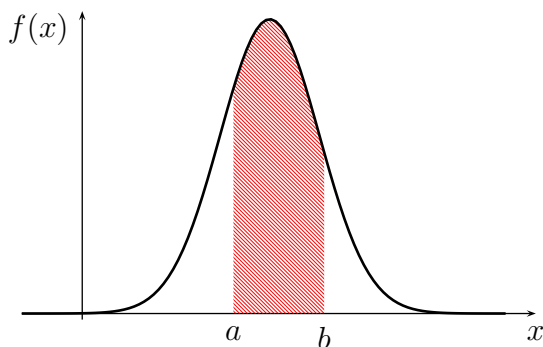


Abbildung 3.8: Dichtefunktion einer stetigen Zufallsvariablen.

Erinnern wir uns, dass die reellen Zahlen \mathbb{R} ‘unendlich dicht gepackt’ sind (d.h. jedes beliebige Intervall enthält überabzählbar viele Werte), die Flächen eines Punktes oder einer Linie in \mathbb{R} sind Null. Deshalb ist die Wahrscheinlichkeit dafür, dass eine Zufallsvariable X einen exakten Wert x annimmt, immer gleich Null – $\Pr(X = x) = 0!$ Man beachte, dass wir $f(x)$ zwar berechnen können und dass $f(x) \geq 0$, dieser Wert darf aber nicht als Wahrscheinlichkeit interpretiert werden; $f(x)$ kann z.B. auch größer als Eins sein!

Die Fläche unter einem Intervall der Dichtefunktion gibt also an, mit welcher Wahrscheinlichkeit Ereignisse, die diesem Intervall der Zufallsvariable zugeordnet sind, eintreten (siehe Abbildung 3.8).

Eine Dichtefunktion muss folgende Bedingungen erfüllen

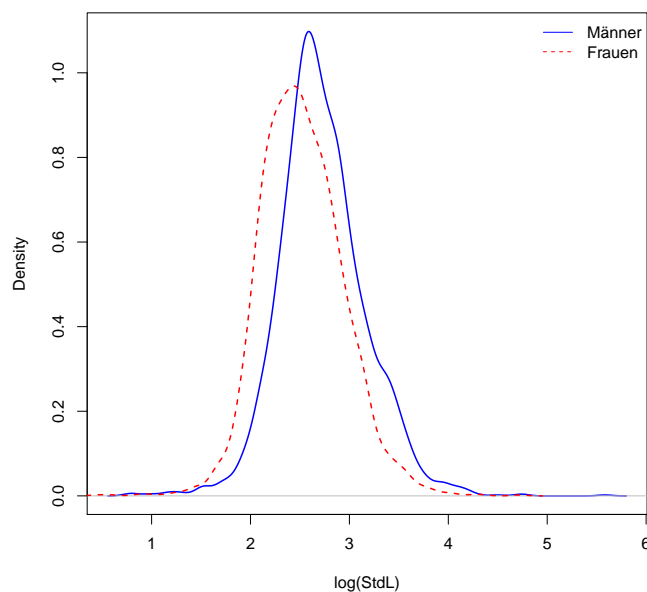


Abbildung 3.9: Dichtefunktion (empirisch) der log(Stundenlöhne) unselbständig beschäftigter Frauen und Männer in Österreich 2012 (Kerndichteschätzung, Datenquelle: EU-SILC Daten)

$$\begin{aligned}
 & f(x) \geq 0 \\
 & \int_{-\infty}^{\infty} f(x) dx = 1 \\
 & \Pr(a \leq X \leq b) = \int_a^b f(x) dx
 \end{aligned}$$

Ein Beispiel mit der Verteilung der log(Stundenlöhne) für Österreich finden Sie in Abbildung 3.9.

Beispiel: Ist die Funktion

$$f(x) = \begin{cases} \frac{1}{9} x^2 & \text{für } 0 \leq x \leq 3 \\ 0 & \text{sonst} \end{cases}$$

eine Dichtefunktion?

1. Offensichtlich ist $f(x) \geq 0$ für alle x im Bereich 0 bis 3.
2. Das Integral von 0 bis 3 ist ⁵

$$\int_0^3 \frac{1}{9} x^2 dx = \frac{1}{27} x^3 \Big|_0^3 = \frac{27}{27} - 0 = 1$$

⁵ $\int X^n dX = \frac{1}{1+n} X^{n+1} + c, \quad (n \neq -1)$

3. Die Wahrscheinlichkeit, dass X zwischen 0 und 1 liegt, ist z.B.

$$\int_0^1 \frac{1}{9} x^2 dx = \frac{1}{27} x^3 \Big|_0^1 = \frac{1}{27} - 0 = \frac{1}{27}$$

¶

Verteilungsfunktion Analog zum diskreten Fall existiert auf für stetige Zufallsvariablen eine *Verteilungsfunktion* $F(x) = \Pr(X \leq x)$.

Eine Verteilungsfunktion

$$F(x) = \int_{-\infty}^x f(v) dv$$

hat im allgemeinen folgende Eigenschaften:

1. $0 \leq F(x) \leq 1$;
2. $F(x)$ ist monoton wachsend, d.h. für $x_1 < x_2$ gilt $F(x_1) \leq F(x_2)$;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$;
4. $\lim_{x \rightarrow +\infty} F(x) = 1$;
5. $F(x)$ ist stetig.

Ein Beispiel mit der empirischen Verteilungsfunktion für die $\log(\text{Stundenlöhne})$ Österreichs finden Sie in Abbildung 3.10.

Beispiel: (aus Bley Müller et al., 2002, 42f) Ist die Funktion

$$f(x) = \begin{cases} 0.5 - 0.125x & \text{für } 0 \leq x \leq 4 \\ 0 & \text{sonst} \end{cases}$$

eine Dichtefunktion? Wie lautet die Verteilungsfunktion?

Offensichtlich ist $f(x)$ eine Dichtefunktion, denn

$$f(x) \geq 0 \quad \text{für alle } x$$

und

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx &= \int_{-\infty}^0 f(x) dx + \int_0^4 f(x) dx + \int_4^{+\infty} f(x) dx \\ &= 0 + \int_0^4 f(x) dx + 0 \\ &= \int_0^4 (0.5 - 0.125x) dx \\ &= \left[0.5x - \frac{0.125}{2} x^2 \right]_0^4 \\ &= 2 - 1 = 1 \end{aligned}$$

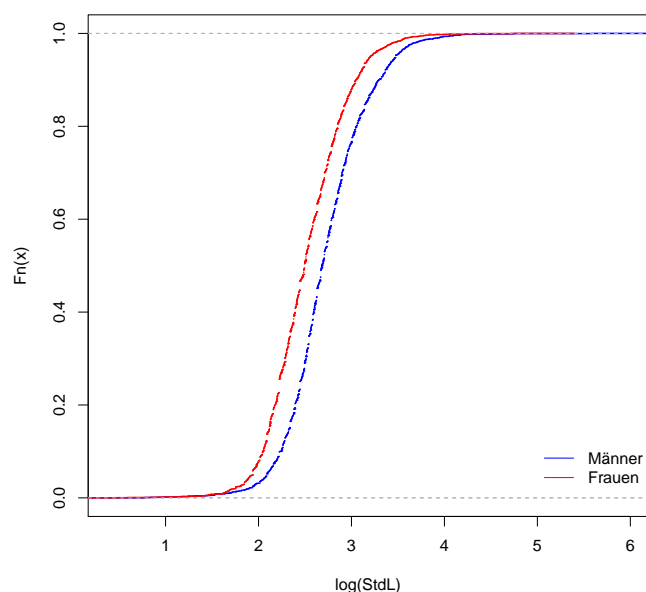


Abbildung 3.10: Verteilungsfunktion für $\log(\text{Stundenlöhne})$ unselbständig beschäftigter Frauen und Männer in Österreich 2012 (Datenquelle: EU-SILC Daten, vgl. Dichtefunktion in Abbildung 3.9).

Die Wahrscheinlichkeit, dass X z.B. einen Wert zwischen 1 und 2 annimmt, ist

$$\begin{aligned} \Pr(1 \leq X \leq 2) &= \int_1^2 f(x)dx \\ &= \int_1^2 (0.5 - 0.125x)dx \\ &= \left[0.5x - \frac{0.125}{2}x^2 \right]_1^2 \\ &= 0.75 - 0.4375 = 0.3125 \end{aligned}$$

Die Verteilungsfunktion $F(x)$ erhält man

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(v)dv = \int_0^x (0.5 - 0.125v)dv \\ &= \left[0.5v - \frac{0.125}{2}v^2 \right]_0^x \\ &= 0.5x - 0.0625x^2 \end{aligned}$$

also

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.5x - 0.0625x^2 & \text{für } 0 \leq x \leq 4 \\ 1 & \text{für } x > 4 \end{cases}$$

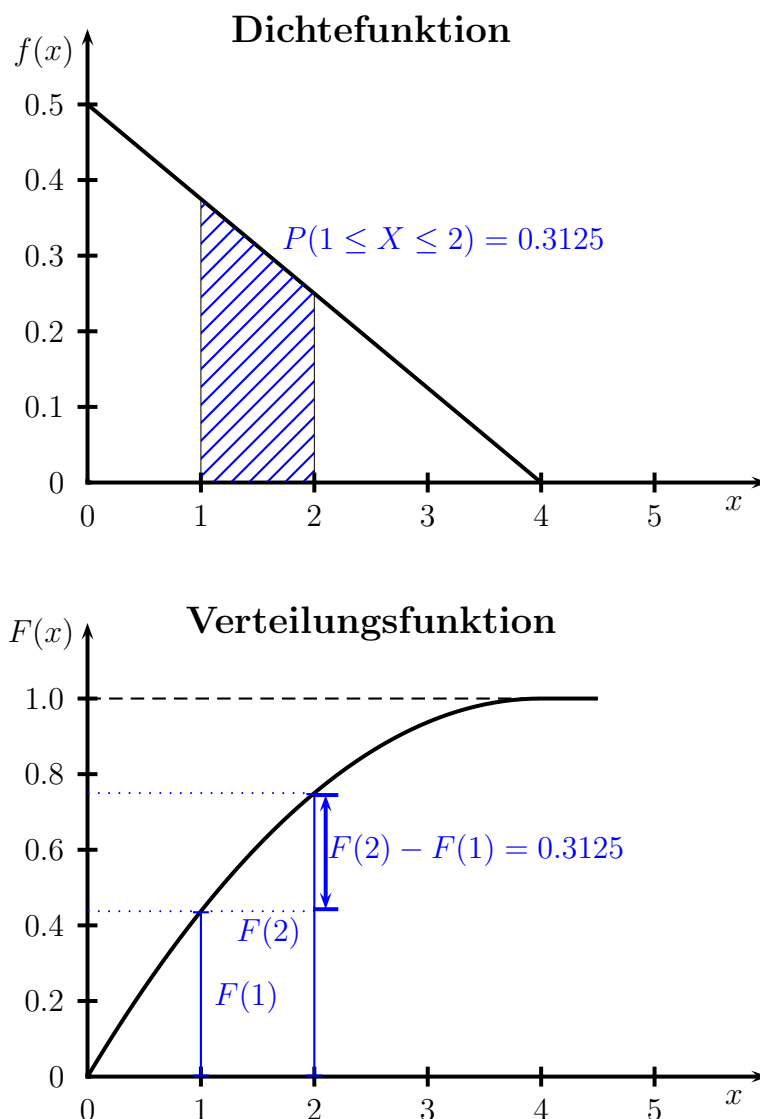


Abbildung 3.11: Dichte- und Verteilungsfunktion einer stetigen Zufallsvariablen.

Die Wahrscheinlichkeit dafür, dass X zwischen 1 und 2 liegt, kann auch mit Hilfe der Verteilungsfunktion berechnet werden:

$$\begin{aligned}
 \Pr(1 \leq X \leq 2) &= F(2) - F(1) \\
 &= 0.75 - 0.4375 \\
 &= 0.3125
 \end{aligned}$$

Dieses Beispiel ist in Abbildung 3.11 dargestellt.

3.4.3 Gemeinsame Wahrscheinlichkeitsfunktionen (*‘Joint Probability Density Function’*)

Die meisten Zufallsexperimente erzeugen mehr als eine Zufallsvariable, und wir interessieren uns in der Ökonometrie vor allem für Zusammenhänge zwischen solchen

Zufallsvariablen, z.B. für den Zusammenhang zwischen Bildung und Einkommen, oder dem Preis und dem Alter von Gebrauchtautos.

Selbst in dem einfachen Zufallsexperiment “zweifacher Münzwurf” können eine Zufallsvariable X “*Wappen beim ersten Wurf*” und eine zweite Zufallsvariable Y “*mindestens ein Wappen bei zwei Würfeln*” definieren, wobei wir den Zufallsvariablen X und Y jeweils den Wert Eins zuordnen, wenn das Ereignis eingetreten ist, und Null sonst. Die Menge der Elementarereignisse Ω ist $\{(ZZ), (WZ), (ZW), (WW)\}$, und wir können sofort die Wahrscheinlichkeitsfunktionen für diese beiden diskreten Zufallsvariablen hinschreiben

$$f_x(x) = \begin{cases} 0.5, & \text{für } X = 0 \\ 0.5, & \text{für } X = 1 \end{cases} \quad \text{und} \quad f_y(y) = \begin{cases} 0.25, & \text{für } Y = 0 \\ 0.75, & \text{für } Y = 1 \end{cases}$$

Aber wir können auch die gemeinsamen Wahrscheinlichkeiten angeben, z.B. ist die Wahrscheinlichkeit beim ersten Wurf *kein Wappen* (d.h. eine Zahl) und bei beiden Würfeln *kein Wappen* zu erhalten gleich 0.25, denn nur das Element $\{(ZZ)\}$ aus Ω erfüllt diese Bedingung, also ist $f(0, 0) = \Pr(X = 0, Y = 0) = 0.25$. Ähnlich können wir die anderen Wahrscheinlichkeiten ermitteln und als gemeinsame Wahrscheinlichkeitsfunktion $f(x, y)$ in Tabellenform anschreiben, wobei die erste Spalte die möglichen Ausprägungen von X und die erste Zeile die möglichen Ausprägungen von Y bezeichnet. Die Elemente ‘innerhalb’ sind die entsprechenden Wahrscheinlichkeiten.

X\Y	0	1
0	0.25	0.25
1	0	0.5

Die Wahrscheinlichkeit dafür, dass wir beim ersten Wurf *ein Wappen* ($X = 1$), und bei beiden Würfeln *kein Wappen* ($Y = 0$) erhalten, ist Null, d.h. $f(1, 0) = \Pr(X = 1, Y = 0) = 0$; oder, $f(0, 1) = \Pr(X = 0, Y = 1) = 0.25$.

Durch Aufsummieren der Wahrscheinlichkeiten erhält man die Randwahrscheinlichkeiten (*‘marginal probability’*), die gemeinsam die Randverteilungen bilden

Randverteilungen (*Marginal Probability Function*)

$$f_x(x) = \sum_y f(x, y) \quad \text{Randverteilung von } X$$

$$f_y(y) = \sum_x f(x, y) \quad \text{Randverteilung von } Y$$

bzw. analog für stetige Zufallsvariablen

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{Randverteilung von } X$$

$$f_y(y) = \int_{-\infty}^{+\infty} f(x, y) dx \quad \text{Randverteilung von } Y$$

Für das obige Beispiel erhalten wir die folgenden Randverteilungen $f_x(x)$ und $f_y(y)$:

X\Y	0	1	$f_x(x)$
0	0.25	0.25	0.5
1	0	0.5	0.5
$f_y(y)$	0.25	0.75	1

Man beachte, dass diese Randverteilungen wieder die univariaten Verteilungen sind, die wir erhalten haben, als wir die beiden Zufallsvariablen X und Y unabhängig voneinander untersucht haben.

In Tabelle 3.1 finden Sie eine etwas allgemeinere Darstellung der gemeinsamen Wahrscheinlichkeitsfunktion zweier diskreter Zufallsvariablen mit $f(x_i, y_j) = \Pr(X = x_i, Y = y_j)$, das heißt, $f(x_i, y_j)$ gibt die Wahrscheinlichkeit dafür an, dass die Zufallsvariable X den Wert x_i und die Zufallsvariable Y gleichzeitig den Wert y_j annimmt. Der Index $i = 1, \dots, n$ läuft über alle möglichen Ausprägungen von X , und $j = 1, \dots, m$ über alle möglichen Ausprägungen von Y .

Tabelle 3.1: Gemeinsame Wahrscheinlichkeitsfunktion zweier diskreten Zufallsvariablen X und Y mit Randverteilungen, wobei n und m die Anzahl der Ausprägungen der Zufallsvariablen X und Y angeben.

$X \setminus Y$	y_1	y_2	\dots	y_m	$f_x(x)$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	\dots	$f(x_1, y_m)$	$\sum_j f(x_1, y_j)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	\dots	$f(x_2, y_m)$	$\sum_j f(x_2, y_j)$
x_3	$f(x_3, y_1)$	$f(x_3, y_2)$	\dots	$f(x_3, y_m)$	$\sum_j f(x_3, y_j)$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
x_n	$f(x_n, y_1)$	$f(x_n, y_2)$	\dots	$f(x_n, y_m)$	$\sum_j f(x_n, y_j)$
$f_y(y)$	$\sum_i f(x_i, y_1)$	$\sum_i f(x_i, y_2)$	\dots	$\sum_i f(x_i, y_m)$	1

Selbstverständlich kann dies auch auf mehr Dimensionen erweitert werden, z.B. $f(x, y, z)$, aber diese Wahrscheinlichkeitsfunktionen können nicht mehr einfach grafisch dargestellt werden.

Natürlich muss auch für gemeinsame Wahrscheinlichkeitsfunktionen wieder gelten

$f(x_i, y_j) \geq 0$ für $i, j = 1, 2, \dots$	
und	$\sum_{i=1}^n \sum_{j=1}^m f(x_i, y_j) = 1$
bzw. für stetige Zufallsvariablen	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$

Gemeinsame Verteilungsfunktion: In analoger Weise ist auch die *gemeinsame Verteilungsfunktion* zweier diskreter Zufallsvariablen definiert,

$$F(x, y) = \Pr(X \leq x, Y \leq y)$$

definiert.

Sie gibt an, mit welcher Wahrscheinlichkeit die Zufallsvariable X *höchstens* den Wert x und die Zufallsvariable Y *höchstens* den Wert y annimmt.

Analog für stetige Zufallsvariablen

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(v, w) dv dw$$

Bedingte Wahrscheinlichkeitsfunktion (*Conditional Probability Density Function*)

Angenommen ein Zufallsexperiment erzeugt zwei Zufallsvariablen X und Y , und wir kennen bereits die Realisation von $X = x$, wissen aber noch nichts über Y . Erlaubt uns dies eine bessere Einschätzung der Wahrscheinlichkeiten für Y ?

Kehren wir noch einmal zurück zu unserem früheren Beispiel mit dem zweifachen Münzwurf, wobei $X = 1$ wenn beim ersten Wurf ein Wappen geworfen wurde und Null sonst, und $Y = 1$ wenn bei beiden Würfeln mindestens ein Wappen geworfen wurde. Die gemeinsame Wahrscheinlichkeitsfunktion mit den Randverteilungen haben wir bereits vorher diskutiert, wir reproduzieren sie hier nochmals für die weiteren Erklärungen

$X \setminus Y$	0	1	$f_x(x)$
0	0.25	0.25	0.5
1	0	0.5	0.5
$f_y(y)$	0.25	0.75	1

Angenommen wir wissen, dass beim ersten Wurf eine Zahl geworfen wurde ($X = 0$), ändert dies unsere Einschätzung für die Wahrscheinlichkeit bei zwei Würfeln mindestens ein Wappen zu werfen? Offensichtlich ja, denn wenn wir bereits mit dem ersten Wurf eine Zahl erhalten haben sind die beiden Ereignisse $\{(WZ)\}$ und $\{(WW)\}$ aus $\Omega = \{(ZZ), (ZW), (WZ), (WW)\}$ unmöglich! Mit diesem Vorwissen $X = 0$ ist die Wahrscheinlichkeit überhaupt kein Wappen zu werfen $Y = 0$ gleich 0.5, wir schreiben dies

$$\Pr(Y = 0 | X = 0) = 0.5 \quad \text{bzw.} \quad \Pr(Y = 1 | X = 0) = 0.5$$

und sagen, die Wahrscheinlichkeit für $Y = 0$ *gegeben* $X = 0$ ist 0.5, oder besser, die *bedingte Wahrscheinlichkeit* für $Y = 0$ *gegeben* $X = 0$ ist 0.5, und analog für $Y = 1$.

Wir können uns dies auch folgendermaßen vorstellen: sobald wir wissen, dass $X = 0$ eingetreten ist, ist in der obigen Tabelle mit der gemeinsamen Wahrscheinlichkeitsfunktion nur noch die *Zeile* mit $X = 0$ relevant, die zweite Zeile mit $X = 1$ ist nicht eingetreten und deshalb irrelevant.

Wie wirkt sich dies auf unser Wissen über Y aus? Wenn wir bereits wissen, dass nur Ereignisse in der Zeile $X = 0$ möglich sind (wir diese Zeile also gewissermaßen festhalten), dann müssen wir nur die Einträge für die gemeinsamen Wahrscheinlichkeiten durch die Randwahrscheinlichkeit für $X = 0$ dividieren, damit wir wieder korrekte Wahrscheinlichkeiten erhalten.

Deshalb können wir die bedingte Wahrscheinlichkeitsfunktion von Y für diskrete Zufallsvariablen etwas allgemeiner schreiben

$$f(y|X = x) = \Pr(Y = y|X = x) = \frac{f(x, y)}{f_x(x)} \quad \text{für } f_x(x) > 0$$

bzw. die bedingte Wahrscheinlichkeitsfunktion für X :

$$f(x|Y = y) = \Pr(X = x|Y = y) = \frac{f(x, y)}{f_y(y)} \quad \text{für } f_y(y) > 0$$

Wir brauchen nur die gemeinsame Wahrscheinlichkeit durch die Randwahrscheinlichkeit der Zufallsvariable dividieren, auf die wir ‘bedingen’ (die also gewissermaßen durch die Bedingung ‘festgehalten’ wird), um die entsprechende bedingte Wahrscheinlichkeit zu erhalten.

Deshalb können wir uns die bedingte Wahrscheinlichkeit $f(y|X = 0)$ als eine *gewichtete* gemeinsame Wahrscheinlichkeit für $f(y, X = 0)$ mit der Randwahrscheinlichkeit $f_x(x)$ als Gewicht vorstellen. Die Gewichtung ist erforderlich, damit die Summe der bedingten Wahrscheinlichkeiten wieder Eins ergibt ($\sum_y f(y|X = 0) = 1$ und $\sum_y f(y|X = 1) = 1$), sonst wäre die Definition einer Wahrscheinlichkeitsfunktion verletzt.

Für unser obiges Münzenbeispiel erhalten wir die bedingten Wahrscheinlichkeiten von Y , gegeben X

$$f(y|X = 0) = \begin{cases} \frac{f(X=0, Y=0)}{f_x(X=0)} = \frac{0.25}{0.5} = 0.5 & \text{für } Y = 0 \\ \frac{f(X=0, Y=1)}{f_x(X=0)} = \frac{0.25}{0.5} = 0.5 & \text{für } Y = 1 \end{cases}$$

bzw.

$$f(y|X = 1) = \begin{cases} \frac{f(X=1, Y=0)}{f_x(X=1)} = \frac{0}{0.5} = 0 & \text{für } Y = 0 \\ \frac{f(X=1, Y=1)}{f_x(X=1)} = \frac{0.5}{0.5} = 1 & \text{für } Y = 1 \end{cases}$$

oder kompakter

Y	0	1
$f(y X = 0)$	0.5	0.5
$f(y X = 1)$	0	1

Stochastische (bzw. statistische) Unabhängigkeit Zwei Zufallsvariablen X und Y sind stochastisch unabhängig, wenn

$$f(x, y) = f_x(x)f_y(y)$$

Für diskrete Zufallsvariablen bedeutet dies, dass für stochastische Unabhängigkeit *alle* gemeinsamen Wahrscheinlichkeiten gleich dem Produkt der Randwahrscheinlichkeiten sein müssen $\Pr(X = x_i, Y = y_j) = \Pr(X = x_i) \Pr(Y = y_j)$.

Unter Verwendung der Definition der bedingten Wahrscheinlichkeit $f(y|X = x) = \frac{f(y, X=x)}{f_x(x)}$ können wir dies alternativ schreiben

$$f(y|X = x) = f_x(x)$$

oder in Worten für diskrete Zufallsvariablen: bei stochastischer Unabhängigkeit ist die bedingte Wahrscheinlichkeit gleich der der unbedingten Wahrscheinlichkeit.

Deshalb können wir aus dem Wissen über das Vorliegen einer Zufallsvariable nur dann etwas über eine andere Zufallsvariable lernen, wenn diese *nicht* stochastisch unabhängig sind.

Beim Roulette sind die einzelnen Durchgänge stochastisch unabhängig, deshalb können wir aus früheren Realisationen nichts lernen, ganz egal wie oft hintereinander die Kugel auf Rot liegen geblieben ist.

Beispiel: Die beiden Zufallsvariablen X und Y mit der folgenden diskreten Wahrscheinlichkeitsverteilung

		Werte von Y		
		0	1	$f_x(x)$
Werte von X	1	0	1/3	1/3
	2	1/3	0	1/3
	3	0	1/3	1/3
$f_y(y)$		1/3	2/3	1

sind *nicht* stochastisch unabhängig, weil z.B. für die Wahrscheinlichkeiten von $X = 1$ und $Y = 0$: $1/3 \times 1/3 \neq 0$

Einige bedingte Wahrscheinlichkeiten sind z.B.

$$\Pr(Y = 0|X = 1) = \frac{\Pr(Y = 0, X = 1)}{\Pr_x(X = 1)} = \frac{0}{1/3} = 0$$

$$\Pr(Y = 0|X = 2) = \frac{\Pr(Y = 0, X = 2)}{\Pr_x(X = 2)} = \frac{1/3}{1/3} = 1$$

$$\Pr(Y = 1|X = 1) = \frac{\Pr(Y = 1, X = 1)}{\Pr_x(X = 1)} = \frac{1/3}{1/3} = 1$$

⋮

Die bedingten Wahrscheinlichkeitsfunktionen sind für Y , gegeben X :

Y	0	1
$f(y X = 1)$	0	1
$f(y X = 2)$	1	0
$f(y X = 3)$	0	1

X , gegeben Y :

X	1	2	3
$f(x Y = 0)$	0	1	0
$f(x Y = 1)$	0.5	0	0.5

3.5 Erwartungswerte (*'expected values'*)

Wahrscheinlichkeitsfunktionen sind wie normale Häufigkeitsverteilungen durch bestimmte Parameter charakterisiert. Die ersten zwei Momente sind der Erwartungswert $E(X)$ und die Varianz $\text{var}(X)$, häufig abgekürzt als μ und σ^2 .

Der **Erwartungswert** einer Zufallsvariable ist die *mit den Eintrittswahrscheinlichkeiten gewichtete Summe aller möglichen Ausprägungen einer Zufallsvariable*.

$$E(X) = \sum_{i=1}^n x_i f(x_i) \quad \text{für diskrete ZV}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{für stetige ZV}$$

Achtung:

1. Beim Erwartungswert wird über *alle möglichen* Ausprägungen der Zufallsvariable aufsummiert, gewichtet mit den Wahrscheinlichkeiten.
2. Erwartungswerte beziehen sich niemals auf Realisationen (z.B. Stichprobenbeobachtungen), sondern auf Zufallsvariablen!
Das Analogon für den Erwartungswert von Zufallsvariablen für Realisationen ist der Mittelwert.

Beispiel für diskrete ZV: Erwartungswert der Augenzahl beim Würfeln

$$E(X) = \mu_X = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3.5$$

Man beachte, dass der Erwartungswert einer Zufallsvariable X , also $E(X)$, *keine* Zufallsvariable ist. Jeder, der den Erwartungswert der Augenzahl berechnet wird auf das gleiche Ergebnis kommen (wenn er sich nicht verrechnet), da ist kein Zufallselement enthalten!

Um zu betonen, dass der Erwartungswert eine feste Zahl ist, wird er häufig mit μ bezeichnet (d.h. $E(X) := \mu_X$); der Mittelwert einer Stichprobe wird als \bar{x} geschrieben.

Für die Erwartungswerte von Zufallsvariablen gilt ebenso wie für Mittelwerte, dass die Summe der Abweichungen davon immer Null ist

$$\sum_i [x_i - E(X)] f(x_i) = \sum_i x_i f(x_i) - E(X) \sum_i f(x_i) = E(X) - E(X) = 0$$

das $E(X)$ eine Konstante ist und $\sum_i f(x_i) = 1$.

Beispiel für stetige ZV: Erwartungswert der Dichtefunktion $f(x) = x^2/9$ für $0 \leq x \leq 3$:

$$E(X) = \int_0^3 x \left(\frac{1}{9} x^2 \right) dx = \int_0^3 \frac{x^3}{9} dx$$

$$= \frac{x^4}{36} \Big|_0^3$$

$$= \frac{81}{36} = \frac{9}{4} = 2.25$$

3.5.1 Rechnen mit Erwartungswerten

Erwartungswerte sind gewichtete Summen, deshalb kann mit dem Erwartungswertoperator $E(\cdot)$ ‘sehr ähnlich’ gerechnet werden wie mit dem Summenzeichen.

- Für eine Zufallsvariable X und $c \in \mathbb{R}$ gilt:

$$\begin{aligned} E(c) &= c && \text{für } c = \text{konst.}, \text{ weil } \sum f(x) = 1 \\ E(cX) &= cE(X) && \text{für } c \text{ const.} \\ E[E(X)] &= E(X) && \text{weil } E(X) := \mu = \text{konst.} \\ E[g(X)] &= \int_{-\infty}^{\infty} g(x)f(x) dx && \text{für eine Funktion } g(\cdot) \end{aligned}$$

Warum? Weil z.B. $E(c) = \sum_i cf(x_i) = c \sum_i f(x_i) = c$; $E(cX) = \sum_i cx_i f(x_i) = c \sum_i x_i f(x_i) = cE(X)$;

$E(X)$ ist eine Konstante, also ist $E[E(X)] = \sum_i E(X)f(x_i) = E(X) \sum_i f(x_i) = E(X)$.

Beispiel: Wenn X die Augenzahl eines fairen Würfels ist, wie groß ist der Erwartungswert von $g(X) = X^2$?

$$E(X^2) = 1^2 \frac{1}{6} + 2^2 \frac{1}{6} + 3^2 \frac{1}{6} + 4^2 \frac{1}{6} + 5^2 \frac{1}{6} + 6^2 \frac{1}{6} = 15.1\dot{6}$$

Natürlich ist $E(X^2) = 15.1\dot{6} \neq [E(X)]^2 = 3.5^2 = 12.25$.

- Für eine diskrete Zufallsvariable X und $a, b \in \mathbb{R}$ gilt:

$$\boxed{E(a + bX) = a + bE(X)}$$

Beweis:

$$\begin{aligned} E(a + bX) &= \sum_{i=1}^m (a + bx_i)f(x_i) \\ &= \sum_i af(x_i) + \sum_i bx_i f(x_i) \\ &= a \sum_i f(x_i) + b \sum_i x_i f(x_i) \\ &= a + bE(X) \end{aligned}$$

Dies gilt analog auch für stetige Zufallsvariablen.

Bislang haben wir ausschließlich univariate Wahrscheinlichkeitsverteilungen untersucht. Nun wollen wir das Konzept für multivariate Fälle erweitern. Angenommen wir haben zwei Zufallsvariablen X und Y mit einer gemeinsamen Verteilung $f(x, y)$.

$$E[g(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) f(x_i, y_j) \quad \text{für diskrete ZV}$$

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad \text{für stetige ZV}$$

- Der Erwartungswert einer Summe von Zufallsvariablen ist gleich der Summe der Erwartungswerte, d.h.

$$E(X + Y + Z + \dots) = E(X) + E(Y) + E(Z) + \dots$$

Warum?

$$\begin{aligned} E(X + Y) &= \sum_i \sum_j (x_i + y_j) f(x_i, y_j) = \sum_i \sum_j x_i f(x_i, y_j) + \sum_i \sum_j y_j f(x_i, y_j) \\ &= \sum_i \left(x_i \sum_j f(x_i, y_j) \right) + \sum_j \left(y_j \sum_i f(x_i, y_j) \right) \\ &= \sum_i x_i f(x_i) + \sum_j y_j f(y_j) \\ &= E(X) + E(Y) \end{aligned}$$

Dies kann einfach verallgemeinert werden. Der Erwartungswert einer Linearkombination von Zufallsvariablen ist gleich der Linearkombination der Erwartungswerte, d.h. für $a_1, \dots, a_n \in \mathbb{R}$ und X_1, \dots, X_n Zufallsvariablen: $E(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \dots + a_n E(X_n)$, bzw.

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

- Wenn X und Y zwei *stochastisch unabhängige* Zufallsvariablen sind, dann ist der Erwartungswert ihres Produktes gleich dem Produkt der Erwartungswerte, d.h.

$$E(XY) = E(X) E(Y)$$

Actung, dies gilt *nur* für stochastisch unabhängige Zufallsvariablen!

Beweis:

$$E(XY) = \sum_i \sum_j (x_i y_j) f(x_i, y_j)$$

wenn X und Y unabhängig sind gilt: $f(x_i, y_j) = f(x_i) f(y_j)$. Deshalb:

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i) f(y_j) = \sum_i x_i f(x_i) \sum_j y_j f(y_j) = E(X) E(Y)$$

Achtung: Der Erwartungswert einer Summe ist *immer* gleich der Summe der Erwartungswerte, hingegen ist der Erwartungswert eines Produktes im allgemeinen nur dann gleich dem Produkt der Erwartungswerte, wenn die Variablen stochastisch unabhängig sind!

Beispiel:

X\Y	0	1	$f_x(x)$
0	0.25	0.25	0.5
1	0	0.5	0.5
$f_y(y)$	0.25	0.75	1

$$E(X) = 0 \times 0.5 + 1 \times 0.5 = 0.5$$

$$E(Y) = 0 \times 0.25 + 1 \times 0.75 = 0.75$$

$$E(XY) = 0 \times 0 \times 0.25 + 0 \times 1 \times 0.25 + 1 \times 0 \times 0 + 1 \times 1 \times 0.5 = 0.5$$

$$\Rightarrow E(XY) \neq E(X)E(Y)$$

Wie wir schon früher gesehen haben sind diese beiden Zufallsvariablen nicht stochastisch unabhängig.

3.5.2 Varianz

Die **Varianz** σ_X^2 einer Zufallsvariablen X ist definiert als

$$\text{var}(X) := \sigma_X^2 = E[X - E(X)]^2$$

d.h. für diskrete Zufallsvariablen $\sigma_X^2 = \sum_i (x_i - \mu)^2 f(x_i)$

Die Varianz kann auch folgendermaßen berechnet werden:

$$\begin{aligned} \sigma_X^2 &= E[X - E(X)]^2 = E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu\mu + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

also

$$\boxed{\text{var}(X) := \sigma_X^2 = E[X - E(X)]^2 = E(X^2) - [E(X)]^2}$$

- Für eine Zufallsvariable X und $a, b \in \mathbb{R}$ gilt:

$$\boxed{\text{var}(a + bX) = b^2 \text{var}(X)}$$

Beweis:

$$\begin{aligned} \text{var}(a + bX) &= E[(a + bX) - E(a + bX)]^2 \\ &= E[a + bX - a - bE(X)]^2 \\ &= E[b(X - E(X))]^2 \\ &= b^2 E[X - E(X)]^2 \\ &= b^2 \text{var}(X) \end{aligned}$$

Die Varianz einer Konstanten a ist natürlich immer Null, $\text{var}(a) = 0$. Dies gilt auch für stetige Zufallsvariablen.

Beispiel: Gesucht ist die Varianz einer Zufallsvariablen, deren Dichtefunktion durch $f(x) = x^2/9$ für $0 \leq x \leq 3$ gegeben ist.

Wir verwenden $\text{var}(X) = E(X^2) - [E(X)]^2$ und berechnen zuerst $E(X^2)$:

$$\begin{aligned} E(X^2) &= \int_0^3 x^2 \left(\frac{1}{9} x^2 \right) dx = \int_0^3 \frac{x^4}{9} dx \\ &= \frac{x^5}{45} \Big|_0^3 \\ &= \frac{243}{45} = 5.4 \end{aligned}$$

Da $E(X) = 9/4$ (siehe voriges Beispiel) ist $[E(X)]^2 = (9/4)^2$. Wir erhalten also

$$\text{var}(X) = E(X^2) - [E(X)]^2 = \frac{243}{45} - \left(\frac{9}{4} \right)^2 = 0.34$$

3.5.3 Kovarianz

Die **Kovarianz** ist definiert als

$$\text{cov}(X, Y) := E[[X - E(X)][Y - E(Y)]] = E(XY) - E(X)E(Y)$$

das zweite Gleichheitszeichen gilt weil

$$\begin{aligned} \text{cov}(X, Y) &= E[[X - E(X)][Y - E(Y)]] \\ &= E[XY - Y E(X) - X E(Y) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

- Wenn X und Y zwei stochastisch unabhängige Zufallsvariablen sind, dann ist die Kovarianz zwischen X und Y immer gleich Null ($\text{cov}(X, Y) = 0$).

Da bei stochastischer Unabhängigkeit gilt $E(XY) = E(X)E(Y)$ folgt dies unmittelbar aus obiger Definition.

Achtung: Eine Kovarianz von Null impliziert aber umgekehrt nicht stochastische Unabhängigkeit, wie man sich anhand des folgenden Beispiels für eine diskrete Wahrscheinlichkeitsverteilung verdeutlichen kann:

Beispiel: Gegeben sei folgende diskrete Wahrscheinlichkeitsverteilung:

		Werte von Y		$f(x)$
		0	1	
Werte von X	1	0	1/3	1/3
	2	1/3	0	1/3
	3	0	1/3	1/3
$f(y)$		1/3	2/3	1

Die Kovarianz ist $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

$$\begin{aligned} E(XY) &= \sum_i \sum_j (x_i y_j) f(x_i, y_j) \\ &= 1 * 0 * 0 + 1 * 1 * \frac{1}{3} + 2 * 0 * \frac{1}{3} + 2 * 1 * 0 + 3 * 0 * 0 + 3 * 1 * \frac{1}{3} \\ &= \frac{4}{3} \\ E(X) &= 1 * \frac{1}{3} + 2 * \frac{1}{3} + 3 * \frac{1}{3} = 2 \\ E(Y) &= 0 * \frac{1}{3} + 1 * \frac{2}{3} = \frac{2}{3} \end{aligned}$$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{4}{3} - 2 * \frac{2}{3} = 0$$

Die Variablen X und Y sind offenbar nicht stochastisch unabhängig, da $f(x_i)f(y_j) \neq f(x_i, y_j)$. Trotzdem ist die Kovarianz Null! Kovarianzen messen nur die lineare Abhängigkeit!

- Wenn X und Y zwei Zufallsvariablen sind gilt

$$\begin{aligned} \text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y) \\ \text{var}(X - Y) &= \text{var}(X) + \text{var}(Y) - 2 \text{cov}(X, Y) \end{aligned}$$

Warum? Erinnern Sie sich, $(a \pm b)^2 = a^2 + b^2 \pm 2ab$

$$\begin{aligned} \text{var}(X - Y) &= E[(X - Y) - E(X - Y)]^2 \\ &= E[(X - E(X)) - (Y - E(Y))]^2 \\ &= E[X - E(X)]^2 + E[Y - E(Y)]^2 - \\ &\quad 2E[(X - E(X))[(Y - E(Y))]] \\ &= \text{var}(X) + \text{var}(Y) - 2 \text{cov}(X, Y) \end{aligned}$$

Wenn X und Y stochastisch unabhängig sind ist $\text{cov}(X, Y) = 0$. deshalb gilt $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$

- Die Varianz einer Summe ist die Summe der Varianzen plus zwei Mal die Summe aller Kovarianzen zwischen den Zufallsvariablen der ursprünglichen Summe (analog zu $(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$).

$$\text{var}\left(\sum_{i=1}^m X_i\right) = \sum \text{var}(X_i) + \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m \text{cov}(X_i, X_j)$$

weil $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$.

Die positive Quadratwurzel der Varianz einer Zufallsvariablen X heißt **Standardabweichung**: $\text{st.dev.}(X) := \sigma_X = +\sqrt{\text{var}(X)}$.

Übungsbeispiele:

1. Zeigen Sie, dass $\text{cov}[X, 2X] = 2 \text{var}(X)$.
2. Zeigen Sie, dass $\text{cov}[X, (Y + Z)] = \text{cov}(X, Y) + \text{cov}(X, Z)$.
3. Zeigen Sie, dass für konstante a und b gilt

$$\text{cov}[X, (a + bX)] = b \text{var}(X)$$

4. Zeigen Sie, dass für konstante a_1, b_1, a_2 und b_2 gilt

$$\text{cov}[a_1 + b_1X, a_2 + b_2Y] = b_1b_2 \text{cov}(X, Y)$$

3.5.4 Korrelationskoeffizient

Die Kovarianz hängt von den Maßeinheiten der Variablen ab und ist deshalb manchmal schwierig zu interpretieren.

Der **Korrelationskoeffizient** (corr) hat diesen Problem nicht, er ist unabhängig von den zugrunde liegenden Maßeinheiten, in denen X und Y gemessen wurde:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Der Korrelationskoeffizient ist eine dimensionslose Zahl, die immer zwischen Null und Eins liegt

$$-1 \leq \text{corr}(X, Y) \leq 1$$

Bei einem perfekten negativen linearen Zusammenhang nimmt er den Wert -1 an, bei einem perfekten positiven linearen Zusammenhang den Wert $+1$. Bei stochastischer Unabhängigkeit ist $\text{cov}(X, Y)$ Null, deshalb ist in diesem Fall auch der Korrelationskoeffizient Null.

Übung: Zeigen Sie, dass $\text{corr}[X, (a + bX)] = 1$, und $\text{corr}[X, (a - bX)] = -1$.

Beweis* für $-1 \leq \text{corr}(X, Y) \leq 1$.

Beginnen wir mit dem Vorzeichen der Varianz

$$\text{var}(X + bY) = \text{var}(X) + 2b \text{cov}(X, Y) + b^2 \text{var}(Y) \geq 0$$

weil Varianzen nie negativ werden können.

Da dies für alle b gilt, muss es auch für ein spezielles b gelten. Der Trick besteht nun in der Wahl eines speziellen b , welches uns neue Einsichten liefert. Ein solches b ist

$$b = -\frac{\text{cov}(X, Y)}{\text{var}(Y)}$$

(mit $\text{var}(Y) > 0$), denn wenn wir dieses in die obige Varianz $\text{var}(X + bY)$ einsetzen folgt

$$\begin{aligned}\text{var}(X + bY) &= \text{var}(X) - \frac{2 \text{cov}(X, Y)}{\text{var}(Y)} \text{cov}(X, Y) + \frac{\text{cov}(X, Y)^2}{\text{var}(Y)^2} \text{var}(Y) \\ &= \text{var}(X) - \frac{2 \text{cov}(X, Y)^2}{\text{var}(Y)} + \frac{\text{cov}(X, Y)^2}{\text{var}(Y)} \\ &= \text{var}(X) - \frac{\text{cov}(X, Y)^2}{\text{var}(Y)} \geq 0\end{aligned}$$

Dies muss wieder größer gleich Null sein, da eine Varianz nicht negativ werden kann.

Daraus folgt aber

$$\text{var}(X) \text{var}(Y) \geq \text{cov}(X, Y)^2$$

oder

$$\frac{\text{cov}(X, Y)^2}{\text{var}(X) \text{var}(Y)} \leq 1$$

Die Wurzel des linken Ausdrucks ist der Korrelationskoeffizient $r = \text{corr}(X, Y) = \text{cov}(X, Y) / \sqrt{\text{var}(X) \text{var}(Y)}$, deshalb muss gelten

$$r = \text{corr}(X, Y) \leq |1|$$

bzw. $-1 \leq \text{corr}(X, Y) \leq 1$. Dies ist ein Spezialfall der Cauchy-Schwarz Ungleichung (siehe Appendix). ■

Da die Varianzen immer positiv sind hat der Korrelationskoeffizient immer das gleiche Vorzeichen wie die Kovarianz.

Weiters gilt für konstante a_1, b_1, a_2 und b_2 , wenn $b_1 b_2 > 0$

$$\text{corr}(a_1 + b_1 X, a_2 + b_2 Y) = \text{corr}(X, Y)$$

und wenn $b_1 b_2 < 0$

$$\text{corr}(a_1 + b_1 X, a_2 + b_2 Y) = -\text{corr}(X, Y)$$

3.5.5 Bedingte Erwartungswerte

Die bedingten Erwartungswerte spielen in der Ökonometrie eine herausragende Rolle, da die gefitteten Werte \hat{y} des Regressionsmodells als bedingte Erwartungswerte interpretiert werden können.

Der bedingte Erwartungswert einer Zufallsvariablen Y wird unter der Voraussetzung berechnet, dass noch zusätzliche Informationen über den Ausgang des zugrunde liegenden Zufallsexperiments verfügbar ist, z.B. dass $X = x$

$$\begin{aligned}E(Y|X = x) &= \sum_{i=1}^m y_i f(y_i|X = x) && \text{für diskrete ZV} \\ E(Y|X = x) &= \int_{-\infty}^{\infty} y f(y|X = x) dy && \text{für stetige ZV}\end{aligned}$$

Im wesentlichen werden sie gleich berechnet wir die unbedingten Erwartungswerte, als gewichtete Summe über alle möglichen Ausprägungen von Y , aber als Gewichte dienen nun die *bedingten Wahrscheinlichkeiten*.⁶ Sie können intuitiv als ein Analogon zu den bedingten Mittelwerten der deskriptiven Statistik angesehen werden.

Beispiel: Angenommen wir interessieren uns für den Zusammenhang zwischen Geschlecht und Einkommen. Dem Geschlecht X sei die Ausprägungen $X = 1$ für weiblich und $X = 0$ für männlich zugeordnet, die Einkommenssituation Y sei $Y = 0$ für ‘mittel’, $Y = 1$ für ‘hoch’, und $Y = 2$ für ‘sehr reich’ zugeordnet.

Die bivariate Wahrscheinlichkeitsverteilung $f(x, y)$ sei bekannt

		Werte von X		$f_y(y)$
		männl.	weibl.	
		0	1	
Werte von Y	0	0.1	0.2	0.3
	1	0.4	0.1	0.5
	2	0.0	0.2	0.2
$f_x(x)$		0.5	0.5	1

Die Wahrscheinlichkeit, dass eine Frau ‘sehr reich’ ist, ist demnach $\Pr(Y = 1, X = 1) = 0.1$.

Die Randwahrscheinlichkeiten sind $f_x(x)$ und $f_y(y)$, z.B. für Y (wenn X m Ausprägungen aufweist)

$$f_y(y) = \Pr(Y = y) = \sum_{i=1}^m \Pr(X = x_i, Y = y) = \begin{cases} 0.3 \\ 0.5 \\ 0.2 \end{cases}$$

Die *unbedingten* Erwartungswerte sind

$$E(X) = \sum_i \sum_j x_i f(x_i, y_j) = \sum_i x_i f_x(x_i) = 0.5$$

$$E(Y) = \sum_j \sum_i y_j f(x_i, y_j) = \sum_j y_j f_y(y_j) = 0.9$$

Die **bedingte Wahrscheinlichkeit** von Y , gegeben $X = x$ ist

$$\Pr(Y = y|X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)} \quad \text{bzw.} \quad f(y|x) = \frac{f(x, y)}{f_x(x)}$$

wir dividieren die gemeinsame Wahrscheinlichkeit durch die Randwahrscheinlichkeit der Variable, die wir ‘festhalten’. Die Wahrscheinlichkeit, dass eine Person Lila wählt, gegeben diese Person ist ein Mann, beträgt z.B. $\Pr(Y = 0|X = 0) = 0.1/0.5 = 0.2$ oder 20%.

Die bedingte Wahrscheinlichkeitsverteilung für Y , gegeben X , ist für dieses Beispiel also

⁶Tatsächlich sind bedingte Erwartungswerte v.a. für stetige Variablen keine so trivialen Gebilde, da jeweils auf die σ -Algebra des zugrunde liegenden Zufallsexperiments Bezug genommen werden muss. Unter anderem muss garantiert sein, dass die Wahrscheinlichkeit des Ereignisses, auf das bedingt wird, positiv ist.

		$f(y X = 0)$	$f(y X = 1)$
Werte von Y	0	0.2	0.4
	1	0.8	0.2
	2	0.0	0.4

Den **bedingten Erwartungswert** von Y , gegeben X , erhalten wir schließlich, indem wir über alle möglichen Ausprägungen von Y aufsummieren, *gewichtet mit den bedingten Wahrscheinlichkeiten*

$$E(Y|X = x) = \sum_{j=1}^J y_j \Pr(Y = y_j|X = x) = \sum_j y_j f(y_j|x)$$

Für obiges Beispiel:

$$\begin{aligned} E(Y|X = 0) &= 0 \times 0.2 + 1 \times 0.8 + 2 \times 0.0 = 0.8 \\ E(Y|X = 1) &= 0 \times 0.4 + 1 \times 0.2 + 2 \times 0.4 = 1 \end{aligned}$$

Die **bedingte Erwartungswertfunktion** (*Conditional Expectation Function*, CEF) von Y ordnet schließlich jeder Ausprägung von X den bedingten Erwartungswert von Y zu

$$E(Y|X = x) = \begin{cases} 0.8 & \text{für } X = 0 \\ 1 & \text{für } X = 1 \end{cases}$$

Übung: Der bedingte Erwartungswert von X , gegeben Y , macht in diesem Beispiel inhaltlich nicht sehr viel Sinn, aber Sie können aber trotzdem versuchen ihn zu berechnen.

Zur Kontrolle

$$E(X|Y = y) = \begin{cases} 2/3 & \text{für } Y = 0 \\ 0.2 & \text{für } Y = 1 \\ 1 & \text{für } Y = 2 \end{cases}$$

Die stochastische bedingte Erwartungswertfunktion

Wir haben bisher die bedingten Erwartungswerte für *gegebene* Ausprägungen von X berechnet, d.h. $E(Y|X = x_i)$. Diese bedingten Erwartungswerte sind wie die unbedingten Erwartungswerte Konstante, d.h. deterministische Größen.

Können wir auch für ein ‘beliebiges’ X die bedingte Erwartung berechnen? Können wir auf eine Zufallsvariable konditionieren?

Die Theorie dahinter ist ziemlich komplex, u.a. weil den einzelnen Ausprägungen x_i unterschiedliche Wahrscheinlichkeiten zugeordnet sind (die Randwahrscheinlichkeiten von X). Deshalb muss man auf die σ -Algebra des zugrunde liegenden Zufallsexperiments Bezug nehmen, doch man kann (mit Hilfe der Maßtheorie und des Satzes von Radon-Nikodým) zeigen, dass eine solche stochastische bedingte Erwartungswertfunktion existiert und einige für das Folgende wichtige Eigenschaften hat.

Um deutlich zu machen, dass hier auf eine Zufallsvariable konditioniert wird schreibt man für den stochastischen bedingten Erwartungswert häufig

$$E(Y|\sigma(X))$$

Da wir in diesem Fall auf eine Zufallsvariable konditionieren ist auch der stochastische bedingte Erwartungswert eine Zufallsvariable!

Eigenschaften der bedingten Erwartungswertfunktion

Die folgenden Eigenschaften gelten auch für die stochastische bedingte Erwartungswertfunktion.

1. **Linearität:** Seien X, Y und Z Zufallsvariablen und $a, b \in \mathbb{R}$ Konstante

$$E(aX + bY|Z = z) = a E(X|Z = z) + b E(Y|Z = z)$$

2. **Das einfache Gesetz der iterierten Erwartungen:** Erinnern wir uns, in der deskriptiven Statistik haben wir gezeigt, dass die mit den Anteilen gewichtete Summe der bedingten Mittelwerte der unbedingte Mittelwert ist.

Wenn z.B. wie in Tabelle 2.7 (Dummy Variablen) der Anteil von Männern und Frauen je 0.5 ist, und der durchschnittliche Stundenlohn von Männern 15 Euro und von Frauen 12.5 Euro ist, dann ist durchschnittliche Stundenlohn über alle Personen $0.5 \times 15 + 0.5 \times 12.5 = 13.75$.

Ein analoges Gesetz gilt auch für Zufallsvariablen. Für zwei Zufallsvariablen Y und X gilt

$$E(Y) = E_x[E(Y|X = x)]$$

d.h. *der (unbedingte) Erwartungswert der bedingten Erwartungswerte ist der unbedingte Erwartungswert* (E_x soll bedeuten, dass der äußere Erwartungswert über die X gebildet wird).

Für obiges Beispiel haben wir bereits den unbedingten Erwartungswert $E(Y) = 0.9$ und die bedingten Erwartungswerte $E(Y|X = 0) = 0.8$ sowie $E(Y|X = 1) = 1$ berechnet. Außerdem haben wir auch die Randverteilungen $f_x(X = 0) = 0.5$ und $f_x(X = 1) = 0.5$.

Daraus folgt

$$E(Y) = E[E(Y|X)] = \sum_i E(Y|X = x_i) f(x_i) = 0.8 \times 0.5 + 1 \times 0.5 = 0.9$$

bzw.

$$\begin{aligned} E(X) = E_y[E(X|Y)] &= \sum_j E(X|Y = y_j) f(y_j) \\ &= 2/3 \times 0.3 + 0.2 \times 0.5 + 1 \times 0.2 = 0.5 \end{aligned}$$

Beispiel: Kehren wir nochmals zu dem früheren Beispiel zurück:

		Werte von X		$f_y(y)$
		männl.	weibl.	
		0	1	
Werte von Y	0	0.1	0.2	0.3
	1	0.4	0.1	0.5
	2	0.0	0.2	0.2
$f_x(x)$		0.5	0.5	1

Der unbedingte Erwartungswert von Y ist $E(Y) = 0 \times 0.3 + 1 \times 0.5 + 2 \times 0.2 = 0.9$.

Die auf X bedingten Erwartungswerte von Y sind $E(Y|X = 0) = 0.8$ und $E(Y|X = 1) = 1$ (siehe oben).

Nach dem Gesetz der iterierten Erwartungen

$$E(Y) = E_x[E(Y|X = x)] = 0.5 \times 0.8 + 0.5 \times 1 = 0.9$$

Auch der $E(X) = 0.5$ kann so berechnet werden

$$E(X) = E_y[E(X|Y = y)] = 2/3 \times 0.3 + 0.2 \times 0.5 + 1 \times 0.2 = 0.5$$

3. **‘Taking out what is known property’:** Seien $g(X)$ und $h(Y)$ Funktionen der Zufallsvariablen X, Y , dann gilt

$$E[(g(X)h(Y)|X = x)] = g(X) E[h(Y)|X = x]$$

Intuitiv können wir uns vorstellen, dass durch die Konditionierung auf $X = x$ gewissermaßen X ‘festgehalten’ wird, und damit auch $g(X)$, weshalb $g(X)$ als Konstante vor den Erwartungswertoperator gezogen werden kann.

Als Spezialfall sehen wir uns $E(XY|X = x)$ für diskrete X und Y an, wobei $i = 1, \dots, m$ die Ausprägungen von X und $j = 1, \dots, m$ die Ausprägungen von Y indexiert.

$$\begin{aligned} E(XY|X = x_i) &= \sum_{j=1}^m x_i y_j \Pr(y_j, x_i|X = x_i) \\ &= x_i \sum_{j=1}^m y_j \Pr(y_j|X = x_i) \\ &= x_i E(Y|X = x_i) \end{aligned}$$

Da wir die Zufallsvariable X bei der Ausprägung x_i ‘festhalten’ ist $\Pr(y_j, x_i|X = x_i) = \Pr(y_j|X = x_i)$, und weil die Summation über j läuft können wir x_i als Konstante vor das Summenzeichen ziehen (wir untersuchen XY für einen spezifischen Wert von X).

Beispiel: wir setzen mit dem vorhergehenden Beispiel fort

$$E(XY|X = 1) = 0 \times 1 \times 0.4 + 1 \times 1 \times 0.2 + 2 \times 1 \times 0.4 = 1.$$

Mit der ‘Taking out what is known property’ für $x_i = 1$:

$$x_i E(Y|X = x_i) = 1 \times E(Y|X = 1) = 1 \times 1 = 1$$

$$\text{Oder } E(XY|Y = 1) = 1 \times 0 \times 0.8 + 1 \times 1 \times 0.2 = 0.2$$

bzw. mit der ‘Taking out what is known property’ für $y_i = 1$:

$$y_i E(X|Y = y_i) = 1 \times 0.2 = 0.2$$

Beispiel: diese Eigenschaft der bedingten Erwartungswertfunktion ist besonders für Regressionsfunktionen mit stochastischen X von Bedeutung (siehe Spanos, 1999, 364f).

Für bivariat normalverteilte Zufallsvariablen X und Y kann man zeigen, die bedingte Erwartungswertfunktion immer linear ist

$$E(Y|\sigma(X)) = \beta_1 + \beta_2 X$$

d.h. die bedingten Erwartungswerte liegen exakt auf einer Geraden. Für nicht normalverteilte Zufallsvariablen gilt dies manchmal zumindest approximativ.

Wir zeigen nun, dass in diesem Fall eine einfache Beziehung zwischen den Parametern β_1 und β_2 und den Momenten der gemeinsamen Verteilung von X und Y existiert.

Aufgrund des einfachen Gesetzes der iterierten Erwartungen gilt $E_x(E(Y|X)) = E(Y) = \beta_1 + \beta_2 E(X)$, oder

$$\beta_1 = E(Y) - \beta_2 E(X)$$

Außerdem folgt aus dem Gesetzes der iterierten Erwartungen und der ‘taking out what is known property’, dass

$$E(XY) = E[E(XY|\sigma(X))] = E[X E(Y|\sigma(X))]$$

Einsetzen von $\beta_1 = E(Y) - \beta_2 E(X)$ gibt

$$\begin{aligned} E(XY) &= E[X(\beta_1 + \beta_2 X)] \\ &= E[X(\underbrace{E(Y) - \beta_2 E(X)}_{\beta_1} + \beta_2 X)] \\ &= E\{X E(Y) + \beta_2 [X^2 - X E(X)]\} \\ &= E(X) E(Y) + \beta_2 [E(X^2) - E(X) E(X)] \\ \underbrace{E(XY) - E(X) E(Y)}_{\text{cov}(X,Y)} &= \beta_2 \underbrace{[E(X^2) - [E(X)]^2]}_{\text{var}(X)} \end{aligned}$$

daraus folgt

$$\beta_2 = \frac{\text{cov}(X, Y)}{\text{var}(X)}$$

Man beachte, dass sich dies auf die PRF und die Momente der Grundgesamtheit bezieht, nicht auf die SRF!

Für die bivariate Wahrscheinlichkeitsverteilung

		Werte von X		$f_y(y)$
		männl.	weibl.	
		0	1	
Werte von Y	0	0.1	0.2	0.3
	1	0.4	0.1	0.5
	2	0.0	0.2	0.2
$f_x(x)$		0.5	0.5	1

haben wir bereits die unbedingten Erwartungswerte $E(Y) = 0.9$ und $E(X) = 0.5$ sowie die bedingte Erwartungswertfunktion

$$E(Y|X = x) = \begin{cases} 0.8 & \text{für } X = 0 \\ 1 & \text{für } X = 1 \end{cases}$$

berechnet.

Wir können nun auch die *lineare Approximation* an diese bedingten Erwartungswerte berechnen; dazu benötigen wir $E(X^2) = 0^2 \times 0.5 + 1^2 \times 0.5 = 0.5$ und $E(XY) = \sum_{i=1}^2 \sum_{j=1}^3 x_i y_j f(x_i, y_j) = 0 \times 0 \times 0.1 + 0 \times 1 \times 0.2 + 1 \times 0 \times 0.4 + 1 \times 1 \times 0.1 + 2 \times 0 \times 0 + 2 \times 1 \times 0.2 = 0.5$.

Also

$$\beta_2 = \frac{\text{cov}(X, Y)}{\text{var}(X)} = \frac{E(XY) - E(X)E(Y)}{E(X^2) - [E(X)]^2} = \frac{0.5 - 0.5 \times 0.9}{0.5 - 0.5^2} = \frac{0.05}{0.25} = 0.2$$

und

$$\beta_1 = E(Y) - \beta_2 E(X) = 0.9 - 0.2 \times 0.5 = 0.8$$

Als lineare Approximation an die bedingte Erwartungswertfunktion erhalten wir also

$$E(Y|\sigma(X)) \stackrel{\text{lin}}{\approx} 0.8 + 0.2X$$

Offensichtlich ist in diesem Fall die bedingte Erwartungswertfunktion tatsächlich linear, denn die lineare Approximation $E(Y|X = 0) = 0.8 + 0.2 \times 0 = 0.8$ und $E(Y|X = 1) = 0.8 + 0.2 \times 1 = 1$ liefert exakt die gleichen Werte wie früher, aber das ist natürlich ein Zufall, oder um den eingangs erwähnten Charlie Chan zu zitieren, *“Strange events permit themselves the luxury of occurring”*. □

4. Die bedingte Erwartungswertfunktion ist der beste ‘*mean squared errors*’ Prediktor

$$E[Y - E(Y|\sigma(X))]^2 \leq E[Y - g(X)]^2 \quad \text{für alle } g(\cdot)$$

Die Distanz $E[Y - g(X)]^2 < \infty$ heißt ‘*mean squared error*’ (MSE). Von allen möglichen Funktionen $g(X)$ liefert der bedingte Erwartungswert $E(Y|\sigma(X))$ den kleinsten MSE.

3.5.6 Bedingte Varianz

Neben den bedingten Erwartungswerten ist in der Ökonometrie v.a. die bedingte Varianz von Bedeutung. Sie ist definiert als der bedingte Erwartungswert der quadratischen Abweichung der Zufallsvariablen von ihrem bedingten Erwartungswert.

$$\begin{aligned} \text{var}(Y|X = x) &= E \{ [Y - E(Y|X = x)]^2 | X = x \} \\ &= \sum_y [Y - E(Y|X = x)]^2 f(Y|X = x) \quad \text{für diskrete ZV} \\ &= \int_{-\infty}^{+\infty} [Y - E(Y|X = x)]^2 f(Y|X = x) \quad \text{für stetige ZV} \end{aligned}$$

Auch für bedingte Varianzen gilt:

- Nichtnegativität: $\text{var}(Y|X = x) \geq 0$
- Lineare Transformationen: $\text{var}(a + bY|X = x) = b^2 \text{var}(Y|X = x)$ für $a, b \in \mathbb{R}$
- Verschiebungssatz: $\text{var}(Y|X = x) = E(Y^2|X = x) - [E(Y|X = x)]^2$

Beispiel Die bedingte Varianz von Y für $X = 0$ ist für obiges Beispiel

$$\begin{aligned} \text{var}(Y|X = 0) &= \sum_j (y_j - E(Y|X = 0))^2 \text{Pr}(y_j|X = 0) \\ &= (0 - 0.8)^2 \times 0.2 + (1 - 0.8)^2 \times 0.8 + (2 - 0.8)^2 \times 0.0 \\ &= 0.16 \end{aligned}$$

Die **bedingte Varianzfunktion** (*'scedastic function'*) ordnet jeder möglichen Ausprägungen von X die entsprechende bedingte Varianz zu.

Hinweis: Wenn X, Y bivariat normalverteilt sind ist die bedingte Varianzfunktion $\text{var}(Y|X = x) = \sigma^2$, d.h. konstant und damit unabhängig von x , d.h. *homoskedastisch*.

Dies gilt fast nur für die Normalverteilung, für die allermeisten anderen theoretischen Verteilungen ändert sich die bedingte Varianz mit den x , d.h. die bedingte Varianzfunktion ist eine Funktion der x (vgl. Spanos, 1999, 342ff). Dieser Fall wird *Heteroskedastizität* genannt und wird uns später noch ausführlich beschäftigen.

Beispiel Kehren wir nochmals zurück zum früheren Zahlenbeispiel

Für die bivariate Wahrscheinlichkeitsverteilung

		Werte von X		$f_y(y)$
		männl.	weibl.	
		0	1	
Werte von Y	0	0.1	0.2	0.3
	1	0.4	0.1	0.5
	2	0.0	0.2	0.2
$f_x(x)$		0.5	0.5	1

haben wir bereits die bedingte Erwartungswertfunktion berechnet

$$E(Y|X = x) = \begin{cases} 4/5 & \text{für } X = 0 \\ 1 & \text{für } X = 1 \end{cases}$$

Die bedingten Varianzen sind

$$\begin{aligned} \text{var}(Y|X = 0) &= E(Y^2|X = 0) - [E(Y|X = 0)]^2 \\ &= \left[0^2 \times \frac{1}{5} + 1^2 \times \frac{4}{5} - 2^2 \times \frac{0}{5} \right] - \left[\frac{4}{5} \right]^2 = \frac{4}{25} = 0.16 \end{aligned}$$

$$\begin{aligned} \text{var}(Y|X = 1) &= E(Y^2|X = 1) - [E(Y|X = 1)]^2 \\ &= \left[0^2 \times \frac{2}{5} + 1^2 \times \frac{1}{5} - 2^2 \times \frac{2}{5} \right] - [1]^2 = \frac{4}{5} = 0.8 \end{aligned}$$

Die bedingte Varianzfunktion ist also

$$\text{var}(Y|X = x) = \begin{cases} 0.16 & \text{für } X = 0 \\ 0.8 & \text{für } X = 1 \end{cases}$$

Da in diesem Beispiel die bedingte Varianz davon abhängt, welche Ausprägung X annimmt, liegt Heteroskedastizität vor.

Übungsbeispiele:

1. Angenommen, ein eigenartiger Würfel mit 3 Seiten werde zweimal geworfen. Die Augenzahl sei 1,2 oder drei. Z_1 sei die Augenzahl des ersten Wurfes, und Z_2 die Augenzahl des zweiten Wurfes. Weiters sei $X = Z_1 + Z_2$, und $Y = Z_1 - Z_2$. (Lösungen ohne Gewähr!)
 - (a) Berechnen Sie die gemeinsame Wahrscheinlichkeitsfunktion und die Randverteilungen.
 - (b) Berechnen Sie den Erwartungswert und die Varianz von Y . (Lös.: 0, 10/9)
 - (c) Berechnen Sie die Kovarianz zwischen X und Y .
 - (d) Sind X und Y statistisch unabhängig? (Lös.: nein)
 - (e) Berechnen Sie den bedingten Erwartungswert von Y für $X = 3$. (Lös.: 0)
 - (f) Berechnen Sie den bedingten Erwartungswert von X für $Y = 0$. (Lös.: 4)
2. Gegeben sei folgende diskrete Wahrscheinlichkeitsverteilung:

		Werte von Y		
		1	3	9
Werte von X	2	1/8	1/24	1/12
	4	1/4	1/4	0
	6	1/8	1/24	1/12

- (a) Berechnen Sie die Randverteilungen von X und Y (d.h. die *marginal probability density functions*).
- (b) Berechnen Sie die bedingte Wahrscheinlichkeitsfunktion von Y für $X = 2$ (d.h. die *conditional probability density function*).
- (c) Berechnen Sie die Kovarianz zwischen X und Y .
- (d) Sind X und Y statistisch unabhängig?

(aus GHJ, S. 59)

Einige Lösungen: $\Pr(X = 2 \text{ und } Y = 9) = 2/24$, $\Pr(X = 2|Y = 9) = \frac{2/24}{4/24} = \frac{1}{2}$, $E(X) = 4$, $E(Y) = 3$, $E(Y|X = 2) = 1 \times \frac{1}{2} + 3 \times \frac{1}{6} + 9 \times \frac{1}{3} = 4$, $\text{var}(X) = 2$, $\text{var}(Y) = 8$, $E(XY) = 12$, $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 12 - 4 \times 3 = 0$ (!), $\text{var}(Y|X = 2) = E(Y^2|X = 2) - (E(Y|X = 2))^2 = 29 - 16 = 13$.

$$\text{var}(Y|X) = \begin{cases} 13 & \text{für } X = 2 \\ 1 & \text{für } X = 4 \\ \dots & \text{für } X = 6 \end{cases}$$

3. Eine Zufallsvariable sei gleichverteilt im Intervall $[0, 1]$: $X \sim G(0, 1)$, d.h. die Dichtefunktion lautet $f(X) = 1$.

- (a) Wie lautet die zugehörige Verteilungsfunktion?
- (b) Berechnen Sie $\Pr(0.1 \leq X \leq 0.9)$
- (c) Berechnen Sie $E[X]$
- (d) Berechnen Sie $\text{var}[X]$
- (e) Berechnen Sie $E[a + bX]$
- (f) Berechnen Sie $\text{var}[a + bX]$.

3.A Appendix

3.A.1 Cauchy-Schwarz Ungleichung

Seien X und Y zwei Zufallsvariablen, dann ist die Cauchy-Schwarz Ungleichung

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

Für den Beweis definieren wir $W = Y + bX$, wobei b eine Konstante ist. Dann ist

$$\mathbb{E}(W^2) = \mathbb{E}(Y^2) + 2b\mathbb{E}(XY) + b^2\mathbb{E}(X^2) \geq 0$$

da $W^2 \geq 0$ und deshalb auch $\mathbb{E}(W^2) \geq 0$. Dies muss für jedes b gelten, also z.B. auch für

$$b = -\frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$$

(Erwartungswerte sind deterministische Größen). Einsetzen gibt

$$\begin{aligned} \mathbb{E}(W^2) &= \mathbb{E}(Y^2) - \left(2\frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}\right)\mathbb{E}(XY) + \left(\frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}\right)^2\mathbb{E}(X^2) \\ &= \mathbb{E}(Y^2) - \frac{2[\mathbb{E}(XY)]^2}{\mathbb{E}(X^2)} + \frac{[\mathbb{E}(XY)]^2}{\mathbb{E}(X^2)} \\ &= \mathbb{E}(Y^2) - \frac{[\mathbb{E}(XY)]^2}{\mathbb{E}(X^2)} \geq 0 \end{aligned}$$

Achtung, wir benötigen nur $\mathbb{E}(W^2) \geq 0$, dies gilt für jedes b , also auch für dieses spezielle b !

Deshalb gilt

$$\mathbb{E}(X^2)\mathbb{E}(Y^2) \geq [\mathbb{E}(XY)]^2$$

bzw.

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

Literaturverzeichnis

Spanos, A. (1999), *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press.