Ben-Alexander Bohnke (2008, modifiziert aus „Integrale Logik“)

WAHRSCHEINLICHKEIT UND KAUSALITÄT

1. Isolierte Gleichgewichts-Relationen
2. Verknüpfte Gleichgewichts-Relationen
3. Ungleichgewichts-Relationen
4. Zufall und Korrelation
5. Kausal-Analyse

Man kann bei den Relatoren unterscheiden zwischen *Gleichgewichts-Relatoren* (wie $\leftrightarrow$) *und Ungleichgewichts-Relatoren* (wie $\rightarrow$). Entsprechend kann man unterscheiden zwischen *Gleichgewichts-Relationen* und *Ungleichgewichts-Relationen*. Diese Relationen werden im Folgenden ausführlich analysiert, vor allem im Hinblick auf p^T.

Ich habe (in der „Integralen Logik“) die *theoretische Wahrscheinlichkeit* bisher in folgenden Funktionenbeschrieben: zuerst natürlich *Wahrscheinlichkeit*, aber auch *Informationsgehalt* und *Tautologie-Grad*.

Hier geht es nun darum, inwieweit die Wahrscheinlichkeit Aussagen machen kann über synthetische *Zusammenhänge*, über *Abhängigkeit* bzw. *Korrelation*. Das betrifft die *empirische Wahrscheinlichkeit* p, aber noch mehr die *theoretische Wahrscheinlichkeit* p^T.

Und dies ist ein wesentlicher Hintergrund der folgenden Überlegungen: Wie kann man *Zusammenhänge* oder *Abhängigkeiten* zwischen Variablen bzw. zwischen Objekten oder Ereignissen feststellen? Dabei sind zwei Schritte zu unterscheiden:

- *funktionale* Abhängigkeit bestimmen

Es wird eine rein quantitative Verteilung festgestellt, die von der *Zufallserwartung* (signifikant) abweicht.

- *ursächliche* Abhängigkeit bestimmen

In einem zweiten Schritt wird erläutert, erklärt oder begründet, warum die funktionale Abhängigkeit auftritt. Als wichtigster Grund kommt hier *Kausalität* in Frage; und auf die will ich mich auch beschränken. Zunächst sollen dabei nur die *logischen* Strukturen von Kausalität analysiert werden; um zu beurteilen, ob wirklich ein Kausalzusammenhang vorliegt, müssen noch *hyper-logische* Faktoren berücksichtigt werden, dies wird in im 5. Punkt erläutert.

1. Isolierte Gleichgewichts-Relationen

1.1 Äquivalenz
1.2 Positiv-Implikation
1.3 Abhängigkeit

Es handelt sich um Relatoren bzw. Relationen wie $X \leftrightarrow Y$, $X >< Y$, $X \lfloor Y$, insgesamt gibt es in der *herkömmlichen* Aussagen-Logik 6 solcher Relatoren.

Es geht aber auch um die *Positiv-Implikation* $X *\!\!\rightarrow Y$, und die wird im Punkt 1 und 2. des Exkurses die wichtigste Rolle spielen.

Ich spreche von ‚*Gleichgewichts-Relatoren’*, weil in der Wahrheitstafel die Anzahl ihrer + und – unter dem Zentral-Relator im *Gleichgewicht* stehen, also *gleich häufig* vorkommen. Damit liegt ihre *strukturelle* theoretische Wahrscheinlichkeit in der *Mitte* zwischen $p^T = 1$ und $p^T = 0$, genau bei $p^T = 0{,}5$. (Ich habe sie an anderer Stelle auch ‘neutral’ genannt.)

1.1 ÄQUIVALENZ: X ↔ Y

X ↔ Y hat *strukturell* eine theoretische Wahrscheinlichkeit p^T = 2/4 = 1/2 = 0,5. D. h. es kommen in der Wahrheitstafel 2+ und 2– vor.

X	↔	Y
+	+	+
+	–	–
–	–	+
–	+	–

Ich berücksichtige hier nur den Normalfall X ↔ Y, also mit 2 *unterschiedlichen* Variablen. Bei 2 *gleichen* Variablen, und somit einer Tautologie X ⇔ X, ergibt sich natürlich ein anderes Ergebnis: p^T[X ⇔ X] = 4/4 = 1. Auch bei *mehr als 2* Variablen ergibt sich ein anderes Ergebnis, z. B. gilt: p^T[X ↔ Y ↔ Z] = 2/8 = 1/4 = 0,25.

In quantifizierter Form gilt für p(X ↔ Y) = r/n: p^T ist maximal 0,5, bei n = 1 oder n = 2.
Z. B. : p^T[p(X ↔ Y) = 1/1] = 1/2 = 0,5. Oder: p^T[p(X ↔ Y) = 1/2] = 2/4 = 0,5

Mit steigendem n fällt p^T. Ab n = 3 liegt der Wert immer unter 0,5 – so gesehen sind alle Verteilungen (ab n = 3) *unwahrscheinlich*. Z. B.:
p^T[p(X ↔ Y) = 1/3] = 0,375 ≈ 0,38. Zur Berechnung von p^T[p(X ↔ Y)] vgl. 3-3-3.

p^T bildet (graphisch) eine *symmetrische* oder zyklische *Kurve*.
Z. B. Werteverlauf für p (X ↔ Y) bei n = 4:
16/256, 64/256, 96/256, 64/256, 16/256 (ungekürzte Werte)
1/16, 4/16, 6/16, 4/16, 1/16 (gekürzte Werte).
D. h. p^T kommt wieder zum Ausgangswert zurück.
z. B.: p^T von p(X ↔ Y), bei n = 4

p^T					
6/16			●		
5/16					
4/16		●		●	
3/16					
2/16					
1/16	●				●
p	0/4	1/4	2/4	3/4	4/4

1.2 POSITIV-IMPLIKATION: X *→ Y

Noch interessanter ist für uns aber die *Positiv-Implikation* X *→ Y. Denn sie steht für die fundamentale *Kopula*-Funktion („ist“), und wir werden sehen, dass sie gravierende Unterschiede zur Normal-Implikation X → Y aufweist – die nämlich *ungleichgewichtig* ist.
In quantitativer Form lautet die Positiv-Implikation: p(X *→ Y) = r/n.
Die folgenden Aussagen (in 1. und 2.) beziehen sich alle auf die *Positiv-Implikation*.

Beispiel: p(Raucher *→ krank) = 75/100 = 0,75, d. h. X = Raucher, Y = Kranker (bzw. Plural: Kranke oder adjektivisch: krank). *Sprachlich*: ‚Die Wahrscheinlichkeit, dass ein Raucher krank ist, beträgt 0,75'. Kürzer: ‚75% der Raucher sind Kranke'. Dieser Wert, wie auch die späteren, ist nur ein Beispielwert, er stammt nicht aus der Medizinstatistik.

Für X *→ Y ergeben sich dieselben Werte von p wie für X ↔ Y. D. h für n = 4:

p(X*→ Y)	p dez.	p^T	p^T dez.	p^T %
4/4	1,0	1/16	0,06	6%
3/4	0,75	4/16	0,25	25%
2/4	0,5	6/16	0,38	38%
1/4	0,25	4/16	0,25	25%
0/4	0,0	1/16	0,06	6%

Die *empirische Wahrscheinlichkeit* p gibt wie beschrieben zunächst die *relative Häufigkeit* an, bei p(X *→ Y) = r/n gibt p an, dass r von n (bzw. soundsoviel Prozent) X auch Y sind.

Genauer müsste man unterscheiden zwischen der *relativen Häufigkeit* einer *Teilmenge* (Stichprobe) und der *Gesamtmenge* (Gesamtheit). Nur wenn man Aussagen über die Gesamtheit macht, sind im strengen Sinn *Wahrscheinlichkeitsaussagen* mit *Prozentangaben* und *Dezimalangaben* zulässig (vgl. 1-3-0-2). Die hier verwendeten *kleinen* Größen, wie z. B. n = 4, lassen natürlich an eine *Stichprobe* denken. Dennoch soll es hier prinzipiell um Aussagen über eine Gesamtheit gehen – die Größen dienen eben nur als Beispiele, und mit den kleinen Größen lassen sich die komplizierten Verhältnisse einfacher darstellen.

Neben dieser *extensionalen* Deutung wäre übrigens auch eine *intensionale* Interpretation möglich, die allerdings formal anders gestaltet sein müsste, z. B. p^{Int}(X *→ Y) = s/m. Dies gibt nicht an, *wie viele* X auch Y sind, sondern *zu* welchem Ausmaß ein (oder jedes) X die Eigenschaft Y besitzt, etwa in Prozent. Z. B. bei p^{Int}(X *→ Y) = 1: „X besitzt zu 100% die Eigenschaft der Gesundheit", kurz: „X ist zu 100% gesund" (vgl. 1-3-5). Diese intensionale Betrachtung ist hier aber sekundär, ich werde sie hier nicht weiter verfolgen, die extensionalen Verhältnisse sind schon kompliziert genug.

Man kann p – extensional – aber auch in *Wahrscheinlichkeits-Begriffen* angeben. Wenn z. B. *alle* X auch Y sind (p = 1), dann ist ein beliebiges X *sicher* ein Y.

p(X*→ Y)	p dez.	Häufigkeit/Wahrscheinlichkeit
4/4	1,0	X ist sicher ein Y
3/4	0,75	X ist wahrscheinlich ein Y
2/4	0,5	X ist vielleicht (zufällig) ein Y
1/4	0,25	X ist wahrscheinlich kein Y
0/4	0,0	X ist sicher kein Y

1.3 ABHÄNGIGKEIT

Die Wahrscheinlichkeit p kann aber auch dazu dienen, die *Abhängigkeit* zwischen X und Y, genauer von Y zu X, anzugeben. Zunächst folgende Übersicht:

p(X*→ Y)	p dez.	Art/Größe der Abhängigkeit
4/4	1,0	größte *positive* Abhängigkeit
3/4	0,75	mittlere *positive* Abhängigkeit
2/4	0,5	keine Abhängigkeit
1/4	0,25	mittlere *negative* Abhängigkeit
0/4	0,0	größte *negative* Abhängigkeit

Es gilt: Falls *positive* Abhängigkeit, dann keine *negative* Abhängigkeit und umgekehrt.

Wenn 100% aller X auch Y sind, dann liegt ein *maximaler* positiver Zusammenhang vor. Dem entspricht der Wert p = 1. In so einem Fall kann man z. B. *Kausalität* annehmen, im Beispiel, dass Rauchen eine *Ursache* von Krankheit ist. Und bei p = 0,75 könnte man annehmen, dass Rauchen ein *wichtiger Faktor* in der Entstehung von Krankheit ist, der aber allein nicht zwangsläufig zum Erkranken führt, eine *Teilursache*, man kann von *‚statistischer Kausalität'* sprechen.

Allerdings sind die hier im Beispiel verwendeten Werte von n = 4 sicher zu klein, um daraus *kausale* Gesetzmäßigkeiten herzuleiten (aber an diesem kleinen Wert von n lässt sich der Sachverhalt übersichtlicher darstellen, und ein großer Wert von n ändert nichts am Prinzip).

Wenn 50% (2/4) aller X auch Y sind, liegt *kein Zusammenhang* vor, insofern verweist der Wert p = 0,5 auf eine *zufällige* Verteilung. Zur Erläuterung: Wenn genau 50% aller X auch Y sind, hält man das im Alltagsdenken meist für einen *mittelgroßen* Zusammenhang. Statistisch ist aber i. allg. gerade *kein* Zusammenhang gegeben, weder ein positiver noch ein negativer. Sondern es ergeben sich einfach die meisten Kombinationsmöglichkeiten bei p = 0,5, daher ist p^T hier maximal. (Ich werde allerdings im nächsten Punkt zeigen, dass unter bestimmten Voraussetzungen p = 0,5 doch für einen mittelgroßen Zusammenhang stehen kann.) Alles über 50% ist ein positiver, alles unter 50% ein negativer Zusammenhang. Zum Beispiel:

p(Raucher *→ Kranker) = 0,5. Bedeutet verkürzt:
50% der Raucher sind Kranke. Das heißt andererseits:
50% der Raucher sind nicht Kranke. Bzw.:
50% der Raucher sind Gesunde.

Dies wird hier so interpretiert, dass *kein* Zusammenhang zwischen Rauchen und Kranksein besteht. Genauso wie man beim Werfen einer Münze (im Idealfall) feststellen kann, dass zu 50% Zahl und zu 50% Bild kommt, die 50%-50%-Verteilung ist eben gerade typisch für eine *zufällige* Verteilung, wie beim *Glücksspiel*. Erst wenn mehr als 50% der Raucher erkranken, sieht man einen positiven Zusammenhang:

p(Raucher *→ Kranker) > 0,5: *positiver* Zusammenhang

Wenn weniger als 50% der Raucher erkranken, müsste man dagegen annehmen, dass Rauchen sogar in gewissem Ausmaß vor Krankheit schützt.

p(Raucher *→ Kranker) < 0,5: *negativer* Zusammenhang

1.3.1 Genereller Abhängigkeits-Wert

Die Wahrscheinlichkeit p gibt die Größe der *Abhängigkeit* nicht generell an, weil *positiver* und *negativer* Abhängigkeit unterschiedliche Werte zugeordnet werden. So ist die Abhängigkeit von Y zu X bei p(X *→ Y) = 3/4 *genauso groß* wie p(X *→ Y) = 1/4, nur im ersten Fall handelt es sich um eine *positive* Abhängigkeit und im zweiten Fall um eine *negative* Abhängigkeit. Man kann daher einen anderen, auf p basierenden quantitativen *Abhängigkeitsbegriff* p^A einführen, der *positive* und *negative* Abhängigkeit integriert. Ursprünglich bestimmte ich: p(positive Abhängigkeit) = 1 – p(negative Abhängigkeit): $p^{A+} = 1 - p^{A-}$. Dies hat sich jedoch als problematisch erwiesen, weil man sinnvollerweise davon ausgeht, dass sich *positive* und *negative* Abhängigkeit vollständig ausschließen.

Ich gehe stattdessen jetzt von einer entsprechenden Beziehung aus, wobei aber p(X * → Y) nicht für positive Abhängigkeit steht und p(X * → Y) nicht für negative Abhängigkeit:

p(X * → Y)	p(X *→ ¬Y)
4/4	0/4
3/4	1/4
2/4	2/4
1/4	3/4
0/4	4/4

Es gilt: p(X * → Y) = 1 − p(X *→ ¬Y). Anstatt p(X *→ ¬Y) könnte man auch einsetzen: p¬(X *→ Y), weil bei der Positiv-Implikation gilt: p(X *→ ¬Y) = p¬(X *→ Y). Die *integrative Abhängigkeit* p^A berechnet man nun aus diesen beiden Werten.

Wenig sinnvoll ist es, die beiden Werte zu addieren und durch 2 zu dividieren, dann erhielte man immer den gleichen Abhängigkeitswert. Stattdessen errechnet man besser den *Betrag der Differenz*: p^A = | p(X * → Y) – p(X *→ ¬Y) |. Genauer gesagt:
p^A [p(X *→ Y)] = | p(X * → Y) – p(X *→ ¬Y) |. Für p^A [p(X *→ ¬Y)] gilt Entsprechendes.

In Worten: ‚Die Abhängigkeit von Y zu X in p(X *→ Y) errechnet sich als *Betrag* der Differenz von p(X * → Y) und p(X *→ ¬Y)'.

p(X * → Y)	p(X * → ¬Y)	integr. Abhängigkeit p^A
4/4	0/4	|4/4 – 0/4| = 4/4 = 1,0
3/4	1/4	|3/4 – 1/4| = 2/4 = 0,5
2/4	2/4	|2/4 – 2/4| = 0/4 = 0
1/4	3/4	|1/4 – 3/4| = 2/4 = 0,5
0/4	4/4	|0/4 – 4/4| = 4/4 = 1,0

So ergeben sich überzeugende Werte:
p^A[X *→ Y) = 4/4] = 1 hat ebenso den höchsten Wert wie p^A[X *→ Y) = 0/4] = 0; im ersten Fall liegt *maximale positive* Abhängigkeit vor, im zweiten Fall *maximale negative* Abhängigkeit. Und bei 50% (im Beispiel p = 2/4) ergibt sich wie gefordert der Wert p^A = 0.

Es sei noch kurz angefügt: Mittels dieses Abhängigkeits-Begriffs kann man auch die integrative Abhängigkeit bei *aussagen-logischen*, *qualitativen* Relationen, konkret bei X *→ Y bestimmen (also ohne dass n bekannt ist). Wie beschrieben, lässt sich X *→ Y auffassen als p(X *→ Y) = 1. Demnach steht X *→ ¬Y für p(X * → ¬Y) = 1.

Es gilt (nach dem Existenz-Modell): p(X * → Y) = 1 ⇔ p(X * → ¬Y) = 0. Somit lässt sich bestimmen: p^A[X *→ Y] = | p(X * → Y) – p(X *→ ¬Y) | = |1 – 0 | = 1.
Entsprechendes gilt für die Berechnung von p^A[¬(X *→ Y)].

Hier ist allerdings noch Folgendes anzumerken: Bei p(X *→ Y) bzw. X *→ Y gibt man prinzipiell nur an, inwieweit *Y von X abhängig* ist. Will man die *Gesamtabhängigkeit* zwischen X und Y angeben, so muss man auch

p(X ←* Y) sowie ggf. p(¬X *→ ¬Y) und p(¬X ←* ¬Y)

hinzunehmen. Das ist in diesem Kontext aber verzichtbar (vgl. unten).

1.3.2 p^T als Angabe der Größe einer Abhängigkeit

Wir haben gezeigt, dass die *empirische Wahrscheinlichkeit* p nicht sehr gut geeignet ist, die *Größe einer Abhängigkeit* anzugeben, der auf p basierende Wert p^A zwar besser, aber es fehlt die Erklärung, warum 100% eine größere (funktionale) Anhängigkeit darstellt als z. B. 75%. Hierauf kann die *theoretische Wahrscheinlichkeit* p^T eine Antwort geben.

Kommen wir zurück zur Tabelle mit n = 4:

p(X*→Y)	p dez.	p^T	p^T dez.
4/4	1,0	1/16	0,06
3/4	0,75	4/16	0,25
2/4	0,5	6/16	0,38
1/4	0,25	4/16	0,25
0/4	0,0	1/16	0,06

p^T gibt ja an, wie viele Möglichkeiten der *Kombination* von X und Y es (z. B. bei X *→ Y) jeweils gibt, wenn man die *Unabhängigkeit* von X und Y voraussetzt. Je mehr Möglichkeiten es für einen Wert von p gibt (also bei p = 4/4, 3/4, 2/4, 1/4 oder 0/4), desto wahrscheinlicher ist, dass diese Verteilung auftritt. Wenn nun real eine (theoretisch) unwahrscheinliche Verteilung auftritt, z. B. p(X *→ Y) = 4/4, so kann man umgekehrt daraus schließen, dass ein *Zusammenhang* zwischen X und Y besteht. Z. B. kann X *kausale* Ursache von Y oder aber ein *Teil* von Y sein. Und *je unwahrscheinlicher* diese Verteilung, *desto größer* vermutlich der *Zusammenhang* oder die Abhängigkeit von X und Y.

Den p-Wert mit der höchsten p^T nennt man ‚*Erwartungs-Wahrscheinlichkeit*' oder ‚*Zufallserwartung'* (vgl. 3-3-0-4). Man kann auch sagen, es ist der Wert mit dem *höchsten Tautologie-Grad*, denn ich habe ja gezeigt, dass man bei *synthetischen* ebenfalls Strukturen von einem Tautologie-Grad sprechen darf. Bei einem *Gleichgewichts-Relator* ist die Zufallserwartung immer p = 0,5, also hier p(X *→ Y) = 0,5.

Es wäre zu diskutieren: Kann die *Zufallserwartung* auch bedeuten, dass ein *geringer* bzw. der geringste Zusammenhang besteht, aber nicht *kein* Zusammenhang, nicht keine Abhängigkeit? Ich halte das letztlich nicht für akzeptabel, denn man geht bei einem Zufallsexperiment, z. B. dem Rollen der Kugel im Roulette, eben davon aus, dass diese Ereignisse *völlig unabhängig* voneinander sind, d. h. also die Zahl, die bei einem Kugellauf kommt, ganz unabhängig ist von den vorausgegangenen und den folgenden („Die Kugel hat kein Gedächtnis."); und dies wird idealerweise von der Zufallserwartung erfüllt. Man könnte einwerfen, dass es in einem *vernetzten Universum* gar keine Ereignisse gibt, die völlig unabhängig voneinander sind, aber dies ist für die logisch-mathematische Betrachtung nicht wirklich relevant, außerdem kann man diese These der universalen Interdependenz auch bestreiten.

Der größte Zusammenhang ist gegeben, wenn p(X *→ Y) = 1, also im *deterministischen* Fall (entsprechendes gilt für p = 0). Hier gilt der geringste Wert von p^T. Deterministischer Fall und Zufallserwartung sind also *Gegensätze*. Generell kann man für p(X *→ Y) sagen: je kleiner p^T, desto größer der Zusammenhang:

p = 1:	höchste empirische (positive) Abhängigkeit	p^T = minimum
p = 0:	höchste empirische negative Abhängigkeit / Gegensatz	p^T = minimum
p = 0,5:	keine Abhängigkeit zwischen X und Y, Zufallserwartung	p^T = maximum

p^T ist also auch geeignet, die *Größe* der Abhängigkeit von X und Y anzugeben. Allerdings differenziert es genau wie p^A nicht zwischen positiver und negativer Abhängigkeit. Z. B. kann man aus p^T = 4/16 nicht sehen, ob gilt p(X *→ Y) = 3/4 oder p(X *→ Y) = 1/4.

Fassen wir zusammen: p und p^T geben beide die *Größe* einer Abhängigkeit an; dabei ergibt sich aber folgender Unterschied:

Die *Objekt-Wahrscheinlichkeit* p gibt die Abhängigkeit *absolut* an. Wenn wir z. B. wissen, dass p = 1 (z. B. 4/4), wissen wir, dass vollständige (positive) Abhängigkeit gegeben ist.

Die *Meta-Wahrscheinlichkeit* p^T gibt dagegen die Abhängigkeit *relativ*, innerhalb einer *Verteilung* an. Wir können nicht sagen, dass ein *bestimmter* Wert von p^T vollständige Abhängigkeit anzeigt; bei p(X *→ Y) = 4/4 ist dieser Wert 1/16, aber bei einer anderen Relation(sgröße) ist es ein anderer Wert. Doch wir können feststellen: Der *niedrigste Wert* von p^T (innerhalb einer Verteilung) zeigt die *größte Abhängigkeit* an (positive oder negative); und der *höchste Wert* von p^T zeigt die geringste bzw. gar *keine Abhängigkeit* an, nämlich die Zufallserwartung. Dies gilt zunächst für die *Positiv-Implikation* X *→ Y; bei anderen Relatoren können sich Unterschiede ergeben.

1.3.3 p^T als Angabe der Sicherheit einer Abhängigkeit

In der Statistik wird sinngemäß zuweilen angegeben: Die *empirische* Wahrscheinlichkeit gibt die *Größe* eines Zusammenhangs an, die *theoretische* Wahrscheinlichkeit die *Sicherheit*. Man kann diese These folgendermaßen zusammenfassen:

– p gibt die *Größe einer Abhängigkeit* an:
  je größer p (über der Zufallserwartung: 0,5), desto größer der *positive* Zusammenhang;
  je kleiner p (unter 0,5), desto größer der *negative* Zusammenhang

– p^T gibt die *Sicherheit einer Abhängigkeit* an (wobei der Wert von p eingeht)
  je kleiner p^T, desto größer die Sicherheit des Zusammenhangs (desto sicherer)
  je größer p^T, desto kleiner die Sicherheit des Zusammenhangs (desto unsicherer)
  Anders gesagt: je größer der Informationsgehalt p^I, desto sicherer der Zusammenhang.

Wie lässt sich dieses These begründen? Man kann darauf hinweisen, dass p^T bei gleichem p unterschiedlich groß ist.
Beispiel: p = 0,5, genauer p(X *→ Y) = 0,5, also *Zufalls-Erwartung*:

n	p = 0,5	p^T	p^T dez.
2	1/2	2/4	0,5
4	2/4	6/16	0,38
6	3/6	20/64	0,31
8	4/8	70/256	0,27
10	5/10	252/1024	0,25

Obwohl p immer gleich ist, nämlich p = 0,5 (und so auch p^A immer gleich, nämlich $p^A = 0$), nimmt p^T je nach der Größe von n einen anderen Wert an. Und zwar *fällt* der p^T-Wert der Zufalls-Erwartung mit *steigendem* n: also je größer n, desto kleiner p^T.

Der Nenner von p^T ist 2^n, entsprechend 2^2, 2^4, 2^6, 2^8 ,2^{10}. Die genaue Berechnung ergibt sich aus den Formeln in 3-3-2.

Dasselbe zeigt sich bei p = 1, im *deterministischen* Fall. Auch hier ist p^T unterschiedlich je nach der Größe von n, obwohl p immer gleich bleibt (p = 1) und auch stets $p^A = 1$.
Im deterministischen Fall *fällt* p^T mit *steigendem* n noch stärker:

p = 1	p^T	p^T dez.
1/1	1/2	0,5
2/2	1/4	0,25
3/3	1/8	0,13
4/4	1/16	0,06
5/5	1/32	0,03

Z. B. ist bei p(X *→ Y) = 3/3 = 1 und p(X *→ Y) = 5/5 = 1 die *Größe* des Zusammenhangs gleich, nämlich n/n = 1. Dagegen differiert p^T:

p^T[p(X *→ Y) = 3/3 = 1] = 1/8 = 0,13
p^T[p(X *→ Y) = 5/5 = 1] = 1/32 = 0,03

Man könnte das so interpretieren, dass hier die *Sicherheit* unterschiedlich ist, mit der p = 1 gilt. Dass im Beispiel der Zusammenhang von X und Y bei p(X *→ Y) = 5/5 = 1 sicherer ist als bei p(X *→ Y) = 3/3 = 1. Allgemein, dass p bzw. p^A eben die *Größe* des Zusammenhangs angeben, p^T dagegen die *Sicherheit* (der Größe) des Zusammenhangs.

Aber diese Interpretation wirft auch Probleme auf, gerade bei der Zufallserwartung p = 0,5. Betrachtet man nur den Wert p = 0,5, z. B. p(X *→ Y) = 0,5, so kann man eben vermuten, dass die Größe des Zusammenhangs 0,5 ist. Wie ich jedoch gezeigt habe, besteht bei p = 0,5 gerade *kein Zusammenhang*. Dies wird aber durch p^T ausgedrückt, weil (in dieser Konstellation) gilt: Wenn p^T *maximal* ist, dann besteht *kein Zusammenhang*. Man müsste vielleicht die Anwendung von p^T danach differenzieren, ob es sich auf ein *bestimmtes* n (z. B. n = 4) bezieht oder auf mehrere Größen von n (1, 2, 3 ...). Innerhalb *einer* Größe von n (z. B. n = 4) kann man offensichtlich auch sagen, dass p^T die *Größe* des Zusammenhangs angibt, und zwar *relativ zu einer Verteilung* (4/4, 3/4, 2/4, 1/4, 0/4). Eine letzte Klärung steht hier noch aus.

2. Verknüpfte Gleichgewichts-Relationen

2.1 Bedingung $p(\neg X *\!\rightarrow Y) = 0$
2.2 Bedingung $p(\neg X *\!\rightarrow Y) = 1$
2.3 Bedingung $p(\neg X *\!\rightarrow Y) = 0{,}5$

Wir haben bisher Relationen der Form $X *\!\rightarrow Y$ bzw. quantitativ $p(X *\!\rightarrow Y) = r/n$ *für sich allein* untersucht. Um aber die *Abhängigkeit*, den *Zusammenhang* von X und Y genauer beurteilen zu können, müssen wir auch noch weitere Relationen heranziehen und miteinander verknüpfen, also eine *komplexere Analyse* vornehmen.

$p(X *\!\rightarrow \neg Y)$ müssen wir nicht unbedingt berücksichtigen, weil ja (nach dem Existenz-Modell) gilt: $p(X *\!\rightarrow Y) = 1 - p(X *\!\rightarrow \neg Y)$.

Dagegen müssen wir $p(\neg X *\!\rightarrow Y)$ bzw. $p(\neg X *\!\rightarrow \neg Y)$ berücksichtigen. Wenn wir z. B. eine Aussage machen wollen, wieweit Rauchen zu Krankheit führt, müssen wir auch berücksichtigen, ob Nicht-Raucher gesünder sind als Raucher.

Die Frage ist, ob wir zusätzlich auch $X \leftarrow\!* Y$ und ggf. $\neg X \leftarrow\!* \neg X$ heranziehen müssen. Bei der normalen *Äquivalenz* $X \leftrightarrow Y$ wäre das unproblematisch, denn es gilt:

$$X \leftrightarrow Y \;\Leftrightarrow\; (X \rightarrow Y) \wedge (X \leftarrow Y)$$

Auch bei der normalen Implikation $X \rightarrow Y$ ist das unproblematisch, weil hier das Gesetz der *Kontraposition* gilt:

$$X \rightarrow Y \;\Leftrightarrow\; \neg X \leftarrow \neg Y \quad \text{bzw.} \quad \neg X \rightarrow Y \;\Leftrightarrow\; X \leftarrow \neg Y$$

Aber wie gezeigt wurde, gilt (bei der vorrangigen Interpretation) der *Positiv-Implikation* die *Kontraposition* nicht. Dennoch ist es normalerweise verzichtbar, $X \leftarrow\!* Y$ heranzuziehen, wenn wir nur die *Abhängigkeit von Y zu X* untersuchen wollen.

Nur wenn wir die *vollständige Korrelation* von X und Y analysieren wollen, brauchen wir auch $X \leftarrow\!* Y$ bzw. $X *\!\leftrightarrow Y$. Insofern man alle möglichen Beziehungen von X und Y berücksichtigt, kann man anstelle von ‚*Abhängigkeit'* von ‚*Korrelation'* sprechen (ich werden diese begriffliche Differenzierung aber nicht immer anwenden).

Allerdings wäre hier eine *qualitative* Darstellung $X *\!\leftrightarrow Y$ bzw. eine *deterministische D*arstellung $p(X *\!\leftrightarrow Y) = 1$ nicht brauchbar. Eine Relation wie Rauchen $\leftrightarrow$ Kranksein würde aussagen, dass Rauchen die *einzige* Ursache von Krankheit ist, was völlig unrealistisch ist.

Bei Krankheiten, aber im Grunde generell haben wir es real mit Sachverhalten zu tun, für die es *mehrere Ursachen* gibt. Rauchen ist nur *eine* Ursache von Krankheit, es gibt viele andere wie Infektion, falsche Ernährung, Bewegungsmangel, Verletzung usw.

Für eine *vollständige* Analyse müsste man noch *zusätzliche Variablen* miteinbeziehen, im Sinne von *Multi-Kausalität*. Hier könnte man dann doch die *Äquivalenz* einsetzen, qualitativ in der Form: $X_1 \wedge X_2 \wedge ... \wedge X_i *\!\leftrightarrow Y$. Oder man müsste $p(X *\!\leftrightarrow Y) = r/n$ einen nur *statistischen* Wert (< 1) zuweisen. Aber die Logik solcher *multivariaten* Verhältnisse zu analysieren, würden den Rahmen dieser Arbeit sprengen und führte auch mehr in das Gebiet der *Statistik*.

Wir wollen nun *mögliche Kombinationen* von $p(X *\!\rightarrow Y)$ und $p(\neg X *\!\rightarrow Y)$ untersuchen.

2.1 BEDINGUNG $p(\neg X \ast\!\rightarrow Y) = 0$

Wenn wir wissen, dass $p(\neg X \ast\!\rightarrow Y) = 0$, ergibt sich eine veränderte Deutung der Werte von $p(X \ast\!\rightarrow Y)$. Nehmen wir als modifiziertes Beispiel: X = Rauchen bzw. Raucher, Y = Bronchitis bzw. Bronchitiker. Dann bedeutetet $p(\neg X \ast\!\rightarrow Y) = 0$:
„Kein Nicht-Raucher hat Bronchitis“ bzw. „alle Nicht-Raucher haben keine Bronchitis“.

Wenn dem so ist, dann signalisiert aber z. B., dass 10% aller Raucher Bronchitis haben, bereits eine *schwache positive Abhängigkeit* zwischen Rauchen und Bronchitis, nicht eine negative Abhängigkeit wie in der bisherigen Analyse. Generell zeigt in diesem Fall jeder Wert über 0% ($p > 0$) eine *positive* Abhängigkeit.

$p(\neg X \ast\!\rightarrow Y) = 0$	$p(X \ast\!\rightarrow Y)$	p dez.	Abhängigkeit von Y zu X
	4/4	1,0	größte Abhängigkeit
	3/4	0,75	große Abhängigkeit
	2/4	0,5	mittlere Abhängigkeit
	1/4	0,25	geringe Abhängigkeit
	0/4	0,0	keine Abhängigkeit

Vergleichen wir noch einmal die beiden Ansätze:

$p(X \ast\!\rightarrow Y)$	Größe der Abhängigkeit (1) $p(\neg X \ast\!\rightarrow Y)$ unbestimmt	Größe der Abhängigkeit (2) $p(\neg X \ast\!\rightarrow Y) = 0$
4/4 (1,0)	größte *positive* Abhängigkeit	größte positive Abhängigkeit
3/4 (0,75)	mittlere *positive* Abhängigkeit	große positive Abhängigkeit
2/4 (0,5)	*keine* Abhängigkeit	mittlere positive Abhängigkeit
1/4 (0,25)	mittlere *negative* Abhängigkeit	geringe positive Abhängigkeit
0/4 (0,0)	größte *negative* Abhängigkeit	keine (positive) Abhängigkeit

Besonders auffallend ist der Unterschied bei $p = 0{,}5$. Betrachtet man $p(X \ast\!\rightarrow Y)$ isoliert, steht das für „keine Abhängigkeit“. Insgesamt ist der Verlauf *non-linear*, verläuft in einer Kurve.

Setzt man zusätzlich $p(\neg X \ast\!\rightarrow Y) = 0$, dann steht $p(X \ast\!\rightarrow Y) = 0$ für „mittlere positive Abhängigkeit“. Es ergibt sich hier also eine *lineare* Darstellung: je größer $p(X \ast\!\rightarrow Y)$, desto größer ist die Abhängigkeit von Y in Bezug auf X.

Beziehen wir jetzt die theoretische Wahrscheinlichkeit p^T mit in die Analyse ein. Wir haben gesehen: $p^T[p(X \ast\!\rightarrow Y) = 4/4 = 1] = 1/16$, entsprechend $p^T[p(\neg X \ast\!\rightarrow Y) = 0/4 = 0] = 1/16$.
Da $p(\neg X \ast\!\rightarrow Y)$ und $p(X \ast\!\rightarrow Y)$ vollständig voneinander *unabhängig* sind, erhält man den Wert von $p^T[p(X \ast\!\rightarrow Y) = 4/4$ und $p^T[p(\neg X \ast\!\rightarrow Y) = 0/4$ durch *Multiplikation*.
$p^T[p(X \ast\!\rightarrow Y) = 4/4 \wedge p(\neg X \ast\!\rightarrow Y) = 0/4] = 1/16 \times 1/16 = 1/256 = 0{,}0039062 \approx 0{,}00$

Insgesamt ergibt sich:

$p(\neg X\ast\!\rightarrow Y)$	p^T	$\wedge$	$p(X\ast\!\rightarrow Y)$	p^T	Produkt	$p^T \wedge$	dezimal
0/4 = 0	1/16		4/4	1/16	$1/16 \times 1/16$	1/256	≈0,00
			¾	4/16	$1/16 \times 4/16$	4/256	0,016
			2/4	6/16	$1/16 \times 6/16$	6/256	0,023
			¼	4/16	$1/16 \times 4/16$	4/256	0,016
			0/4	1/16	$1/16 \times 1/16$	1/256	≈0,00
				16/16		16/256	

Übertragen wir den resultativen Wert von p^T in die oben gezeigte Aufstellung:

$p(\neg X *\rightarrow Y) = 0$	$p(X *\rightarrow Y)$	p dez.	Abhängigkeit von Y zu X	p^T
	4/4	1,0	größte Abhängigkeit	1/256
	3/4	0,75	große Abhängigkeit	4/256
	2/4	0,5	mittlere Abhängigkeit	6/256
	1/4	0,25	geringe Abhängigkeit	4/256
	0/4	0,0	keine Abhängigkeit	1/256

Die p^T-Werte bestätigen die Analyse:

• Am unwahrscheinlichsten ist: $p(\neg X *\rightarrow Y) = 0/4 \;\wedge\; p(X *\rightarrow Y) = 4/4$

Insofern liegt hier eine *maximale Abhängigkeit* von Y zu X vor. Man kann zuvorderst annehmen, dass X eine *kausale Ursache* von Y ist.

• Ebenfalls den kleinsten Wert $p^T = 1/256$ hat: $p(\neg X *\rightarrow Y) = 0/4 \;\wedge\; p(X *\rightarrow Y) = 0/4$

Hier gilt also im Beispiel: „Kein Nicht-Raucher hat Bronchitis und kein Raucher hat Bronchitis“. Das bedeutet: Wenn man davon ausgeht, dass sich die Menschen in Raucher und Nicht-Raucher untergliedern, dann käme die Krankheit Bronchitis beim Menschen gar nicht vor (ggf. wäre zu analysieren, ob sie bei anderen Lebewesen vorkommt – sonst wäre es ja nur eine virtuelle Krankheit). Die Deutung von p^T ist hier schwierig (vgl. später).

• Den größten Wert von p^T hat die *mittlere Abhängigkeit* von Y zu X:

$$p^T[p(\neg X *\rightarrow Y) = 0/4 \;\wedge\; p(X *\rightarrow Y) = 2/4] = 6/256$$

Hier zeigt sich: Genauso wie $p(X *\rightarrow Y) = 2/4$ für sich alleine den höchsten Wert von p^T hat, so besitzt auch $p(\neg X *\rightarrow Y) = 0/4 \;\wedge\; p(X *\rightarrow Y) = 2/4$ den höchsten Wert. Nur steht eben $p(X *\rightarrow Y) = 2/4$ isoliert für „keine Abhängigkeit“, $p(X *\rightarrow Y) = 0/4 \;\wedge\; p(\neg X *\rightarrow Y) = 2/4$ steht dagegen für „mittlere Abhängigkeit“. *Mittlere Abhängigkeit* mag bedeuten: X ist eine *Teilursache* von Y.

Die weiteren Werte brauchen wir hier nicht zu analysieren, denn es zeigt sich ja eine genaue Parallele zu den schon analysierten Werten. Nur während bei $p^T[p(X *\rightarrow Y)]$ der *Nenner* die Größe 16 besitzt, ist eben bei $p^T[p(\neg X *\rightarrow Y) \;\wedge\; p(X *\rightarrow Y)]$ der Nenner 16 x 16 = 256.

2.2 BEDINGUNG $p(\neg X *\rightarrow Y) = 1$

Nachdem wir eben analysiert haben, was sich für $p(X *\rightarrow Y)$ ergibt, wenn $p(\neg X *\rightarrow Y) = 0$, fragen wir uns jetzt nach den Folgen, wenn $p(\neg X *\rightarrow Y) = 1$, im Beispiel 4/4.

Hier können wir uns eine genaue Berechnung bzw. Darstellung der p^T-Werte ersparen, sie stimmen nämlich genau mit den oben genannten (bei $p(\neg X *\rightarrow Y) = 0$) überein. Was ergibt sich aber für die *Interpretation*?

$p(\neg X *\rightarrow Y) = 1$	$p(X *\rightarrow Y)$	p dez.	Abhängigkeit von Y zu X	p^T
	4/4	1,0	keine Abhängigkeit	1/256
	3/4	0,75	geringe negative Abhängigkeit	4/256
	2/4	0,5	mittlere negative Abhängigkeit	6/256
	1/4	0,25	große negative Abhängigkeit	4/256
	0/4	0,0	größte negative Abhängigkeit	1/256

Zur Erläuterung am Beispiel: Wenn 100% (4/4) der Nicht-Raucher eine Bronchitis haben und ebenfalls 100% (4/4) der Raucher, dann besteht *kein Zusammenhang* zwischen Rauchen und Bronchitis, sie stehen in einem *zufälligen* Verhältnis. Man müsste ggf. nach anderen Kausalfaktoren suchen, welche Bronchitis auslösen.

Allerdings ist die Deutung von p^T hier problematisch. Wir haben bisher gesehen, dass kein Zusammenhang als *Zufalls-Erwartung* mit dem *höchsten* Wert von p^T einhergeht. Hier ergibt sich aber der *niedrigste* Wert, nämlich $p^T = 1/256$. Eine vollständige Erklärung für dieses Phänomen steht noch aus. Eventuell ist die Verwendung von p^T nur unter bestimmten Rahmenbedingungen sinnvoll.

Eine andere, bessere Interpretation ist die folgende: Man geht von der Menge aller Menschen aus, die sich in Raucher und Nicht-Raucher unterteilen (in welchem Prozentsatz, ist hier irrelevant). *Alle* Raucher und Nicht-Raucher leiden unter Bronchitis. Zwar ist das Verhältnis zwischen Rauchen und Bronchitis somit *zufällig*, aber nicht das Verhältnis zwischen *Mensch und Bronchitis*. Da gilt sogar der *All-Satz*, die deterministische Abhängigkeit: p(Mensch *→ Bronchitis) = 8/8 = 1. Und dafür ergibt sich eine $p^T = 1/256$, und für p = 1 ist die niedrige p^T auch plausibel. Man könnte folgern, dass das Menschsein selbst eine kausale Ursache der Bronchitis ist, z. B. wegen einer Erbkrankheit.

Mensch

Raucher	Nicht-Raucher
100% Bronchitis	100% Bronchitis

Wenn 100% (4/4) der Nicht-Raucher husten, aber nur 75% (3/4) der Raucher, dann müsste Rauchen sogar eine gewisse Schutzwirkung gegen diese Krankheit haben. Und wenn sogar nur 0% der Raucher an Bronchitis erkranken, dagegen 100% der Nicht-Raucher, dann wäre umgekehrt gerade Nicht-Rauchen die Ursache von Bronchitis (was natürlich den *realen medizinischen Daten* völlig widerspricht).

2.3 BEDINGUNG p(¬X *→ Y) = 0,5

Hier erhalten wir andere Werte von p^T:

p(¬X*→Y)	p^T	∧	p(X*→Y)	p^T	Produkt	p^T ∧	dezimal
2/4 = 0,5	6/16		4/4	1/16	6/16×1/16	6/256	0,02
			¾	4/16	6/16×4/16	24/256	0,09
			2/4	6/16	6/16×6/16	36/256	0,14
			¼	4/16	6/16×4/16	24/256	0,09
			0/4	1/16	6/16×1/16	6/256	0,02
				16/16		96/256	

Übertragen wir den resultativen Wert von p^T (p^T ∧) wieder in unsere Aufstellung:

p(¬X *→ Y) = 0,5	p(X *→ Y)	p dez.	Abhängigkeit von Y zu X	p^T
	4/4	1,0	mittlere positive Abhängigkeit	6/256
	3/4	0,75	geringe positive Abhängigkeit	24/256
	2/4	0,5	keine Abhängigkeit	36/256
	1/4	0,25	geringe negative Abhängigkeit	24/256
	0/4	0,0	mittlere negative Abhängigkeit	6/256

Zur Erläuterung an einem Beispiel: Wenn 50% aller Nicht-Raucher Bronchitis haben und ebenso 50% aller Raucher, dann kann man wieder davon ausgehen, dass *kein Zusammenhang* zwischen Rauchen und Bronchitis besteht. Liegt aber der Anteil der kranken Raucher *über 50%*, ist eine *positive Abhängigkeit* gegeben (d. h. Rauchen fördert Bronchitis); liegt der Anteil der kranken Raucher *unter 50%*, ist eine *negative Abhängigkeit* gegeben (d. h. Rauchen verringert Bronchitis).

Entsprechende Verhältnisse ergeben sich, wenn wir andere Werte heranziehen, z. B.:

$$p^T[p(\neg X *\!\!\rightarrow Y) = 3/4 \;\wedge\; p(X *\!\!\rightarrow Y) = 1/4] = 16/256$$

Manches musste hier vereinfacht werden, um den Rahmen des Textes nicht zu sprengen. Eine umfangreichere Tabelle findet sich im Kapitel 5: System: 5-3-3-5.

Ein wesentliches Resultat soll noch festgehalten werden: Immer wenn

$$p(\neg X *\!\!\rightarrow Y) = p(X *\!\!\rightarrow Y)$$

liegt ein *zufälliges* Verhältnis vor, also keine Abhängigkeit zwischen X und Y (auch wenn das von p^T her nicht immer plausibel zu begründen ist.)

3. Ungleichgewichts-Relationen

3.1 Non-lineare Werte
3.2 Zufalls-Erwartung
3.3 Abhängigkeit

Ungleichgewichts-Relatoren sind Relatoren wie Implikator $\rightarrow$ und Konjunktor $\wedge$. Hier ist in der Wahrheitstafel die Anzahl der + (plus) und der − (minus) *ungleich*, steht also im *Ungleichgewicht*. Ich konzentriere mich hier auf den *Implikator* bzw. die *Implikation* $X \rightarrow Y$.

3.1 NON-LINEARE WERTE

$X \rightarrow Y$ hat *strukturell* eine theoretische Wahrscheinlichkeit $p^T = 3/4 = 0{,}75$. In *quantifizierter* Form gilt: p^T ist maximal 0,75, aber nur bei n = 1. Bei n = 2 gibt es noch den Wert 9/16 = 0,56. Auch hier gilt natürlich: Verwendet man nur *eine* Variable $X \Rightarrow X$ oder *mehr als zwei* Variablen, z. B. $X \rightarrow Y \rightarrow Z$, ergeben sich andere Werte.

n	$p(X \rightarrow Y)$	p^T	p^T dez.
1	1/1	3/4	0,75
	0/1	1/4	0,25
2	2/2	9/16	0,56
	1/2	6/16	0,38
	0/2	1/16	0,06

Mit steigendem n fällt p^T. Ab n = 3 liegt $p^T[X \rightarrow Y]$ immer unter 0,5: so gesehen sind auch hier alle Verteilungen (theoretisch) *unwahrscheinlich*.

p^T bildet bei $p(X \rightarrow Y)$ keine *symmetrische* Kurve, es liegt aber auch keine streng lineare Abhängigkeit vor: z. B. die Werte für $p^T[p(X \rightarrow Y) = r/n]$ bei n = 4:

$p(X \to Y)$	p^T	p^T dez.
4/4	81/256	0,32
3/4	108/256	0,42
2/4	54/256	0,21
1/4	12/256	0,05
0/4	1/256	≈0,00

3.2 ZUFALLS-ERWARTUNG

Anders als bei den *Gleichgewichts-Relatoren* liegt die *Zufallserwartung* (p-Wert mit dem höchsten Wert von p^T) hier also nicht bei p = 0,5 (im Beispiel p = 2/4), sondern bei p = 0,75 bzw. p = 3/4.

Allgemein gilt für $p(X \to Y)$ bei n = 4:
p = 1 p^T < maximum (hier: 81/256)
p = 0 p^T = minimum (hier: 1/256)
0 < p < 1 p^T = maximum (hier: 108/256)

Vergleichen wir unterschiedliche *Zufallserwartungen* von $p(X \to Y)$:

n	p	p dez.	p^T	p^T dez.
1	1/1	1,0	3/4	0,75
2	2/2	1,0	9/16	0,56
3	3/3	1,0	27/64	0,42
4	3/4	0,75	108/256	0,42
5	4/5	0,8	405/1024	0,40
6	5/6	0,83	1458/4096	0,36

Man sieht: bei n = 1, n = 2 und n = 3 ist sogar p = 1 die *Zufallserwartung* von $p(X \to Y)$, während bei der *Positiv-Implikation* $p(X *\!\!\to Y)$ der Wert p = 1 und die Zufallserwartung gerade *Gegensätze* sind. Bei $p(X \to Y)$ sind dagegen p = 0 und die Zufallserwartung *Gegensätze*.

Die Zufallserwartung von $p(X \to Y)$ hat maximal einen Wert von $p^T = 0{,}75$. (Hinzufügen muss ich noch: bei n = 3 besteht für p = 2/3 ebenfalls eine theoretische Wahrscheinlichkeit von $p^T = 27/64$, man kann p = 2/3 hier also ebenfalls als Zufallserwartung ansehen.).

Insgesamt sind die Werte bei *Ungleichgewichts-Relatoren* wie $X \to Y$ viel ungeregelter als bei *Gleichgewichts-Relatoren*. Der Nenner berechnet sich 4^n, für die Gesamtberechnung verweise ich auf 3-3-1.

Vergleichen wir noch verschiedene *deterministische* Werte von $p(X \to Y) = 1$.

n	p = 1	p^T	p^T dez.
1	1/1	3/4	0,75
2	2/2	9/16	0,56
3	3/3	27/64	0,42
4	4/4	81/256	0,32
5	5/5	243/1024	0,24

Hier ist die Berechnung einfach, wie schon früher aufgezeigt: $p^T = 3^n/4^n$, genauer:

$$p^T[p(X \rightarrow Y) = n/n] = (3/4)^n$$

Dies zeigt noch einmal, dass wir in diesen Fällen immer die *ungekürzten* p-Werte angeben müssen: p = 1/1, 2/2, 3/3 usw. und nicht einfach pauschal p = 1 einsetzen dürfen.

3.3 ABHÄNGIGKEIT

Bei der *normalen Implikation* X → Y ist die Abhängigkeit viel schwieriger zu bestimmen als bei der *Positiv-Implikation* X *→ Y.

Bei der *Positiv-Implikation* X *→ Y stehen *empirische* und *theoretische* Wahrscheinlichkeit in einem klaren Verhältnis, vereinfacht:

p	p^T
maximal (1)	minimal
mittig (0,5)	maximal
minimal (0)	minimal

Dagegen stehen p und p^T bei X → Y in einem komplizierten Verhältnis (z. B. bei n = 4):

p(X → Y)	p^T	p^T dez.
4/4	81/256	0,32
3/4	108/256	0,42
2/4	54/256	0,21
1/4	12/256	0,05
0/4	1/256	≈0,00

Es gibt somit 2 Methoden, die Abhängigkeit von X → Y zu bestimmen, über die *Objekt-Wahrscheinlichkeit* p und über die *Meta-Wahrscheinlichkeit* p^T:

• Bestimmung der Abhängigkeit von Y zu X in X → Y: über p
Hier entspricht die Abhängigkeit von X → Y der von X *→ Y.

p(X → Y)	Abhängigkeit gemäß p
4/4	größte positive Abhängigkeit
3/4	mittlere positive Abhängigkeit
2/4	keine Abhängigkeit
1/4	mittlere negative Abhängigkeit
0/4	größte negative Abhängigkeit

Im Gegensatz zu X *→ Y ergeben sich aber bei X → Y Diskrepanzen. Z. B. liegt hier der größte Wert von p^T bei p = 3/4; somit wäre also 3/4 die *Zufallserwartung* (= keine Abhängigkeit), während es – von p aus betrachtet – eine *positive Abhängigkeit* ist.

• Bestimmung der Abhängigkeit von Y zu X in X → Y über p^T
Hier wäre die größte Abhängigkeit bei 0/4 gegeben, weil dort p^T am geringsten ist.
Im Einzelnen zeigt das die folgende Übersicht:

$p(X \to Y)$	p^T	p^T dez.	Abhängigkeit gemäß p^T
4/4	81/256	0,32	geringe Abhängigkeit
3/4	108/256	0,42	keine Abhängigkeit
2/4	54/256	0,21	mittlere Abhängigkeit
1/4	12/256	0,05	große Abhängigkeit
0/4	1/256	≈0,00	größte Abhängigkeit

Da p^T nicht zwischen positiver und negativer Abhängigkeit unterscheidet, ist es schwierig, hier jeweils anzugeben, ob eine *positive* oder *negative* Abhängigkeit vorliegt, dies geht nur im Rückgriff auf p. Zwar mag man z. B. $p(X \to Y) = 4/4$ eindeutig als *positive* Abhängigkeit bestimmen. Und bei p = 1/4 und p = 0/4 liegt eine *negative* Abhängigkeit vor. Aber beim Wert $p(X \to Y) = 3/4$ ist es schwierig: gemäß p ist es eine *positive Abhängigkeit*, gemäß p^T ist es gar *keine Abhängigkeit*; ähnlich ist der Wert p = 2/4 schwierig einzuordnen.

Da die *Abhängigkeit* von Y zu X sich hier also nur schwer genau bestimmen lässt, ist die Implikation X → Y auch viel problematischer für eine *Kausalanalyse* zu handhaben als die Positiv-Implikation X *→ Y.

Erstens, weil – wie beschrieben – der Verlauf des Wertes von p^T nicht so symmetrisch und gleichmäßig verläuft. Wenn man die Abhängigkeit mit p^T darstellen will, so ergibt sich z. B. bei n = 4 (vgl. die Tabelle oben): ↑ ↓ ↓ ↓, d. h. p^T geht erst 1x hoch (von 81/256 auf 108/256) und dann 3x runter, es ist schwer, so eine Entwicklung der Abhängigkeit zu begründen. Vor allem auch, weil p dagegen 4x fällt (von 4/4 auf 3/4 auf 2/4 usw.), was eben eine Zuordnung von p^T zu p schwierig macht.

Zweitens, die Werte von p und p^T stehen bei X → Y in keinem direkt einleuchtenden Verhältnis. Anders bei X *→ Y, hier weisen p = 1 und p = 0 beide den niedrigsten Wert von p^T auf, was plausibel ist und unseren normalsprachlichen Intuitionen entspricht (jedenfalls nach einigem Nachdenken). Bei X → Y besitzen dagegen p = 1 und p = 0 unterschiedlichen Werte von p^T; außerdem ist (bei n = 4) p = 3/4 (0,75) die *Zufalls-Erwartung*, bei der also keine Abhängigkeit bestehen soll, was auch problematisch ist. Das alles ergibt sich zwar aus der *Definition* der Implikation, aber es zeigt auch noch mal, welche Probleme diese Definition aufwirft, auch in der *Wissenschaftstheorie*; denn man formalisiert *wissenschaftliche Sätze* normalerweise mittels der Implikation.

Drittens, X → Y und ¬X → Y sind voneinander logisch *abhängig*, wogegen X *→ Y und ¬X *→ Y logisch *unabhängig* sind. Das wird sofort deutlich, wenn man die entsprechenden (aus den Wahrheitstafeln gebildeten) Brüche einsetzt:

$$p(X * \to Y) = \frac{a}{a+b}$$

$$p(\neg X * \to Y) = \frac{c}{c+d}$$

$$p(X \to Y) = \frac{a+c+d}{a+b+c+d}$$

$$p(X \to Y) = \frac{a+b+c}{a+b+c+d}$$

$(X *\to Y) \wedge (\neg X *\to Y)$ bzw. $p(X *\to Y) = r/n \wedge p(\neg X *\to Y) = s/m$ ist somit *synthetisch* (oder aber eine pseudo-analytische Relation). Denn in den Brüchen kommen nur *unterschiedliche* Variablen vor, im ersten Bruch a und b, im zweiten Bruch c und d.

Dagegen ist $p(X \rightarrow Y) = r/n \;\wedge\; p(\neg X \rightarrow Y) = s/n$ eine *semi-analytische* Relation. Hier kommen in beiden Brüchen alle Variablen a, b, c und d vor; die Nenner sind gleich, die Zähler überschneiden sich.

Der semi-analytische Status zeigt sich vor allem auch daran, dass $X \rightarrow Y$ und $\neg X \rightarrow Y$ *nicht beide falsch* sein können.
$\neg(X \rightarrow Y) \wedge \neg(\neg X \rightarrow Y)$ bzw. $p(X \rightarrow Y) = 0 \;\wedge\; p(\neg X \rightarrow Y) = 0$ ist eine *Kontradiktion*.
Anders gesagt: $\neg(\neg X \rightarrow Y) \Rightarrow (X \rightarrow Y)$, quantitativ: $p(\neg X \rightarrow Y) = 0 \;\Rightarrow\; p(X \rightarrow Y) = 1$.
Im Beispiel gilt: „Kein Nicht-Raucher hat Bronchitis ⇒ alle Raucher haben Bronchitis“. Auch das widerspricht unseren Intuitionen.

Aus all diesen Gründen werde ich hier eine *Kausalanalyse* mittels der Normal-Implikation $X \rightarrow Y$ nicht weiter untersuchen.

4. Zufall und Korrelation

Im Punkt 4 sollen die Aussagen der ersten drei Punkten einerseits zusammengefasst, vor allem aber noch einmal vertieft, systematisiert und erweitert werden.
Dies geschieht in folgender Reihenfolge, anhand eines Beispiels erläutert:

- Objekt-Ebene

Wir stellen z. B. fest: $p(X \ast\!\!\rightarrow Y) = 7/10$, Beispiel: p(Sportler $\ast\!\!\rightarrow$ gesund) = 7/10. Also: 7 von10 Sportlern sind gesund. Dies sei eine *repräsentative Stichprobe*. Wir gehen davon aus, dass in der *Gesamtheit* dasselbe Verhältnis gilt (zur Problematik vgl. unten).

- Abhängigkeit auf der Basis der empirischen Wahrscheinlichkeit p

Wir deuten das obige Ergebnis so, dass eine *Abhängigkeit* von Y zu X besteht, weil 70% der Sportler gesund sind. Bestände ein *zufälliger* Zusammenhang, dann würden wir (zunächst) erwarten, dass etwa 50% der Sportler gesund sind.

- Abhängigkeit auf der Basis der theoretischen Wahrscheinlichkeit p^T

Für p = 7/10 besteht eine theoretische Wahrscheinlichkeit von $p^T = 120/1024 = 0{,}12$. Auch dies deuten wir so, dass eine Korrelation besteht und zwar mit gewisser Sicherheit. Denn z. B. für die Zufallserwartung p = 5/10 bestände eine deutlich höhere theoretische Wahrscheinlichkeit von $p^T = 252/1024 = 0{,}25$. Und je niedriger p^T, desto höher und sicherer der Abhängigkeit.

- Kausalität

Um zu testen, ob hier aber wirklich ein *Kausal-Zusammenhang* vorliegt, müssen wir weitere Tests vornehmen. Z. B. muss man auch die Relation $p(\neg X \ast\!\!\rightarrow Y) = s/m$ heranziehen, man muss prüfen, ob X *zeitlich vor* Y stattfindet (oder umgekehrt) u.v.m.

4.1 OBJEKT-EBENE

Die Objekt-Ebene ist die Ebene, auf der wir die *relative Häufigkeit* einer Relation feststellen bzw. sie mit der *empirischen Wahrscheinlichkeit* angeben.

Dabei sei noch einmal daran erinnert: für p(X *→ Y) = r/n können wir vor allem zwei *äquivalente* Deutungen abgeben, am Beispiel p(X → Y) = 7/10:

- *Klasse* mit Angabe der *relativen Häufigkeit*: 7 von 10 (oder 70% der) X sind Y
- *Individuum* mit Angabe der *Wahrscheinlichkeit*: ein (beliebiges) X ist mit einer Wahrscheinlichkeit von 0,7 (oder 70%) ein Y

D. h. ich verwende das Symbol ‚p' sowohl für die *relative Häufigkeit* als auch für die *empirische* Wahrscheinlichkeit (man kann zwar auch dazwischen differenzieren, aber dies macht es unnötig noch komplizierter).

4.1.1 Endliche Gesamtheit

Die Grundgesamtheit (oder kurz Gesamtheit) umfasst *alle* Elemente einer Klasse = n.
Wenn r X von der Gesamtheit n auch Y sind, schreibe ich: p/Gesamtheit(X *→ Y) = r/n
Beispiel: Angenommen, die Gesamtheit aller (deutschen) Sportler sei 100.000.
Man untersucht *alle* Sportler (100.000) und stellt fest: 70.000 sind gesund.
Hier kann man zu Recht eine *dezimale* oder *prozentuale* Angabe machen, z. B.:

p(Sportler *→ gesund) = 0,7 oder p(Sportler *→ gesund) = 70%

Aber es ist normalerweise völlig unrealistisch, eine so große Anzahl wie 100.000 Sportler oder andere Individuen zu untersuchen. D. h. bei sehr *großen Gesamtheiten* bzw. Mengen ist es in der Wissenschaftspraxis kaum möglich, *alle* Elemente zu untersuchen.

Und es gibt hier noch zwei besondere Probleme: Wenn man (zum heutigen Zeitpunkt) Aussagen über *alle Elemente einer Menge*, z. B. alle Menschen machen will, dann hat man es auch mit *vergangenen* Elementen (also bereits verstorbenen Menschen) und *zukünftigen* Elementen (Menschen, die erst noch geboren werden) zu tun.

Erstens, wie will man Elemente aus der *Vergangenheit* untersuchen? Gerade, wenn es sich vor langer Zeit Verstorbene handelt, etwa Menschen aus der Steinzeit, ist es undenkbar, diese Menschen und ihre Eigenschaften (genau) zu erfassen. Ein zusätzliches Problem ist hier, ab wann man in der Evolution von Menschen und nicht mehr von Vor-Menschen ausgeht.

Zweitens, noch problematischer sind Aussagen über *zukünftige* Menschen, denn hier stoßen wir auch auf physikalische Probleme, inwieweit die Zukunft determiniert ist. (Und es ergibt sich das zusätzliche Problem, ab wann wir, bei fortschreitender Evolution, noch von ‚Menschen' sprechen dürfen.) Man kann nicht mit Sicherheit etwas über zukünftige Menschen aussagen. Zwar wird diese Beschränkung normalerweise nicht explizit angemerkt, aber im Grunde sind angebliche Aussagen über eine *Gesamtheit* immer auf ein *Zeitfenster* limitiert, von *heute* bis in eine noch *überschaubare Vergangenheit*. Oder man limitiert solche Aussagen direkt auf eine *Zeitzone* bzw. grenzt sie in anderer Weise ein.

So gesehen sind Aussagen über Mengen bzw. Gesamtheiten eigentlich immer nur über Aussagen über *Teilmengen*, was natürlich auch logische Konsequenzen hat.

4.1.2 Teilmenge oder Stichprobe

Es geht hier um eine *beliebige* Stichprobe oder Teilmenge der Gesamtheit.

p/Stichprobe(X * → Y) = r/n

Man untersucht z. B. 10 Sportler uns stellt fest: 7 von ihnen sind gesund.

p/Stichprobe(Sportler *→ gesund) = 7/10

Hier sollte man besser keine *dezimale* oder *prozentuale* Angabe machen, weil diese Angaben sich auf die Gesamtheit (All-Menge bzw. Klasse) beziehen. Oder man muss klar einschränken: „70% der Sportler *der Stichprobe* sind gesund". (Im logischen Kontext kann man allerdings den Unterschied Gesamtheit versus Stichprobe schon einmal vernachlässigen.)

Es gibt *sichere* und *unsichere* Schlüsse von der Stichprobe auf die Gesamtheit.

• Sicherer, strenger Schluss

Angenommen, wir untersuchen nur die Sportler einer Mannschaft *A*. Das seien 20 Personen. Man hat 10 Sportler – als Stichprobe – untersucht, 7 davon sind gesund, also p = 7/10.

p/Stichprobe(Sportler *→ gesund) = 7/10

Dann kann man für die Gesamtheit n = 20 ableiten:

p/Stichprobe(Sportler *→ gesund) = 7/10

⇒ p/Gesamtheit(Sportler *→ gesund) ≥ 7/20 ∧ ≤ 17/20.

Denn wenn 7 von 10 Sportlern bei der *Stichprobe* gesund sind, dann müssen es in der *Gesamtheit* natürlich auch mindestens 7 sein. Es können aber nicht alle = 20 sein, sondern maximal 17, weil ja schon in der Stichprobe 3 Sportler nicht gesund sind.

Der Wert ≥ 7/20 ∧ ≤ 17/20 ist allerdings nicht sehr aussagekräftig. Und während p dezimal in der Stichprobe genau 0,7 ist, ist es in der Gesamtheit nur zu bestimmen als ≥ 0,35 ∧ ≤ 0,85, also der Wert liegt zwischen 0,35 und 0,85, was auch nicht sehr exakt ist.

• Unsicherer, semi-analytischer Schluss

Hier verwende ich nur einfache Relationen der Form Fx_i, die Individuen-Variable ‚x' sei hier nur auf Sportler eingegrenzt, also z. B. x_1 = $\text{Sportler}_{1.}$ Es wird jeweils die theoretische Wahrscheinlichkeit des Schlusses p^T angegeben (dies ist ein Vorgriff auf Kapitel 4):

$$p^T[Fx_1 \overset{*}{\longrightarrow} Fx_1 \wedge Fx_2 \wedge \ldots \wedge Fx_n] = 1/2^{n-1}$$
$$p^T[Fx_1 \wedge Fx_2 \overset{*}{\longrightarrow} Fx_1 \wedge Fx_2 \wedge \ldots \wedge Fx_n] = 1/2^{n-2}$$
$$p^T[Fx_1 \wedge Fx_1 \wedge Fx_2 \overset{*}{\longrightarrow} Fx_1 \wedge Fx_2 \wedge \ldots \wedge Fx_n] = 1/2^{n-3}$$

Bei Verwendung der *Normal-Implikation* → ergibt sich:

$$p^T[Fx_1 \longrightarrow Fx_1 \wedge Fx_2 \wedge \ldots \wedge Fx_n] = (2^n - 2^{n-1} + 1)/2^n$$
$$p^T[Fx_1 \wedge Fx_2 \longrightarrow Fx_1 \wedge Fx_2 \wedge \ldots \wedge Fx_n] = (2^n - 2^{n-2} + 1)/2^n$$
$$p^T[Fx_1 \wedge Fx_1 \wedge Fx_2 \longrightarrow Fx_1 \wedge Fx_2 \wedge \ldots \wedge Fx_n] = = (2^n - 2^{n-3} + 1)/2^n$$

Beispiel: Man hat 1 (einen) Sportler untersucht, er ist gesund. Wie wahrscheinlich ist dann, dass 2 Sportler (also n = 2) gesund sind? $p^T = (2^2 - 2 + 1)/2^2 = 3/4$ bei der *normalen* Implikation. Bei der *Positiv*-Implikation $1/2^{2-1} = 1/2$.

Aber es wird sofort deutlich: Wenn die Stichprobe wesentlich kleiner als die Gesamtheit ist – und das ist der Normalfall –, dann kann man auf diese Weise nur sehr begrenzte Rückschlüsse auf die Gesamtheit ziehen. Das führt uns zu der *repräsentativen* Stichprobe.

4.1.3 Repräsentative Stichprobe

Eine repräsentative Stichprobe soll quasi ein *verkleinertes Abbild der Gesamtheit* sein.

Die Theorie ist: Man kann von einer *Teilmenge* (*Stichprobe*) auf die *Allmenge* (*Gesamtheit*) schließen, wenn die Stichprobe *repräsentativ* ist

Angenommen, man hat eine Stichprobe von 100 Sportlern. 70 davon sind gesund, 30 nicht. Es besteht also ein Prozentsatz von 70%. Man würde dann entsprechend schließen, dass in der Gesamtheit von 100.00 Sportlern ebenfalls 70%, also 70.000 gesund sind.

Auf wie wenige Elemente darf die Stichprobe zurückgehen? Prinzipiell könnte man auch nur 10 Sportler untersuchen, von denen dann 7 gesund sind.

Wenn man allerdings nur 1 (einen) oder jedenfalls weniger als 10 Sportler untersucht, könnte man das Verhältnis nicht herausfinden.

p/Stichprobe(Sportler → gesund) = 7/10 ⟶
p/Gesamtheit(Sportler ⟶ gesund) = 70.000/100.000

Ich verwende hier den Pfeil ⟶ für den *semi-analytischen* Schluss, da dies nur ein *induktiver*, kein sicherer deduktiver Schluss ist.

Die eine Frage ist: Wie erhält man eine *repräsentative Stichprobe*? Man erhält so eine Stichprobe, in der man eine *Zufallsziehung* macht, aber das ist noch keine Gewähr dafür, wirklich ein genaues Abbild zu gewinnen.

Außerdem ist der Schluss nur mit Fehlerrisiken möglich, die man einzugrenzen versucht. Das zu erläutern, führte aber zu weit in die *Statistik*.

Festzuhalten bleibt noch: Auch die *repräsentative* Stichprobe ist keine überzeugende Antwort auf das Problem zukünftiger (und vergangener) Elemente der zu untersuchenden Gesamtheit.

4.1.4 Unendliche Gesamtheit

Gleichgültig wie groß die *endliche* Teilmenge oder Stichprobe ist, dies erlaubt keinerlei relevanten Schluss auf die Verteilung in der *Unendlichkeit*. Denn jede *endliche Menge* ist vernachlässigbar gering im Vergleich zu einer *unendlichen Menge* (vgl. 1-3-0-4).

Vor allem infinite *statistische* Aussagen (0 < p < 1), werfen Probleme auf. Z. B. für p = 0,5. Wie ist eine solche Aussage überhaupt zu formalisieren? Und wie zu verstehen? Als:

$p(X \to Y) = r/n = 1/2$? $p(X \to Y) = 0{,}5$? $\lim_{n \to \infty}[p(X \to Y) = \frac{n}{2n}] = 0{,}5$?

Jedenfalls lassen sich *infinite* statistische Aussagen gar nicht logisch beweisen, weder *verifizieren* noch *falsifizieren*, sondern nur nach *pragmatischen* Kriterien prüfen.

Dagegen lassen sich *deterministische* Aussagen (mit p = 1) immerhin falsifizieren:

- *infiniter Allsatz*: $\Lambda x(Fx \to Gx)$

‚Für alle x gilt: wenn sie die Eigenschaft F haben, dann auch die Eigenschaft G', z. B. ‚für alle x gilt, wenn sie Körper sind, dann unterliegen sie der Schwerkraft'.

Dann genügt logisch *ein* x (x_i), das zwar die Eigenschaft F hat, aber nicht die Eigenschaft G, und der Satz ist widerlegt: $(Fx_i \wedge \neg Gx_i) \Rightarrow \neg\Lambda x(Fx \to Gx)$

- *negativer infiniter Allsatz* $\Lambda x(Fx \to \neg Gx)$

‚Für alle x gilt: wenn sie die Eigenschaft F haben, dann haben sie nicht die Eigenschaft G'.
Auch hier genügt *ein* x, für das dies nicht gilt, also $(Fx_i \wedge Gx_i)$, und der Satz ist widerlegt.

$(Fx_i \wedge Gx_i) \Rightarrow \neg\Lambda x(Fx \to \neg Gx)$

4.2 ABHÄNGIGKEIT AUF BASIS DER EMPIRISCHEN WAHRSCHEINLICHKEIT p

Man will nun wissen, inwieweit die obigen Werte auf eine *Abhängigkeit* hinweisen (dabei vernachlässigen wir jetzt die Unterscheidung von Gesamtheit und Stichprobe).
Man geht von der Theorie aus:

- *Gleichverteilung*: spricht für *Zufall*, keine Abhängigkeit
- *Ungleichverteilung*: spricht für *Abhängigkeit,* Korrelation

Je mehr ein Wert von der Gleichverteilung abweicht, desto größer die Abhängigkeit. Man spricht auch von *Korrelation* (‚Korrelation' lässt sich allerdings auch umfassender verstehen, als Abhängigkeit zwischen allen Ausprägungen der Variablen).

Jedoch, wenn ‚n' ungerade ist, dann ist keine völlige Gleichverteilung möglich, also etwa bei n = 1, n = 3, n = 5, n = 7 usw. Denn z. B. bei = 3 müsste eine Gleichverteilung 1,5 zu 1,5 bedeuten, aber *halbe* X (z. B. halbe Sportler) gibt es nicht.

4.2.1 *Zwei* Variablen
Angenommen, man hat 2 Variablen, X und Y. Bei der uns hier interessierenden Struktur der *Implikation* gibt es 2 Möglichkeiten:

$X *\!\rightarrow Y$: X ist ein Y (z. B. der Sportler ist gesund)
$X *\!\rightarrow \neg Y$: X ist *kein* Y bzw. X ist *nicht* Y (z. B. der Sportler ist nicht gesund)

Dies kann nun für unterschiedlichen *Größen* bzw. *Häufigkeiten* gelten. Es kann also für *alle* X gelten, für *einige* usw., oder genau *quantifiziert* für r von n X. Dabei gilt folgender Zusammenhang: $p(X *\!\rightarrow Y) = r/n$, dann gilt: $p(X *\!\rightarrow \neg Y) = 1 - r/n$.

Man kann auch formulieren: X ist Y_1 oder X ist Y_2, oder noch kürzer: Man unterschiedet nur Y_1 und Y_2.

Immer lässt sich dabei differenzieren zwischen folgenden Werten:
- $p = 1$ (alle / immer)
- $0 < p < 1$ (einige / manchmal)
- $p = 0$ (alle nicht / nie)

Diese werden wie folgt gedeutet (vgl. aber auch Punkt 2 über verknüpfte Relationen):

- $p = 1$ (alle /immer)	vollständige positive Abhängigkeit
- $0 < p < 1$ (einige/manchmal)	
$p > 0,5$	positive Abhängigkeit
$p = 0,5$	keine Abhängigkeit („Zufall“)
$p < 0,5$	negative Abhängigkeit
- $p = 0$ (alle nicht / nie)	vollständige negative Abhängigkeit

Beispiel: Abhängigkeit von Y zu X bei n = 4

4/4	vollständige positive Abhängigkeit
3/4	positive Abhängigkeit
2/4	keine (positive oder negative) Abhängigkeit
1/4	negative Abhängigkeit
0/4	vollständige negative Abhängigkeit

4.2.2 Verschiedene Variablen
Wenn auch die Unterscheidung von 2 (*zwei*) Variablen die Basis ist und man Unterscheidungen von mehr als 2 Variablen prinzipiell immer auf 2 Variablen zurückführen kann (wie es ja auch beim *binären* Code, bei der *Digitalisierung* geschieht), so lassen sich natürlich auch mehr als 2 Variablen unterscheiden.

Wir können z. B. sagen: X ist Y_1, Y_2 oder Y_3. $(X *\!\rightarrow Y_1) \vee (X *\!\rightarrow Y_2) \vee (X *\!\rightarrow Y_3)$.

Oder wir unterscheiden einfach nur zwischenY_1, Y_2 und Y_3.
Entsprechendes gilt übrigens für die *intensionale Quantifizierung*, wo X und Y unterschiedliche intensionale Werte annehmen, z. B. X ist zu 100% Y, X ist zu 50% Y, X ist zu 0 % Y.

Man kann generell sagen, dass *Zufall* immer mit *Ausgleich* verbunden ist.
Bei 2 Variablen bedeutet dies: p = 0,5, also: 50% aller X sind Y, 50 % aller X sind nicht Y.
Bei mehr als 2 Variablen ergeben sich allerdings andere Verhältnisse:

Ausgleich bei unterschiedlichem n:

bei 2;	1/2 -1/2	50% - 50%
bei 3:	1/3 - 1/3 – 1/3	33,3% - 33,3% - 33,3%
bei 4:	1/4- 1/4 - 1/4 – 1/4	25% - 25% - 25% - 25%
bei 5:	1/5 – 1/5 – 1/5 – 1/5 – 1/5	20% - 20% - 20% - 20% - 20%

Z. B. ergäbe sich bei 3 Variablen folgender Ausgleich:

$$p(X \mathbin{*\!\!\rightarrow} Y_1) = 0{,}33 \;\wedge\; p(X \mathbin{*\!\!\rightarrow} Y_2) = 0{,}33 \;\wedge\; p(X \mathbin{*\!\!\rightarrow} Y_3) = 0{,}33$$

4.2.3 Ungleichgewichts-Relatoren

Dies sind Relatoren wie $\rightarrow$ bzw. Relationen wie $X \rightarrow Y$ (vgl. 3.) Wo liegt hier der Ausgleich?

Wenn man nur nach der *empirischen* Wahrscheinlichkeit ‚p' vorgeht, muss man wohl wie bei der Positiv-Implikation $X \mathbin{*\!\!\rightarrow} Y$ argumentieren.

Also läge der Ausgleich bei $p(X \rightarrow Y) = r/n$ dann vor, wenn $p(X \rightarrow Y) = 0{,}5$ und ebenfalls $p(X \rightarrow \neg Y) = 0{,}5$.

Überzeugend kann man allerdings den Ausgleich bei Ungleichgewichts-Relatoren nur darstellen, wenn man auch die *theoretischen Wahrscheinlichkeit* p^T mit einbezieht. Hier ergeben sich andere Werte, denn nur bei $n = 2$ ist $p(X \rightarrow Y) = 0{,}5$ die Zufallserwartung (wie früher schon gezeigt worden ist).

4.3 ABHÄNGIGKEIT AUF BASIS DER THEORETISCHEN WAHRSCHEINLICHKEIT

4.3.1 Empirische Beziehung und Zufalls-Beziehung

Wir haben uns bisher auf empirische Beziehungen konzentriert. Genauer müssen wir aber unterscheiden zwischen einer *empirischen Beziehung* und einer *Zufalls-Beziehung.*

Bei einer *empirischen* Beziehung zwischen X und Y steht nicht von vorneherein fest, ob zwischen X und Y eine *Korrelation* oder sogar eine *Kausal-Beziehung* besteht. Z. B. müssen wir erst untersuchen, wie viele Sportler gesund sind (bzw. wie viele nicht) und ob die Ursache ihrer Gesundheit ist, dass sie Sport treiben.

Bei einer *Zufalls-Beziehung* wissen wir dagegen *von vorneherein*, dass zwischen X und Y nur eine *zufällige* Relation besteht.

Z. B. bei einem *Glücksspiel* wie dem *Roulette*. X sei hier der *Kugelwurf*, Y die *Farbe rot* für rotes Feld (¬Y: schwarzes Feld). Rein physikalisch bestehen zwar gewisse Beziehungen zwischen dem Kugelwurf und der erzielten Farbe, aber faktisch können wir den Zusammenhang als *zufällig* interpretieren.

Und da es *gleichviel* rote wie schwarze Zahlen gibt (nämlich 18), sind die Chancen gleich, dass rot oder schwarz kommt. Von daher erwarten wir von vorneherein, dass ungefähr gleich oft rot wie schwarz kommt. (Von der Zahl 0, die teils rot, teils schwarz ist, sehe ich ab.)

Dennoch ist nicht ausgeschlossen, dass z. B. sehr viel häufiger *rot* als *schwarz* kommt (vgl. 4.4). Die Frage ist, ob man dann von einer *Korrelation* sprechen soll oder besser einer *Schein-Korrelation*. Jedenfalls, in keinem Fall werden wir den Kugelwurf als kausale *Ursache* für das Erscheinen einer roten Zahl deuten (wir können die Zufalls-Beziehung auch als Sonderfall einer *theoretischen Beziehung* deuten, bei der eben vor der empirischen Prüfung feststeht, welche Beziehung zwischen X und Y herrscht).

Zwar ist kaum zu behaupten, dass wir *a priori* – vor aller Erfahrung – wussten, dass *rot* und *schwarz* beim Roulette normalerweise *gleichhäufig* auftreten. Auch dies musste erst durch Beobachtung der *empirischen Häufigkeiten* herausgefunden werden (z. T. sind ja die Gesetze der Wahrscheinlichkeitstheorie gerade an Glücksspielen entwickelt worden). Aber von unserm heutigen Stand aus, wissen wir, dass beim Roulette, beim Würfeln oder auch bei einem Zufallsgenerator *zufällige* Häufigkeiten entstehen, die eine *kausale* Beziehung ausschließen.

4.3.2 Empirische Beziehung
Wir haben bisher nicht genauer unterschieden zwischen *Objekt-Ebene* und *empirischer Ebene*. Es lässt sich aber doch differenzieren:
– *empirische Ebene*
Auf der empirischen Ebene ist nur *ein* Wert richtig (jedenfalls unter denn gleichen Bedingungen, zur gleichen Zeit), z. B. $p(X \rightarrow Y) = 3/4$, also 3 von 4 X sind auch Y.
– *Objekt-Ebene*
Auf der Objekt-Ebene sind dagegen alle *möglichen* Ausprägungen repräsentiert, z. B. bei n = 4 sind das: 4/4, 3/4, 2/4, 1/4, 0/4. Konkret: $p(X \rightarrow Y) = 4/4$, $p(X \rightarrow Y) = 3/4$ usw. bis schließlich zu $p(X \rightarrow Y) = 0/4$. Dass es diese Möglichkeiten gibt, braucht man nicht *empirisch* festzustellen, sondern ergibt sich *analytisch* oder theoretisch. So gesehen können wir die vollständige Reihe 4/4, 3/4, 2/4, 1/4, 0/4 eine *theoretische* Verteilung, aber auf der Objekt-Ebene nennen, es geht jeweils um empirische Wahrscheinlichkeiten p.

Auf der *Meta-Ebene* gibt es dagegen nur theoretische Verteilungen bzw. Wahrscheinlichkeiten, die sich auf die Objekt-Ebene beziehen, z. B. 1/16, 4/16, 6/16, 4/16, 1/16).

In diesem Punkt 4.3.2 geht hier um die Untersuchung der Abhängigkeit auf der Basis von p^T
– *Korrelation* bedeutet, dass ein relativ niedriger Wert von p^T vorliegt, z. B. bei der deterministischen Relation $p(X * \rightarrow Y) = 1$.
– *Keine Korrelation* bedeutet, dass ein relativ hoher Wert von p^T vorliegt, bei einer sogenannte *Zufallsverteilung* oder einem *singulären Zufallswert (Einzelwert)*. Auf diese *Zufallsergebnisse* konzentrieren wir uns zunächst. Wir können hier unterscheiden:

• *Einzelwert*
Wir machen z. B. eine Untersuchung von 4 Sportlern und stellen fest: 2 von den 4 Sportlern sind gesund. Es gilt also: p(Sportler *→ gesund) = 2/4 bzw. p(Sportler *→ gesund) = 0,5, wenn wir den Wert 2/4 dezimal ausdrücken.

Allgemein gilt $p(X * \rightarrow Y) = 0{,}5$, also 50% aller X sind auch Y.
$p(X * \rightarrow Y) = 0{,}5$ ist der *Zufallswert,* der Wert mit der *höchsten theoretischen Wahrscheinlichkeit* p^T, im Beispiel $p(X * \rightarrow Y) = 2/4$, mit der $p^T = 6/16$. Man kann den Zufallswert auch die *‚Erwartungswahrscheinlichkeit'* nennen.

Wenn der Zufallswert p = 0,5 gegeben ist, liegt *keine Abhängigkeit* von Y zu X vor (zur vollständigen Bestimmung von Korrelation vgl. oben).

• *Verteilung*
Eine Verteilung erfasst nicht nur einen *Einzelwert* (z. B. 2/4), sondern *alle möglichen* Werte (im Beispiel 0/4, 1/4, 2/4, 3/4, 4/4). Zwar schließen sich die Einzelwerte ja gegenseitig aus, aber es mögen sich bei *verschieden Untersuchungen* verschiedene Einzelwerte ergeben: zu unterschiedlichen *Zeiten*, an unterschiedlichen *Orten*, unter unterschiedlichen *Bedingungen*. (Das bedeutet allerdings, dass der Einzelwert auch nicht verabsolutiert werden darf.)

Angenommen, wir machen 16 Untersuchungen (bzw. Stichproben) von jeweils 4 X bzw. 4 Sportlern – und erhalten das folgende Ergebnis:

1mal: 4 von 4 X sind auch Y
4mal: 3 von 4 X sind auch Y
6mal: 2 von 4 X sind auch Y
4mal: 1 von 4 X ist auch Y
1mal: 0 von 4 X ist auch Y

Wenn man diese Fälle alle addiert, ergeben sich 64 X.

Von den 64 X sind 32 auch Y, also wiederum 50%.

$p(X^* \rightarrow Y)$	p dez.	p^T	p^T dez.	p^T %
4/4	1,0	1/16	0,06	6%
3/4	0,75	4/16	0,25	25%
2/4	0,5	6/16	0,38	38%
1/4	0,25	4/16	0,25	25%
0/4	0,0	1/16	0,06	6%

Wenn alle Werte in der genannten Häufigkeit auftreten, liegt *keine Korrelation* vor. Wir können dagegen von einer *empirischen zufälligen* Verteilung sprechen, das ist eine Verteilung, die genau der Zufallserwartung bzw. einer *Zufallsverteilung* entspricht. Bei der Zufallsverteilung handelt es sich um eine theoretische, analytische Verteilung, die unabhängig von der Empirie ist.

Man kann zum einen die *theoretische Verteilung* 1/16, 4/16, 6/16, 4/16, 1/16 ‚Zufallsverteilung' nennen, man kann aber auch die konkrete *empirische Verteilung* ‚zufällig' nennen, weil sie nämlich exakt der theoretischen Verteilung entspricht. Demnach liegt keine Korrelation zwischen X und Y vor.

Somit kann eine *Einzeluntersuchung* durchaus zu falschen Ergebnissen führen, wenn wir z. B. den Wert p = 4/4 erhalten und den dann unzulässigerweise verallgemeinern.

4.3.3 Zufalls-Beziehung

Eine *Zufalls-Beziehung* liegt wie beschrieben typischerweise beim *Glücksspiel* vor. Z. B. kann man sagen, beim *Roulette* ist es zufällig, ob bei einem Wurf (bzw. Lauf) der *Kugel* die Farbe „rot" oder „schwarz" kommt, da es gleich viel, nämlich jeweils 18 Felder mit *rot* und mit *schwarz* gibt; von dem Feld 0 sehe ich wie gesagt ab, es hat auch keinen Einfluss auf das Verhältnis von *rot* und *schwarz*.

Wir können folgende Faktoren bei einem Zufalls-System wie dem Roulette unterscheiden:

• *theoretische Wahrscheinlichkeit* p^T

Die gibt die *Chance* (bzw. die Erwartung) an, dass *rot* oder *schwarz* kommt.
Sie beträgt jeweils 18/36, also 1/2 bzw. dezimal 0,5 oder 50 %. p^T[rot] = p^T[schwarz] = 0,5.
Damit ist die theoretische Wahrscheinlichkeit definiert.

• *meta-theoretische Wahrscheinlichkeit* p^{TT}

Die theoretische Wahrscheinlichkeit p^T gibt an, dass es am wahrscheinlichsten ist, dass in einer Verteilung zu 50% rot und zu 50% schwarz auftreten. Aber *wie wahrscheinlich* genau ist das? Dafür kann man eine erneute Wahrscheinlichkeit berechnen. Dabei wird p^T[rot] = 0,5 bzw. p^T[schwarz] = 0,5 quasi wie die *empirische Wahrscheinlichkeit* behandelt, der dann eine theoretische Wahrscheinlichkeit zugewiesen wird, im Grunde eine *Meta-meta-Wahrscheinlichkeit* p^{TT}. Man kann aber einfacher die theoretische Wahrscheinlichkeit hier konkret als relative Häufigkeit p ansehen, so dass man mit *einer* theoretischen Wahrscheinlichkeit p^T auskommt. Die hängt aber davon ab, wie groß n = Anzahl der Kugelwürfe ist. Bei n = 4 ist z. B. die Wahrscheinlichkeit p^T[p(rot) = 2/4 = 0,5] = 6/16 = 0,38. Bei n = 6 gilt dagegen für p = 0,5: p^T[p(rot) = 3/6 = 0,5] = 20/64 = 0,31.

Es lässt sich aber auch eine vollständige *Verteilung* angeben, bei der nicht nur p = 0,5, sondern allen möglichen Werten eine (meta-)theoretische Wahrscheinlichkeit zugeordnet wird.

Bei n = 4 ergeben sich folgende Werte (vgl. Tabelle oben):

p(rot)	p^T
4/4	1/16
3/4	4/16
2/4	6/16
1/4	4/16
0/4	1/16

Man kann das auch so deuten, dass hier 16 *Spiele* zu je 4 *Würfen* gespielt werden. Dabei kommt 1mal eine Verteilung mit 4 rot (r r r r) vor, 4mal eine Verteilung mit 3 rot usw. Multipliziert und addiert man das, dann ergibt sich:

$$
\begin{aligned}
4/4 \times 1/16 &= 4/64 \\
3/4 \times 4/16 &= 12/64 \\
2/4 \times 6/16 &= 12/64 \\
1/4 \times 4/16 &= 4/64 \\
0/4 \times 1/16 &= \underline{0/64} \\
&\quad 32/64 = 1/2 = 0{,}5
\end{aligned}
$$

Man könnte das mit der *Meta-Meta-Wahrscheinlichkeit* fortsetzen, z. B.
$p^T[p = 1/2] = 2/4$, $p^{TT}[p^T = 2/4] = 6/16$, $p^{TTT}[p^{TT} = 6/16] =$ …. usw.
Aber dies ist nicht sinnvoll, es führte zu einem *infiniten Regress*.

• *empirische relative Häufigkeit* p
Auch bei einer *Zufalls-Beziehung* wie beim Roulette kann man eine *empirische relative Häufigkeit* beschreiben bzw. untersuchen. Man kann sich z. B. an den Roulette-Tisch stellen und zählen, wie oft an einem Abend *rot* und *schwarz* kommen. Die Anzahl der untersuchten Fälle kann unterschiedlich sein, ist aber immer endlich. Z. B. mag sich ergeben, dass in 100 Fällen 6omal *rot* und 40mal *schwarz* kommt, die relative Häufigkeit für *rot* also p = 60/100 war. Man erwartet, dass die *empirische Wahrscheinlichkeit* in etwa der *theoretischen Zufallsverteilung* entspricht, weil es sich hier eben um ein *Glückspiel* handelt. Aber die Realität muss sich nicht an die theoretische Wahrscheinlichkeit halten, wie wir noch diskutieren werden.

• empirische allgemeine Wahrscheinlichkeit p
Im obigen Fall hat man an *einem* Abend mitgezählt. Man kann aber auch für viele Spiele über *lange Zeiten* Beobachtungen machen und die dann in einem *empirischen Gesetz* formulieren. Beim „Gesetz der großen Zahl" ist es umstritten, ob es empirisch oder doch analytisch ist. Dagegen gilt das „Zwei-Drittel-Gesetz" als empirisch. Es besagt, dass bei *einem* Durchgang im Roulette (36 bzw. 37 Würfe) etwa 2/3 aller Zahlen vorkommen.

4.4 AUSGLEICH

Ich habe die *Gleichverteilung* bzw. *Gleichwahrscheinlichkeit* erläutert. Wir haben gezeigt, die Wahrscheinlichkeit für *rot* und *schwarz* beträgt beim Roulette 50: 50. Da *rot* und *schwarz* also die gleiche Chance haben, vermutet man, dass es *real* zu einem *Ausgleich* kommen muss, d. h. dass bei einem Roulette-Spiel (ungefähr) zu 50% *rot* und zu 50% *schwarz* auftreten muss.

Ausgleich ist gewissermaßen *Gleichverteilung in der Zeit*. Aber was heißt das konkret? Wie stellt sich der Ausgleich dar?

4.4.1 Ausgleich auf kurze Sicht

- Ausgleich in 1 Wurf

Die Meta-Wahrscheinlichkeit dafür: $p^T = 0$. Denn bei *einem* Wurf kann ja nur entweder *rot*, also p(rot) = 1 oder schwarz, d. h. p(rot) = 0 kommen, es kann keinen Ausgleich geben.

- Ausgleich in 2 Würfen

Spiel	W_1	W_2	p^T
1.	r	r	1/4
2.	r	s	1/4
3.	s	r	1/4
4.	s	s	1/4

Hier bedeutet Ausgleich p = 1/2 = 0,5. Bei 2 Würfen sind 4 verschiedene Spiele möglich. Beim 2. und 3. Spiel kommt 1mal r *und* 1mal s vor. Hier liegt also ein *Ausgleich* vor. Man muss die *theoretischen* Wahrscheinlichkeiten dieser beiden Folgen addieren: 1/4 + 1/4 = 2/4. Es gibt also bei 2 Würfen eine Wahrscheinlichkeit p^T = 2/4 = 1/2 = 50%, dass zu 50% *rot* und *schwarz* auftreten. Aber eben auch zu 50%, dass dies nicht der Fall ist. Man keinesfalls sagen, dass hier mit Sicherheit ein Ausgleich, also eine Kette ‚r s' oder ‚s r' auftritt. Ein Ausgleich ist hier genauso wahrscheinlich wie kein Ausgleich. Also p^T[p(rot) = p(schwarz) = 1/2] = 0,5.

Aber wie käme ggf. der Ausgleich? Hier kann man die *Zufallsverteilung* nennen. Am wahrscheinlichsten wäre, bei 4 Spielen zu je 2 Würfen kämen die nachfolgenden 2er-Ketten je einmal vor: r r / r s / s r / s s. Das wären also zusammen 8 Würfe, mit insgesamt 4 r und 4 s Also hier ergäbe sich wiederum das Verhältnis 50:50 für r : s.

4.4.2 Ausgleich in einer ganzen Kette

Es geht jetzt um den Ausgleich *auf lange Sicht*. Der Ausgleich bei 2 Würfen ist natürlich nicht wirklich relevant, dies sollte nur das Problem am einfachen Beispiel erläutern.

Gemäß dem „Gesetz der großen Zahl“ ist *auf lange Sicht*, bei einer Vielzahl von Würfen ein Ausgleich zu erwarten. Allerdings muss dabei nicht ein *absoluter* Ausgleich stattfinden, es geht vorrangig um einen *prozentualen* Ausgleich: Je weiter eine Chance, eine Farbe zurückliegt, desto höher ist ihre prozentuale Zunahme, wenn sie dann auftritt.

Hier ist grundsätzlich zweierlei zu unterscheiden:

– man betrachtet eine Wurf-Kette *als ganze*

– man setzt *innerhalb einer Kette* von Würfen ein

Hier geht es zunächst um die Analyse einer *Kette von Würfen als ganze*.

Untersuchen wir eine Kette von 14 Würfen. (Das ist zwar auch noch keine wirklich große Zahl, aber sonst werden die anzugebenden Wahrscheinlichkeiten zu gering.)

Die folgende Tabelle gibt an, wie wahrscheinlich ein Ausgleich von rot und schwarz ist, bei einer Folge (oder Kette) von 2 bis zu 14 Würfen.

Nr.	Würfe	p(rot)	p^T	p^T dez.
1.	2.	1/2	2/4	0,5
2.	4.	2/4	6/16	0,38
3.	6.	3/6	20/64	0,31
4.	8.	4/8	70/256	0,27
5.	10.	5/10	252/1024	0,25
6.	12.	6/12	924/4096	0,23
7.	14.	7/14	3432/16384	0,21

Für den Ausgleich bei n = 14 gilt also: p^T[p(rot) = p(schwarz) = 7/14] = 0,21.

Die Zählung 2. 4. 6. usw. erklärt sich wie folgt: Ich zähle nur die *geraden* Würfe auf, bei denen möglich ist, dass genau 50% *rot* und *schwarz* auftreten. Bei 5 Würfen z. B. ist das nicht möglich, wenn p(rot) = 3, kann p(schwarz) nur 2 sein und umgekehrt.

Hier zeigt sich: mit zunehmendem n (Anzahl der Würfe) nimmt die Wahrscheinlich eines *genauen* Ausgleichs 50:50 gerade *ab*, anders als vielleicht erwartet und oft behauptet. Bei dem 14. Wurf der Kugel besteht also nur eine Wahrscheinlichkeit von $p^T = 0{,}21$, dass *rot* und *schwarz* genau übereinstimmen.

Zwar nehmen die *absoluten* Zahlen zu: 2, 6, 20, 70 usw., aber die entscheidenden *relativen* Zahlen sinken, die Wahrscheinlichkeit nimmt ab: von 0,5 über 0,38 usw. bis zu 0,21.

Wenn man allerdings nicht einen *genauen* Ausgleich 50:50 berücksichtigt, sondern nur einen *ungefähren* Ausgleich, dann nimmt die Wahrscheinlichkeit eines Ausgleichs zu: z. B. wenn man nur berücksichtigt, wie wahrscheinlich ist, dass *nicht immer* rot (oder nicht immer schwarz) kommt. Das will ich nun für den folgenden Fall eines *begrenzten Ausgleichs* zeigen.

Wir unterscheiden dabei nur:

kein Ausgleich: es kommt *immer* „rot“, nie „schwarz“, also p(rot) = 1 und p(schwarz) = 0

Ausgleich: das ist falsch, also p(rot) ≠ 1 bzw. p(rot) < 1, damit p(schwarz) > 0

Dies ist allerdings ein sehr bescheidener Quasi-Ausgleich.

Wurf	p(rot) < 1	p^T	p^T dez.
2.	< 2/2	3/4	0,75
4.	< 4/4	15/16	0,94
6.	< 6/6	63/64	0,98
8.	< 8/8	255/256	≈1,0
10.	< 10/10	1023/1024	≈1,0
12.	< 12/12	4093/4096	≈1,0
14.	< 14/14	16383/16384	≈1,0

Hier steigt die Wahrscheinlichkeit eines „Ausgleichs“ bis fast auf $p^T = 1$, und zwar schnell. Das erklärt sich folgendermaßen: mit steigendem n nimmt die Wahrscheinlichkeit p^T ab, dass nie *schwarz*, also immer *rot* kommt, quantitativ dass p(rot) = 1. Dabei ergänzt sich:
$p^T[p(rot) = 1] + p^T[p(rot) < 1] = 1$

Wurf	p(schwarz) = 0	p(rot) = 1	p^T	p^T dez.
2.	0/2	2/2	1/4	0,25
4.	0/4	4/4	1/16	0,26
6.	0/6	6/6	1/64	0,02
8.	0/8	8/8	1/256	≈ 0,0
10.	0/10	10/10	1/1024	≈ 0,0
12.	0/12	12/12	1/4096	≈ 0,0
14.	0/14	14/14	1/16384	≈ 0,0

Wir können den Ausgleich auch noch etwas *erweitern*:
Ausgleich bedeutet hier nicht nur:

p(rot) < 1, sondern auch p(rot) > 0.

Somit ebenfalls: p(schwarz) < 1, sondern auch p(schwarz) > 0.
Ausgleich heißt dann eben: *rot* kommt *nicht immer* und *nicht nie* vor und entsprechend auch *schwarz*.

Man könnte meinen, das müsste (bei n = 2) folgendermaßen berechnet werden:

$p^T[p(rot) < 1]$ x $p^T[p(rot) > 0]$, z. B. 3/4 x 3/4 = 9/16

Aber das wäre ein Irrtum, weil p(rot) < 1 und p(rot) > 0 *nicht unabhängig* voneinander sind.

Stattdessen berechnet man p^T[p(rot) < 1] und *substrahiert* davon den Fall p^T[p(rot) = 0], z. B. 3/4 - 1/4 = 2/4

Wurf	p(rot) < 1	p^T	p(rot) = 0	p^T	$p^T - p^T$	dez.
2.	< 2/2	3/4	0/2	1/4	2/4	0,5
4.	< 4/4	15/16	0/4	1/16	14/16	0,88
6.	< 6/6	63/64	0/6	1/64	62/64	0,97
8.	< 8/8	255/256	0/8	1/256	254/256	0,99
10.	< 10/10	1023/1024	0/10	1/1024	1022/1024	≈1,0
12.	< 12/12	4095/4096	0/12	1/4096	4094/4096	≈1,0
14.	< 14/14	16383/16384	0/14	1/16384	16382/16384	≈1,0

Auch hier geht die Wahrscheinlichkeit eines *Ausgleichs* gegen 1, obwohl nicht ganz so schnell wie bei der Tabelle oben mit nur p(rot) < 1. Allerdings darf man das ≈ 1 nicht falsch interpretieren: Es wird eben kein $p^T = 1$, *keine Sicherheit* erreicht, der Ausgleich muss sich real nicht einstellen.

4.4.3 Analyse innerhalb einer Kette von Würfen

Eben haben wir untersucht: Wie wahrscheinlich ist ein *Ausgleich* zwischen „rot“ und „schwarz“, wenn wir eine Kette von Würfen *als ganze* betrachten?

Jetzt geht es um die Frage: Wenn innerhalb einer Kette von Würfen schon eine Anzahl von Würfen gefallen ist, wie wahrscheinlich ist dann ein Ausgleich?

Betrachten wir wieder unsere Kette von 14 Würfen. Angenommen wir sind auf Position 7, und bisher kam nur *rot* vor: r r r r r r r

Wie wahrscheinlich ist dann, dass noch ein Ausgleich mit schwarz stattfindet? Wenn es ein vollständiger Ausgleich sein soll, müsste nun also 7mal *schwarz* kommen, so dass wir insgesamt folgende Kette bekämen:

r r r r r r r s s s s s s s

Hier ist ganz wichtig: *Jeder Wurf zählt wieder neu*. Die Wahrscheinlichkeit eines Wurfes ist unabhängig von den Würfen, die vorher gekommen sind. Wie man auch sagt: „Die Kugel hat kein Gedächtnis“. Auch wenn 100mal rot gekommen ist, ist es genau gleich wahrscheinlich, dass nochmals *rot* kommt, wie dass jetzt *schwarz* kommt. Viele Spieler denken anders, sie meinen, der Ausgleich wäre jetzt überfällig, aber das ist ein Irrtum.

Wir brauchen somit die vorausgegangene Folge von 7mal *rot* gar nicht zu berücksichtigen. Wir müssen jetzt nur die Wahrscheinlichkeit berechnen, dass 7mal *schwarz* kommt, also dass p(schwarz) = 7/7 = 1.

Dabei erhalten wir insgesamt folgende Zahlen:

p(schwarz)	p^T
7/7	1/128
6/7	7/128
5/7	21/128
4/7	35/128
3/7	35/128
2/7	21/128
1/7	7/128
0/7	1/128
	128/128

D. h., wenn schon 7 mal rot gekommen ist, ist die Wahrscheinlichkeit, dass bei insgesamt 14 Würfen ein exakter *50:50-Ausgleich* erzielt wird, dass also jetzt 7mal *schwarz* kommt, 1/128 = 0.0078, gerundet auf 2 Stellen 0,01.

Vergleichen wir dies mit der Wahrscheinlichkeit der *ganzen* Kette: Dort ergibt sich bei n = 14 eine Wahrscheinlichkeit $p^T = 0{,}21$ für einen rot-schwarz-Ausgleich, also wesentlich höher.

Natürlich hängt die Ausgleichswahrscheinlichkeit *in* einer Kette auch davon ab, an welchem Punkt der Kette man ansetzt und wie groß der bisherige Ausgleich ist. Angenommen, es waren von den bisherigen 7 Würfen nicht 7, sondern nur 5 rot. Ein *vollständiger* Ausgleich wäre also dann gegeben, wenn in den nächsten 7 Würfen nur noch 2mal rot aufträte. Die Wahrscheinlichkeit kann man der obigen Tabelle entnehmen: $p^T[p(\text{rot}) = 2/7] = 21/128 = 0{,}16$; dieser Wert liegt natürlich deutlich über den 0,01. Wenn man nicht an der Position 7 in der Kette ansetzt, sondern z. B. an der Position 3, ergeben sich wieder andere Verhältnisse.

4.4.4 Ausgleich in der Unendlichkeit

Wir untersuchen hier, inwieweit sich beim Roulette zwischen zwei gleichen Chancen rot und schwarz (*2-wertige Zufallsvariable*) ein Ausgleich einstellen muss. Wir haben bisher 1, 2 und 14 Würfe untersucht, aber wie sieht es mit *unendlich* vielen Würfen aus?

Nun gibt es Probleme, einen Quotient p mit unendlichen Mengen zu bilden (vgl.1-3-0-4). Stattdessen wird behauptet: Der *Grenzwert* der relativen Häufigkeit ist hier 0,5, wenn n gegen unendlich geht. Bei *unendlichen* Würfen ergibt sich sicher ($p^T = 1$) ein *genauer* Ausgleich.

Für *Zahlenfolgen* lässt sich ein Grenzwert berechnen, z. B. $\lim_{n \to \infty} n/(2n + 1) = 0{,}5$

1/3	2/5	3/7	4/9	5/11	6/13	7/15	8/17	9/19	10/21	
0,33	0,4	0,43	0,44	0,45	0,46	0,47	0,47	0,47	0,48	(gerundet)

Hier nähert sich die Folge *sicher* und *stetig* dem Grenzwert 0,5. Beim Roulette haben wir dagegen zunächst eine (empirische) Folge von *Ereignissen*/Würfen; denen können wir zwar Zahlen, p-Werte zuordnen. Aber die bewegen sich keineswegs sicher und linear auf 0,5 zu.

Wie aufgezeigt, nimmt die theoretische Wahrscheinlichkeit p^T für einen *vollständigen* Ausgleich p(rot) = p(schwarz) mit steigendem n sogar ab. Und es ist zwar extrem unwahrscheinlich, aber nicht unmöglich, dass in einer *unendlichen* Folge von Würfen z. B. die schwarze Farbe nur *endlich* oft oder sogar *kein einziges Mal* auftritt, sondern nur rot.

Höchstens *indirekt* könnte man von einem absoluten Ausgleich in der Unendlichkeit sprechen. Denn *infinite abzählbare* Mengen besitzen die gleiche *Mächtigkeit*. Wenn also „rot" und „schwarz" *unendlich* oft geworfen wurden, dann gelten diese zwei Mengen in jedem Fall als *gleich groß*, auch wenn z. B. rot viel öfter geworfen wurde: $\infty + a = \infty$

4.4.5 Ausgleich und Reihenfolge

Auch wenn wir schon von *Ketten* von Würfen gesprochen haben, wir haben die Würfe bisher überwiegend als *Mengen / Teilmengen* betrachtet, z. B. 4 Würfe (Teilmenge) von 8 Würfen (Menge). Aber in einem *konkreten* Spiel treten die Würfe als *geordnete* Menge auf, d. h. in einer *Reihenfolge*. Mathematisch spricht man hier von einer *Folge* (nicht von einer *Reihe*); zwar meint man i. allg. *Zahlen*-Folgen, aber wir können auch *Ereignis*-Folgen verwenden.

Nur bei p = 1 ist die Reihenfolge klar, hier stimmt die Menge mit der geordneten Menge überein: z. B. 4mal *rot* kann eben nur in *einer W*eise geordnet sein. Das Problem der Reihenfolge stellt sich auch in anderer Weise: Es geht darum, ob *konkrete* Individuen erfasst werden oder nur eine *Menge* von Individuen, wobei man nicht genau sagen kann, *welche* Individuen dazugehören, nur *wie viele* Individuen.

Um die Bedeutung der Reihenfolge klarzumachen, betrachte man zwei Folgen von 14 Kugelwürfen: (r = rot, s = schwarz):

r r r r r r r r r r r r r r Folge mit 14mal rot
r s r s s r s r r r s r s s Folge mit 7mal rot, 7mal schwarz

erste Folge: relative Häufigkeit p(rot) = 14/14, p(schwarz) = 0/14
zweite Folge: relative Häufigkeit p(rot) = 7/14, p(schwarz) = 7/14

Welche Folge ist theoretisch wahrscheinlicher? Wer sich nicht genauer auskennt, wird sagen, die zweite Folge. Dies ist aber nicht der Fall. Beide Folgen haben genau die gleiche theoretische Wahrscheinlichkeit, nämlich $p^T = 1/2^{14} = 1/16384$.

Man betrachte zur Erläuterung folgendes Diagramm mit nur 4 Würfen (zur Vereinfachung):
Es beinhaltet 16 Spiele zu je 4 Würfen, wobei jedes dieser Spiele für sich alleine betrachtet die gleiche theoretische Wahrscheinlichkeit $p^T = 1/16$ besitzt.

SPIEL	W_1	W_2	W_3	W_4
1.	+	+	+	+
2.	+	+	+	–
3.	+	+	–	+
4.	+	+	–	–
5.	+	–	+	+
6.	+	–	+	–
7.	+	–	–	+
8.	+	–	–	–
9.	–	+	+	+
10.	–	+	+	–
11.	–	+	–	+
12.	–	+	–	–
13.	–	–	+	+
14.	–	–	+	–
15.	–	–	–	+
16.	–	–	–	–

Wie können diese 16 Spiele (Möglichkeiten) aber in folgender Weise ordnen:

Objekt-Wahrscheinlichkeit: p		*Meta-Wahrscheinlichkeit* p^T
0 von 4 WÜRFEN sind rot:	0/4	1/16
1 von 4 WÜRFEN sind rot:	1/4	4/16
2 von 4 WÜRFEN sind rot:	2/4	6/16
3 von 4 WÜRFEN sind rot:	3/4	4/16
4 von 4 WÜRFEN sind rot:	4/4	1/16
		16/16 = 1

Man fasst jetzt also Würfe mit *gleichviel rot* zusammen. Und da zeigt es sich: Dass 2 von 4 Würfen *rot* sind, kommt theoretisch am häufigsten vor, nämlich $p^T = 6/16$. Dass 4 von 4 Würfen *rot* sind, kommt dagegen nur 1mal vor, daher $p^T = 1/16$. Der Fehler, den man begeht, ist, dass man z. B. die *konkrete* Folge r r r r mit der *Zusammenfassung* von 6 Folgen vergleicht, in denen 2mal *r* vorkommt, also den Folgen: r r s s / r s r s / r s s r / s s r r / s r s r / s r r s.

Dann erscheint natürlich die Folge mit 2mal *rot* viel wahrscheinlicher als die mit 4mal *rot*; vergleicht man aber zwei *konkrete Folgen*, z. B. r r r r, mit r s r s, dann sind beide Folgen *gleichwahrscheinlich*.

Hier liegt eine Art (optischer) *Täuschung* vor: 4mal *rot*: r r r r fällt uns auf; dagegen ist eine Folge wie r s r s weniger auffällig und leicht mit z. B. r s s r zu verwechseln. Daher glaubt man, r r r r kommt selten vor, z. B. r s s r viel häufiger, aber man denkt eben nicht daran, dass hier 6 Reihen mit 2mal r und 2mal s zusammengefasst werden.

Bei den *konkreten* Folgen spielt die *Reihenfolge* eine Rolle, bei den *Zusammenfassungen* aber nicht, hier wird von der Reihenfolge abstrahiert. Wie wichtig die Reihenfolge ist, zeigt sich z. B. am Roulette-Spiel. Es nützt dem Spieler gar nichts, wenn ich ihm sage, dass es 6x wahrscheinlicher ist, dass (in 16 Spielen) 2mal *rot* und 2mal *schwarz* auftritt als 4mal *rot*. Denn um zu gewinnen, muss er wissen, in welcher Reihenfolge die *Rots* und *Schwarzs* kommen. Dass ein Ausgleich stattfindet, ist zwar wahrscheinlich, aber man kann nicht sagen, *wie*.

Wenn man das zu Ende denkt, kommt man zu einem erstaunlichen Ergebnis: Dass *immer* rot kommt, und zwar *unendlich* oft, ist genauso wahrscheinlich bzw. besser genauso unwahrscheinlich wie *jede andere konkrete* Folge.

Um nur eine ganz schwache Vorstellung von einer unendlichen Folge zu geben:

r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r r

r r s r s s r s r s r r r s s r s r s s s r r s s r s r r r r s s r s s r r s s s s r s r r s s s s r r s r s s r s s s

Beide Folgen umfassen 60 Elemente, in der oberen kommt nur r (rot) vor, in der unteren 30mal rot und 30mal schwarz, also genau die *Zufalls-Erwartung*.

Die obere Folge mit ausschließlich *rot* ist aber genau so wahrscheinlich bzw. genauso unwahrscheinlich wie die untere bzw. wie jede wie jede andere *konkrete* Folge. Wenn eine solche Folge wie die obere r-Folge auftritt (von einer unendlichen gar nicht zu reden), sind wir höchst verwundert und denken: was für ein unglaublicher Zufall! Wenn dagegen die untere Folge auftritt, verwundert sich kein Mensch darüber. Aber man könnte hier genauso sagen: Was für ein *Zufall*, dass gerade diese Folge auftritt!

Wenn z. B. 1000mal die Kugel nacheinander auf einem roten Feld landet, wäre das natürlich eine hochgradige Korrelation zwischen dem Kugelwurf und dem *rot*. Obwohl ein Zufallsprozess vorliegt, kann sich eben – zufällig – doch eine Korrelation einstellen. Aber (wenn wir von einem regulären Glücksspiel ausgehen), ist diese Korrelation kausal gesehen *zufällig*, es besteht kein K*ausalzusammenhang*.

Man muss ich eins klarmachen: *Wahrscheinlichkeit bedeutet eben nicht Sicherheit*. Und so gesehen kann eben auch das Unwahrscheinlichste eintreten. Nein, eigentlich tritt so gar *notwendig* das Unwahrscheinlichste auf, weil *alle* Folgen maximal unwahrscheinlich sind und zwar *gleich* unwahrscheinlich sind. (vgl. 4.5). Bei allen Folgen geht p^T gegen 0, aber *eine* Folge ist ja *real* gegeben, also kann jedenfalls für sie gar nicht gelten, dass $p^T = 0$.

4.5 WAHRSCHEINLICHKEIT UND WIRKLICHKEIT

Ich möchte hier über unsere *reale* Welt schreiben, was natürlich aus der Logik im engeren Sinn hinausführt.

4.5.1 Unsere Welt beinhaltet Zufall und Korrelation (Ordnung)

Diese Aussage bezieht sich zunächst auf den *statistischen Zufall*. Es stehen *nicht alle* Ereignisse / Variablen in Korrelation miteinander. Aber es sind auch *nicht alle* Ereignisse in zufälliger Beziehung. Somit stehen *manche* Ereignisse in Korrelation und andere nicht.

Eine rein *zufällige* Welt könnte stark vereinfacht folgendermaßen aussehen:

X 100%

Y 50% ¬Y 50% 50%

Z 50% ¬Z 50% Z 50% ¬Z 50% 25%

Eine ganz geordnete, *korrelative* Welt könnte vereinfacht folgendermaßen aussehen:

X 100%

Y 100% ¬Y 0% 100%

Z 100% ¬Z 0% Z 0% ¬Z 0% 100%

Eine *zufällig und korrelativ gemischte* Welt könnte z. B. folgendermaßen aussehen:

X

Y 75% ¬Y 25%

Z 50% (37,5%) ¬Z 50% (37,5%) Z 60% (15%) ¬Z 40% (10%)

Noch deutlicher wird die rein zufällige Welt vielleicht in folgender Darstellung:

X	50%	Y
50%	50%	50%
V	50%	W

Das ist zu lesen: 50% aller X sind Y, 50% aller Y sind X.

50% aller X sind V, 50% aller V sind X usw. usw. Alle Variablen stehen zueinander im Verhältnis 50%.

Für die anderen Welt-Modelle gelten entsprechende Diagramme.

Natürlich kann man nicht für jedes Ereignis in der Welt prüfen, ob es mit irgendeinem anderen korreliert. Zwar korreliert sicher nicht jedes Ereignis mit *jedem* anderen. Aber es ist schon denkbar, dass es für jedes Ereignis mindestens *ein* anderes gibt, das mit ihm korreliert. Ja, es ist darüber hinaus denkbar, dass es für jedes Ereignis eine *kausale Ursache* gibt.

4.5.2 Unsere Welt ist theoretisch unwahrscheinlich

Wir haben uns oben mit der *empirischen* Wahrscheinlichkeit p beschäftigt. Und bei ihr gibt es in der realen Welt offensichtlich alle Werte: von 100% über 50% bis zu 0%.

Ganz anders sieht es mit der theoretischen Wahrscheinlichkeit p^T aus. Betrachten wir zur Verdeutlichung eine *Kette von singulären Ereignissen.*

Folge n	p^T	mögliche Ereignisse	reales Ereignis	p^T
1	1/2	1	1	1/2
2	3/4	3	1	1/4
3	7/8	7	1	1/8
4	15/16	15	1	1/16
5	31/32	31	1	1/32
6	63/64	63	1	1/64

Was zeigt diese Tabelle?

- Erstes Ereignis (1): Hier gibt es genau 2 Möglichkeiten: das Ereignis X ist real (und das Ereignis $\neg X$ ist möglich) oder $\neg X$ ist real (und X ist möglich). X und $\neg X$ haben beide eine Erwartungswahrscheinlichkeit von $p^T = 1/2$. Nehmen wir an, X ist real.
- Zweites Ereignis (2): Hier gilt, real kann Y eintreten oder nicht. Angenommen, Y tritt ein. Dann haben wir real bisher $X \wedge Y$, also 1 reale *Kombination*. Möglich sind aber bisher schon folgende 3 Kombinationen gewesen: $X \wedge \neg Y$, $\neg X \wedge Y$, $\neg X \wedge \neg Y$. Somit hat die reale Kombination den Wert $p^T = 1/4$, die möglichen (irrealen) Ereignis-Kombinationen aber zusammen den Wert $p^T = 3/4$, nämlich 1/4 + 1/4 + 1/4.
- Sechstes Ereignis (6): Bis hierhin hat sich schon ergeben, das die Folge singulärer Ereignisse nur noch eine theoretische Wahrscheinlichkeit von $p^T = 1/64 = 1{,}6$ % hat, also theoretisch *sehr unwahrscheinlich* ist (bzw. war). Dagegen haben die alternativen, möglichen Ereignisse den Wert $p^T = 63/64 = 98{,}4$ %.

Also nur wenn man ein *einzelnes reales* Ereignis betrachtet, hat es eine theoretische Wahrscheinlichkeit von $p^T = 1/2 = 50\%$. Es ist also *weder wahrscheinlich noch unwahrscheinlich*, sondern *neutral*. Aber schon *ab zwei* Ereignissen gibt es nur noch Unwahrscheinlichkeit. Da wir unsere *Welt* letztlich als Ketten (bzw. Systeme) extrem vieler, u. U. *unendlich* vieler Ereignisse begreifen können, ergeben sich also auch extrem hohe Grade von Unwahrscheinlichkeit. Ja, die theoretische Wahrscheinlichkeit, die *Erwartungswahrscheinlichkeit* unserer Welt tendiert gegen 0. Eigentlich dürfte es unsere Welt gar nicht geben.

Umgekehrt, der Bereich des *nur-Möglichen*, des Nicht-Realen erzielt immer höhere Werte von theoretischer Wahrscheinlichkeit, je mehr Ereignisse bzw. je längere Folgen wir berücksichtigen. Unsere reale Welt ist in Relation zu diesen (logisch) *möglichen Welten* extrem unwahrscheinlich. Eine andere, genau bestimmte Welt wäre allerdings auch nicht wahrscheinlicher als unsere reale.

Natürlich, wir können Ereignisse *disjunktiv zusammenfassen* (so wie dies im Grunde auch bei den möglichen Alternativ-Welten gemacht wird); dadurch erreichen wir hohe p^T -Werte bis hin zu $p^T = 1$, der Tautologie: z. B. $p^T[X \vee Y] = 0{,}75$ oder z. B. $p^T[X \vee \neg X] = 1{,}0$. Aber das ist eben nicht mehr die primäre Realitäts-Ebene.

Wir können diese Aussagen noch zuspitzen:
Auch das Wahrscheinlichste ist noch unwahrscheinlich.
Betrachten wir uns noch einmal die einfache Verteilung:

$p(X^* \rightarrow Y)$	p dez.	p^T	p^T dez.	p^T %
4/4	1,0	1/16	0,06	6%
3/4	0,75	4/16	0,25	25%
2/4	0,5	6/16	0,38	38%
1/4	0,25	4/16	0,25	25%
0/4	0,0	1/16	0,06	6%

Wie man sieht: der *empirische* Wert p mit der höchsten *theoretischen* Wahrscheinlichkeit p^T ist $p = 2/4$. Hier besteht eine theoretische Wahrscheinlichkeit $p^T = 6/16 = 38\%$. Das ist höher als alle anderen p^T-Werte, der nächst folgende Wert ist $p^T = 4/16 = 25\%$. Dennoch ist ein Wert von 38% unwahrscheinlich. Denn wir hatten ja definiert:

unwahrscheinlich: $p^T < 0{,}5$ bzw. $p^T < 50\%$

Und bei höherem *n* nimmt die *maximale Wahrscheinlichkeit* von p^T noch weiter ab.
Das veranschaulicht die folgende Tabelle:

n	p = 0,5	p^T	p^T dez.
2	1/2	2/4	0,5
4	2/4	6/16	0,38
6	3/6	20/64	0,31
8	4/8	70/256	0,27
10	5/10	252/1024	0,25

Wir erinnern uns, dass bei $p = 0{,}5$ immer der höchste Wert von p^T gegeben ist.
Bei $n = 10$ ist z. B. der höchste Wert nur noch 0,25, also 25%.

4.5.3 Veränderte Definition von Wahrscheinlichkeit ?

Diese Ergebnisse könnten zu der Frage führen, ob man „wahrscheinlich" neu und anders definieren sollte.

Ich hatte – wie üblich – bestimmt:

1. $p^T > 0{,}5$: wahrscheinlich
2. $p^T < 0{,}5$: unwahrscheinlich
3. $p^T = 0{,}5$: kontingent (neutral).

Man könnte dabei noch ergänzen:

4. $p^T = 1$: sicher
5. $p^T = 0$: sicher nicht

Wie gezeigt, sind bei dieser Definition fast alle (quantifizierten) Strukturen bzw. Verteilungen theoretisch *unwahrscheinlich*. Man kann dies als interessantes Faktum unserer Wirklichkeit ansehen. Man kann aber auch fragen, ob hier ein *zusätzlicher Wahrscheinlichkeitsbegriff* sinnvoll wäre, der sich auf eine *Gesamtverteilung* bezieht, der relativ zur Gesamtverteilung gilt. Das könnte dann z. B. folgendermaßen aussehen:

p(X*→ Y)	p dez.	p^T	p^T dez.	p^T %	
4/4	1,0	1/16	0,06	6%	unwahrscheinlich
3/4	0,75	4/16	0,25	25%	neutral
2/4	0,5	6/16	0,38	38%	wahrscheinlich
1/4	0,25	4/16	0,25	25%	neutral
0/4	0,0	1/16	0,06	6%	unwahrscheinlich

Allerdings gäbe es dann keine *absolut* gültigen Wahrscheinlichkeits-Werte mehr. Man könnte zwar festlegen, dass der *größte Wert* in einer Verteilung immer als *wahrscheinlich* gilt, der *kleinste* als *unwahrscheinlich*, aber die genaue Größe hinge von der Verteilung ab. Und wie die Werte zwischen dem höchsten und niedrigsten Wert einzustufen wären, bliebe auch noch offen.

Ich meine daher, dass man die *absolute* Definition von Wahrscheinlichkeit beibehalten sollte. *Wahrscheinlich* ist demnach nur ein Wert *größer 50%,* was eben bedeutet, dass unsere Welt – theoretisch betrachtet – sehr unwahrscheinlich ist.

5. KAUSAL-ANALYSE

Generell gilt für *synthetische* Aussagen / Relationen, gleichgültig ob mit *Gleichgewichts-Relatoren* oder mit *Ungleichgewichts-Relatoren*: Man will *Zusammenhänge* oder *Abhängigkeiten* zwischen Variablen bzw. zwischen Objekten oder Ereignissen feststellen.
Dabei sind zwei Schritte zu unterscheiden:

- *funktionale* Abhängigkeit bestimmen

Es wird geprüft, *ob* eine Verteilung vorliegt, die von der *Zufalls-Erwartung* (signifikant) abweicht: eine *funktionale Abhängigkeit* oder *Korrelation*.

- *ursächliche* Abhängigkeit bestimmen

In einem zweiten Schritt wird erläutert, erklärt oder begründet, *warum* die funktionale Abhängigkeit auftritt. Als Gründe kommen hier in erster Linie in Frage:

> Kausalität, Zielgerichtetheit, Identität, zeitliches Enthaltensein, räumliches Enthaltensein, logisches Enthaltensein

Der wichtigste Grund ist *Kausalität*, auf die ich mich im Folgenden auch fast ausschließlich beschränke. Dabei spielen *logische* und *nicht-logische* (bzw. *hyper-logische*) Faktoren eine Rolle.

5.1 KORRELATION UND KAUSALITÄT

5.1.1 Der Zufalls-Begriff

Ich habe den Begriff ‚Zufall' bisher nicht ausführlich differenziert. Der Zufalls-Begriff ist aber zumindest *doppeldeutig*; er kann bedeuten:

- keine Korrelation
- keine Kausalität (aber Korrelation)

Was unter *Korrelation* (Abhängigkeit) bzw. keiner Korrelation zu verstehen ist, wurde oben ausführlich erläutert.

Es geht jetzt darum, nur kurz einführend zu erläutern, wie *Kausalität* gegenüber dem *Zufallsbegriff* begrifflich zu bestimmen ist (bevor dies später genauer erläutert wird).

5.1.2 Keine Kausalität

Normalerweise denkt man: Korrelation = Kausalität, keine Korrelation (Zufallswert) = keine Kausalität.
Aber hier sind folgende Einschränkungen zu machen:

Erstens, anstatt Kausalität kann Korrelation auch durch *Identität*, *Teleologie* o.ä. (vgl. oben) bedingt sein, also irgendeine eine andere *über-korrelative* Abhängigkeit. Um es aber nicht zu kompliziert zu machen, will ich mich wie gesagt hier weitgehend auf *Kausalität* beschränken.

Zweitens, wenn sich eine Korrelation ergibt, *muss* die nicht auf einen Kausal-Zusammenhang hinweisen, sondern es *kann* dennoch nur eine Zufalls-Beziehung vorliegen. Eine Zufalls-Verteilung beinhaltet eben nur Wahrscheinlichkeiten, es kann – zufällige – Abweichungen geben, so dass sich eine „zufällige" Korrelation zeigt. Von *Störvariablen* und anderen speziellen statistischen Problemen sehe ich hier ab.

5.1.3 Beziehung von Kausalität und Korrelation

Es gilt: Kausalität $\Rightarrow$ Korrelation, keine Korrelation $\Rightarrow$ keine Kausalität.
Aber: Korrelation $\neg\Rightarrow$ Kausalität, d. h. aus dem Vorliegen einer Korrelation kann man nicht sicher auf Kausalität schließen.

Man sollte daher besser differenzieren zwischen:
statistisch zufällig (= nicht korrelativ) und *kausal zufällig* (= nicht ursächlich)
Dabei gilt: statistisch zufällig $\Rightarrow$ kausal zufällig, aber kausal zufällig $\neg\Rightarrow$ statistisch zufällig
In einer Tabelle zusammengefasst:

statistisch	Kausal	Ergebnis
zufällig	zufällig	reiner Zufall
zufällig	nicht zufällig	(nicht möglich)
nicht zufällig	zufällig	Korrelation
nicht zufällig	nicht zufällig	Kausalität

Entsprechend können wir drei *Stufen* unterscheiden:
– *statistisch zufällig:* keine Korrelation, keine Kausalität, also auch kausal zufällig
– *statistisch nicht zufällig:* Korrelation, aber keine Kausalität, d. h. kausal zufällig
– *nicht zufällig:* nicht statistisch und nicht kausal zufällig, Korrelation und Kausalität

Wenn X und Y statistisch zufällig *zueinander* stehen, heißt das allerdings nicht, dass Y prinzipiell zufällig ist. Es mag einen Faktor Z geben, der mit Y korreliert, der sogar kausale Ursache von Y sein kann.

5.2 KAUSAL-BEGRIFF

Unter *Kausalität* versteht man, dass eine *Ursache* eine *Wirkung* auslöst.

Genauer: Eine *Kausal-Beziehung* X ⇨Y bedeutet: X ist Ursache von Y.
Bzw.: X ist Ursache der Wirkung Y.

Wir müssen also unterscheiden zwischen einer (implikativen) *Wenn-dann-Relation* und einer *Kausal-Relation*.

- implikativ: $X \rightarrow Y$ „*wenn* X, *dann* Y“
  Z. B „Wenn jemand Sportler ist, dann ist (bzw. wird) er gesund“.
- kausal: X ⇨Y „*weil* X, *darum* Y“.
  Z. B „Weil jemand Sportler ist, darum ist (bzw. wird) er gesund“.

Unter einer *kausalen Erklärung* versteht man normalerweise folgenden *Schluss*:
$(X \rightarrow Y) \wedge X \Rightarrow Y$. Oder prädikaten-logisch: $\Lambda x(Fx \rightarrow Gx) \wedge Fx_i \Rightarrow Gx_i$

In anderer Form:

$\Lambda x(Fx \rightarrow Gx)$	Ursache-Wirkungs-Gesetz	alle Sportler sind gesund
Fx_i	singuläre Randbedingung	John ist Sportler
Gx_i	zu erklärendes singuläres Ereignis	John ist (wird) gesund

Genauer:
Für alle x (Menschen) gilt: wenn sie Sport treiben, werden sie gesund.
John treibt Sport.
John wird gesund.

Hier verwendet man also doch die *Implikation* $X \rightarrow Y$, allerdings muss an diese Implikation besondere Bedingungen geknüpft werden, die unten genannt werden.

So wäre z. B. folgende Ableitung *keine kausale Erklärung*: „Alle Münchner sind Deutsche, Franz ist Münchner, darum ist Franz Deutscher“. Denn „alle Münchner sind Deutsche“ ist kein Kausalgesetz.

Man könnte eventuell auch vertreten, anstatt der Implikation (wenn – dann) *direkt* ein *Kausalgesetz* (weil – darum) zu formulieren:

Λx(Fx ⇨ Gx): Für alle x gilt: *weil* sie die Eigenschaft F haben, *darum* haben sie auch die Eigenschaft G. (Aber eine solche Relation X ⇨ Y ist in der Logik nicht definiert.)

Doch normalerweise macht man die Kausalität in der *kausalen Erklärung* an der logischen *Ableitung* fest, z. B.:

Frage: *Warum* ist John gesund?
Antwort: *Weil* alle Sportler gesund sind und John ein Sportler ist.

Allerdings besteht hier die Gefahr von Missverständnissen. Denn natürlich wäre es falsch zu behaupten: „alle Sportler sind gesund und John ist ein Sportler“ ist die *Ursache* von „John ist gesund“ (als Wirkung). Die Kausalität im engeren Sinne ist eine reale Beziehung zwischen Ereignissen (Eigenschaften usw.), aber keine logische Beziehung zwischen Aussagen.
Auf die kausale Erklärung möchte ich hier aber nicht im Einzelnen eingehen.

Kausal-Relationen lassen sich prinzipiell auf zweierlei Weise feststellen:

– *Beobachtung* bzw. allgemein Wahrnehmung und Messung

Man beobachtet Phänomene der empirischen Welt, im Hinblick auf verschiedene Faktoren.
Man stellt z. B. fest, dass Rauchen zu Bronchitis führt und stellt die Kausal-Relation Rauchen ⇨ Bronchitis auf.

– *Experiment*

Bei Beobachtungen in der empirischen Welt besteht aber immer die Gefahr, dass man bestimmte Faktoren bzw. Variablen übersieht. Z. B. könnte es sein, dass Raucher sich häufiger verkühlen und daher Bronchitis bekommen. Im *Experiment* kann man Variablen kontrollieren und eine Vergleichsgruppe heranziehen.

Zur Bestimmung einer Kausal-Relation benötigt man *logische* (bzw. funktionale) und *außerlogische* Kriterien. Beide seien im nachfolgend aufgezählt.

5.3 KAUSALITÄT

• Determinismus: $p = 1$
Normalerweise versteht man unter Kausalität eine *deterministische* Beziehung, die also für *alle* X gilt: $p(X *\rightarrow Y) = 1$, z. B. p(Sportler *→ gesund) = 1.

Dass $p(X *\rightarrow Y) = 1$ gilt, schließt nicht aus, dass auch $p(Z *\rightarrow Y) = 1$. Logisch gesehen geht es bei X und Z um *hinreichende* (aber nicht notwendige) Bedingungen. Z. B. könnte ebenfalls gelten: p(Vegetarier *→ gesund) = 1.

Wie schon zuvor in diesem Exkurs verwende ich weiterhin überwiegend die *Positiv-Implikation* $X *\rightarrow Y$ anstatt der *Normal-Implikation* $X \rightarrow Y$, weil deren Paradoxien für eine Kausal-Analyse problematisch sind.

Man kann allerdings auch eine *statistische Kausalität* akzeptieren, d. h. $0 < p < 1$. Im Beispiel kann man sicher *nicht für alle* Sportler sagen, dass sie gesund sind. Dass hier wenigstens $p > 0$ gelten muss, ist trivial; man wird aber normalerweise einen höheren Wert fordern, z. B. $p > 0{,}75$. Bei statistischer Kausalität sind ggf. verschiedene Ursachen zu berücksichtigen (Multi-Kausalität), bei deren Kombination man dann doch den deterministischen Wert $p = 1$ erhält (vgl. später). Allerdings steht dahinter die Frage, ob die Welt grundsätzlich – in ihrer Tiefenstruktur – *deterministisch* strukturiert ist, was aber in diesem Rahmen nicht weiter verfolgt werden kann.

• hohes n
$p = 1$ kann für absolut ganz unterschiedlichen Brüche stehen:

1/1, 2/2, 3/3, 4/4, 5/5, 6/6, 7/7 ... n/n

Natürlich spricht normalerweise mehr für einen Kausal-Zusammenhang, wenn die Anzahl der untersuchten Objekt *n* möglichst hoch ist, wenn man also z. B. möglichst viele Sportler – als Stichprobe – untersucht bzw. es möglichst viele Sportler gibt.

Bei einem *singulären* Zusammenhang wäre es besonders problematisch, eine Kausal-Beziehung anzunehmen. Dass z. B. 1 Sportler gesund ist, reicht sicher nicht für ein Kausalgesetz.

Aber gerade in der Alltagssprache verwendet man oft singuläre Kausalsätze wie z. B.: „*Weil* ich hungrig bin, esse ich“. Wie ist das zu verstehen? Vereinfacht gesagt verwendet man hier *implizit* eine kausale Erklärung:

Für *alle* Menschen gilt: wenn sie hungrig sind, essen sie	$\Lambda x(Fx \rightarrow Gx)$
<u>Ich bin hungrig</u>	$\underline{Fx_i}$
Also (darum) esse ich	Gx_i

D. h. man bezieht sich doch auf ein *allgemeines* Gesetz, nur die Ableitung ist *singulär*.
Allerdings ist zu berücksichtigen, dass der Mensch – als *Individuum* – u. U. ein ganz spezielles Verhalten zeigt, das genau in der Weise *nur er selbst* so ausübt. Hier ist eine Erklärung mit allgemeinen Gesetzen sehr schwierig, dennoch muss man davon ausgehen, dass allgemeine Gesetze gültig sind, nur die Randbedingungen sind so extrem, dass sie letztlich zu einer singulären Erklärung führen.

• Negative Prämisse p = 0

Für eine Kausal-Relation „weil X, darum Y" muss man fordern, dass gilt: $p(\neg X *\rightarrow Y) = 0$ bzw. aussagen-logisch: $\neg(\neg X *\rightarrow Y)$, was äquivalent ist mit $\neg X *\rightarrow \neg Y$.
Also: p(nicht-Sportler *→ gesund) = 0.

Wenn nämlich z. B. gälte: p(nicht-Sportler *→ gesund) = 1, dann wäre das Sportler sein sicher nicht die Ursache von Gesundheit. Denn dann wären gleichermaßen alle Sportler und alle Nicht-Sportler gesund, das Sporttreiben wäre also nicht spezifisch für Gesundheit. Man muss allerdings nicht unbedingt $p(\neg X *\rightarrow Y) = 0$ fordern, aber $p(\neg X *\rightarrow Y) < p(X *\rightarrow Y$.

• theoretische Wahrscheinlichkeit p^T niedrig

Die theoretische Wahrscheinlichkeit ist (bei einer Struktur wie X *→ Y) am höchsten, wenn die *Zufallserwartung* vorliegt: p = 0,5. Umgekehrt ist sie am niedrigsten, wenn p = 1. Da wir unter Kausalität primär eine deterministische Beziehung mit p = 1 (oder p = 0) verstehen, spricht also ein niedriger Wert von p^T für Kausalität.

• Linearität

Wir haben bisher die Kausal-Relation *extensional* (normalerweise) *quantitativ* angegeben, also p(X *→ Y) = 1, Beispiel: „*alle* Sportler sind gesund".

Dabei haben wir aber die *Eigenschaften* nur *qualitativ* erfasst, d. h. jemand ist Sportler (ja) oder nicht (nein), er ist gesund (ja) oder nicht (nein).

Aber es ist auch möglich, die Eigenschaften – *intensional* – zu quantifizieren. Also im Beispiel: „wenn jemand zu r % sportlich ist, dann ist er zu s% sportlich". In diesem Fall muss man aber eine *lineare* Beziehung fordern, d. h.: „je sportlicher jemand ist, desto gesünder ist er".

• Unbegrenztheit

Grundsätzlich sollte ein Kausalgesetzt unbegrenzt sein, d. h. vor allem *zeitlich* und *räumlich unbegrenzt*. Prinzipiell sollte ein Kausalgesetz also *infinit* sein, für eine unendliche große Menge gelten.

Mit einem solchen Gesetz ergeben sich allerdings besondere Schwierigkeiten. Und in vielen Bereichen ist es auch angemessen, mit einem *finiten* Gesetz zu arbeiten. Nur darf ein Gesetz nicht durch künstliche Einschränkungen zu stark eingeengt sein. Z. B. „Alle Sportler, die sich zum Zeitpunkt i am Ort j befinden". So lässt sich normalerweise kein Kausalgesetz begründen. Auch die Stichprobe, jedenfalls als Zufallsstichprobe, darf nicht durch willkürliche Festsetzungen eingeschränkt sein.

Angenommen, eine Gruppe von Sportler trifft sich zu einem bestimmten Zeitpunkt in einem Zimmer, um Dopingmittel zu nehmen. Die Aussage, „alle Sportler, die zum Zeitpunkt i am Ort j befinden, sind gedopt" mag zwar wahr sein; man mag auch korrekt ableiten: „Sportler John ist zum Zeitpunkt i am Ort j", „also ist John gedopt". Aber dies taugt nicht als kausale Erklärung, hat keinen wirklichen Erklärungswert.

• Zeit

Die Ursache muss der Wirkung *zeitlich vorausgehen*.
Erst ist jemand Sportler, *dann* wird eher gesund.
Es ist zu diskutieren, ob es Kausalität gibt, bei der kein Zeitabstand zwischen Ursache und Wirkung besteht. Das wird einerseits im Bereich der *Mikrophysik* diskutiert, wo es z. B. Informationsübertagung ohne Zeitverzug zwischen Teilchen geben soll. Andrerseits wird im geistigen bzw. psychischen Bereich postuliert, dass die Wirkung *zeitgleich* bzw. gleichzeitig mit der Ursache auftritt. C. G. Jung nannte das *Synchronizität*. Beide dieser bizarren Formen von Kausalität sind allerdings umstritten.

• Raum
Klassisch versteht man unter einer Kausal-Relation eine Beziehung, wo das Ursachen-Objekt durch eine *Kraft Information*, *Energie* oder *Materie* auf ein Wirkungs-Objekt überträgt und bei diesem eine Veränderung hervorruft. Das heißt, die Übertragung erfolgt nicht nur in der Zeit, sondern auch innerhalb des Raums.
Allerdings ist oft eine räumliche Übertragung schwer nachvollziehbar, z. B. auch beim dem Beispiel: Sportler → gesund.
Vor allem, man kann auch im *psychischen* Bereich von Kausalität sprechen, die psychische Welt, die Welt des Bewusstseins ist aber unräumlich. Z. B. „Weil er wütend ist, schreit er". Hier ist es nicht sinnvoll, eine Übertragung durch den *Raum* zwischen dem Gefühl „Wut" und dem Verhalten „schreien" anzunehmen (wenn man nicht Transportwege über die Nervenbahnen annehmen will).

• Kausalrichtung
Manchmal ist nicht ganz eindeutig, was *Ursache* und was *Wirkung* ist.
– Z. B. ist (bzw. wird) jemand gesund, *weil* er Sportler ist?
– Oder ist (bzw. wird) jemand Sportler, *weil* er gesund ist?
Die beste Möglichkeit, das zu unterscheiden, ist die Zeit: das, was *zeitlich vorausgeht*, ist die *Ursache*, was zeitlich folgt, ist die *Wirkung*.

• Mono-Kausalität
Der Begriff der *Mono-Kausalität* ist nicht eindeutig, man kann zumindest folgende zwei Bedeutungen darunter verstehen:

– Erstens, man hat *eine* Ursache X, die vollständig ausreicht, die Wirkung Y zu erklären (es kann aber auch andere Ursachen geben). Logisch ergibt sich hier: $p(X *\rightarrow Y) = 1$, X ist also *hinreichende* Bedingung für Y, und zwar in einer *deterministischen* Beziehung. (Dieser Fall ist oben im Punkt Determinismus schon abgehandelt.)
– Zweitens, es gibt für ein Phänomen *nur eine einzige* Ursache: *strenge* Mono-Kausalität. Auf diese Deutung will ich mich im Folgenden beschränken.
In diesem Fall, wenn z. B. das Sportlersein die *einzige Ursache* von Gesundheit wäre, dann käme als logische Struktur die *Äquivalenz* in Frage.

Bei Verwendung der *Positiv-Implikation* bzw. Positiv-Äquivalenz ergäbe sich:
Sportler *↔ gesund. D. h. wenn jemand Sportler ist, ist er gesund. Und wenn jemand gesund ist, ist er Sportler. Allerdings müsste zusätzlich gelten: ¬Sportler *↔ ¬gesund.
Bei der *Normal-Implikation* gilt dagegen: $X \leftrightarrow Y \Leftrightarrow \neg X \leftrightarrow \neg Y$.
Hier ist (bei $X \leftrightarrow Y$) X also automatisch *notwendige* und *hinreichende* Bedingung für Y.
Man kann das so formulieren: „*Dann und nur dann*, wenn jemand sportlich ist, ist er auch gesund". Und es gilt ebenfalls das Umgekehrte. Somit reicht es hier, nur $X \leftrightarrow Y$ anzugeben bzw. quantitativ $p(X \leftrightarrow Y) = 1$.
Nun ist ja denkbar, dass auch gilt: $Z \leftrightarrow Y$. Ist dann die Mono-Kausalität aufgehoben?
In dem Fall gilt der Schluss: $(X \leftrightarrow Y) \wedge (Z \leftrightarrow Y) \Rightarrow (X \leftrightarrow Z)$. D. h. X und Z müssen auch äquivalent sein.
Z. B. angenommen „Sportler sein" ist äquivalent mit „Gesund sein". Und „Vollwertköstler sein" ist ebenfalls äquivalent mit „Gesund sein". Dann müsste gelten, dass alle Sportler Vollwertköstler sind und alle Vollwertköstler Sportler sind. Man könnte allerdings fragen, was hier die e*igentliche Ursache* ist (vgl. Schein-Kausalität); wenn man nur *eine* Ursache (z. B. X) als eigentliche Ursache von Y bestimmt, so bliebe letztlich doch die Mono-Kausalität erhalten.

Obwohl das paradox klingen mag, ist bei einer mono-kausalen *Äquivalenz* $X \leftrightarrow Y$ dennoch eine *Multi-Kausalitat* nicht zwingend ausgeschlossen (wie im nächsten Punkt gezeigt wird).

• Multi-Kausalität
Auch hier kann man wieder zwei Varianten unterscheiden:
– Erstens, es gibt *viele einzelne* Ursachen für Y: $(X_1 *\!\rightarrow Y) \wedge (X_2 *\!\rightarrow Y) \wedge ... \wedge (X_n *\!\rightarrow Y)$
– Zweitens, erst wenn viele Ursachen *zusammen* realisiert sind, tritt die Wirkung Y auf:
$X_1 \wedge X_2 \wedge ... \wedge X_n *\!\rightarrow Y$
Es mag quantitativ z. B. gelten: $p(X_1 *\!\rightarrow Y) = 0{,}4 \;\wedge\; p(X_2 *\!\rightarrow Y) = 0{,}6 \;\wedge\; p(X_3 *\!\rightarrow Y) = 0{,}8$
So könnte man denken, dass hier gar keine Kausalbeziehung vorliegt. Fasst man allerdings die verschiedenen Variablen zusammen, so mag sich doch eine *deterministische* Beziehung ergeben: $p(X_1 \wedge X_2 \wedge ... \wedge X_n *\!\rightarrow Y) = 1$
Multi-Kausalität ist eigentlich der Normalfall, aber sie ist oft implizit. Wir stellen eine einfache Relation $p(X *\!\rightarrow Y) = 1$ auf, sind uns aber nicht bewusst, dass dabei andere Bedingungen miteingehen.
Denkbar ist auch $p(X_1 \vee X_2 \vee ... \vee X_n *\!\rightarrow Y) = 1$. Hier reicht also, dass *eine von mehreren* Ursachen verwirklicht ist, damit die Wirkung eintritt.
Wie ist aber eine Struktur wie $X_1 \wedge X_2 \wedge ... \wedge X_n *\!\leftrightarrow Y$ zu interpretieren?
Hier ergeben sich zwei Möglichkeiten:
Erstens, man sieht diese Struktur als *multi-kausalen* Zusammenhang, da *mehrere* Ursachen $X_1 \wedge X_2 \wedge ... \wedge X_n$ die Wirkung Y hervorbringen.
Zweitens, man fasst $X_1 \wedge X_2 \wedge ... \wedge X_n$ als *eine komplexe* Ursache zusammen, die dann als *mono-kausale* Ursache von Y gilt (eben auf Grund der Äquivalenz-Beziehung). Damit würde das Konzept der Mono-Kausalität jedoch relativiert.

5.4 NICHT-KAUSALITÄT

Es gibt Relationen und Zusammenhänge, die Ähnlichkeit mit Kausal-Relationen besitzen, die man ggf. mit ihnen verwechseln kann, die aber nicht eine echte Ursache-Wirkungs-Struktur besitzen, jedenfalls keine einfache Struktur.

• Identität
Bei Identität findet man die logische Struktur der *Äquivalenz*: $X \leftrightarrow Y$ bzw. $X *\!\leftrightarrow Y$. Und es gibt keinen zeitlichen Unterschied zwischen „Ursache" und Wirkung". Hierzu muss man sich klarmachen: Bei Identität kann man real gar nicht zwischen 2 Objekten unterscheiden, wir nehmen dasselbe Objekt nur unter verschiedenen Bedingungen wahr, oder wir haben eben nur zwei Namen und stellen dann fest, dass diese dasselbe Objekt bezeichnen.
Es ist auch möglich, dass nur eine *Teilidentität* besteht. Z. B. könnte es sein, dass Gesundheit ein inhärenter Teil vom Sportlersein ist, dass Gesundheit in die Definition von Sportler mit eingeht. In diesem Fall besteht eine Teilidentität, die sich logisch durch die Implikation darstellt: Sportler $*\!\rightarrow$ gesund. Wenn jemand Sportler ist, muss er hier notwendig auch gesund sein, weil dies eben Teil der Sportleridentität ist, er also sonst kein Sportler wäre.

• Ableitung
Eine *tautologische Relation*, ein logischer Schluss, kann keine Kausal-Relation sein.
Z. B. $X \Rightarrow X$ bzw. $X \wedge Y \Rightarrow X$. Allein schon deshalb, weil X natürlich nicht (dem identischen) X *zeitlich* vorausgehen kann. Und weil ein Schluss keine neue Information aussagt.
Z. B. „Er treibt Sport, *darum* wird er gesund".

Hier noch einmal ein Verweis auf die *kausale Erklärung*: $\Lambda x(Fx \rightarrow Gx) \wedge Fx_i \Rightarrow Gx_i$
„Wenn alle Sportler gesund sind und John Sportler ist, dann ist John gesund".
Hier spielt zwar ein logischer Schluss eine Rolle, Gx_i wird *logisch abgeleitet*, aber das Kausal-Gesetz $\Lambda x(Fx \rightarrow Gx)$ selbst ist *synthetisch*, also kein Schluss.

• Wechsel-Wirkung
Bei der Wechsel-Wirkung ergibt sich im deterministischen Fall: $p(X \leftrightarrow Y) = 1$
Eine Veränderung von X führt zu einer Veränderung von Y. Diese Veränderung von Y führt wiederum zu einer Veränderung von X. X wirkt auf Y ein. Y wirkt dann auf X zurück (*Feedback*).
Dann gibt es zwei Möglichkeiten:
– X wird durch das Feedback verstärkt, der Wert von X wird erhöht (*positives Feedback*)
– X wird abgeschwächt, der Wert von X wird erniedrigt (*negatives Feedback*)
Hier muss der 1. Zustand von X (X_1) dem ersten Zustand von Y (Y_1) zeitlich vorausgehen, darauf folgt wiederum ein Zustand von X (X_2) usw. Eine genaue Analyse der Wechselwirkung erfordert eine aufwendigere quantitative Analyse.

• Kausal-Reihe
Die Struktur bei 3 Variablen sei: $X \rightarrow Z \rightarrow Y$.
Z. B. Sport → Muskeltraining → Gesundheit.
Also: Sport führt dazu, dass die Muskeln trainiert werden, und das hat dann den positiven Effekt auf die Gesundheit. Mit dieser Kausalreihe haben wir wohl keine Probleme, wir können dennoch Sport als die *primäre* Ursache akzeptieren. Aber was ist mit der folgenden Reihe?
Sport → Beliebtheit → Gesundheit
Nehmen wir einmal vereinfachend an, alle Sportler seien beliebt; und dass Beliebtheit ein psychosomatischer Faktor ist, der die Gesundheit fördert, die körperliche Bewegung selbst aber keinen besonderen Gesundheitseffekt besitzt. Hier hätten wir sicher Probleme, Sport als die (oder eine) *echte* Ursache von Gesundheit zu akzeptieren.

• Schein-Kausalität
Es gibt aber noch ein größeres Problem bei der Struktur $X \rightarrow Z \rightarrow Y$, z. B.:
Sport → Vollwertkost → Gesundheit
Es gilt im Beispiel: 70% Sportler sind gesund, es gibt also eine hohe Korrelation von Sport zu Gesundheit. Wir vermuten daher, dass Sport ein wichtiger Kausalfaktor für Gesundheit ist. Aber die Sportlichkeit ist nicht die eigentliche Ursache des Gesundseins. Es gibt einen *zusätzlichen Faktor*, z. B. gesunde Ernährung (Vollwertkost).
Und zwar korrelieren Sport und Vollwertkost total in unserem Beispiel, d. h. 100% der Sportler sind Vollwertköstler.
Der Kausal-Zusammenhang lässt sich folgendermaßen prüfen: Wir untersuchen die Nicht-Sportler. Wir stellen fest, dass 40 % der Nicht-Sportler auch Vollwertköstler sind; von denen sind aber 90% gesund. (Von den Nicht-Vollwert-Essern seien nur 20% gesund.). Es zeigt sich also, dass Sport hier gar nicht die Gesundheit befördert, sondern sogar ein ungünstiger Faktor ist. Der Sport ist hier also nur die *Schein-Ursache* der Gesundheit, die wahre Ursache ist die Ernährung.

• Zufall
Ist es denkbar, dass bei hoher Korrelation keinerlei Kausalität besteht (aber auch nicht Identität usw.), sondern *purer Zufall*? Offensichtlich ja.

Ähnliche Verhältnisse kennen wir beim Glücksspiel, das ja nur auf *Zufall* beruht. Z. B. beim Roulette. Jeder Kugelwurf beim Roulette ist ein singuläres Ereignis, das unabhängig ist von vorhergegangen Kugelwürfen. Dort haben die Farben *rot* und *schwarz* die gleichen Chancen, da es 18 rote und 18 schwarze Zahlen gibt. Von daher ist ein Ausgleich z. B. von *rot* und *schwarz* immer das wahrscheinlichste. Aber es kann eben auch 10mal oder mehr nacheinander *rot* kommen, obwohl das sehr unwahrscheinlich ist. Wenn man so will, besteht hier eine „Korrelation" zwischen *Kugelwurf* und *Farbe*, aber es kann unmöglich Kausalität bestehen (das habe ich bereits ausführlich beschrieben).

Im Grunde sind die Verhältnisse bei der Kausal-Analyse noch viel komplizierter, man muss immer auch die *negativen* Fälle prüfen, sollte immer auch *Test-Variablen* einführen usw., aber ich wollte hier nur das Wesentliche darstellen, außerdem nicht zu sehr in die *Statistik* übergleiten.

5.5 KAUSAL-FEHLER

Wie oben schon angedeutet, gibt es verschiedene *Fehlermöglichkeiten* bei der Kausalanalyse. Ein sehr wichtiger Aspekt der Kausalanalyse ist daher die Vermeidung von Fehlern. Dabei muss man sich bei der Kausalanalyse vor allem vor zwei *Fehlern* hüten:

- Fehler 1: eine *Abweichung* von der Zufalls-Erwartung irrtümlich für eine *Abhängigkeit* zu halten
- Fehler 2: das *Eintreffen* der Zufalls-Erwartung irrtümlich für eine *Unabhängigkeit* zu halten

5.5.1 Kausal-Fehler 1

Zu *Fehler 1*: Angenommen man wirft 10mal eine Münze, dann wäre die Zufalls-Erwartung, dass 5 mal Zahl und 5mal die Rückseite (Bild) kommt. Macht man diesen Versuch aber real, so zeigt sich, dass man kaum je dieses Ergebnis erhält. Angenommen es kommt 7mal Zahl und 3mal Bild, ist das schon ein Zeichen, dass hier eine kausale Abhängigkeit vorliegt? Offensichtlich nicht, denn es handelt sich beim Werfen einer Münze um ein *Glücksspiel*. Zwischen den Würfen besteht normalerweise *Unabhängigkeit*. Das Ergebnis weicht zufällig von der Zufalls-Erwartung ab. Man kann erwarten, dass sich bei sehr großem ‚n' (bzw. auf lange Sicht) die Treffer mit Münze und Bild in etwa ausgleichen, dies ist aber nicht gesichert. Und während sich die Zufalls-Erwartung rein mathematisch fest bestimmen lässt, erfordert die Beschreibung des realen Auftretens von Häufigkeiten *empirische* Gesetze, die nie so exakt sind wie mathematische Gesetze.

Auch bei der Untersuchung empirischer Ereignisse, zwischen denen eine (z. B. kausale) Abhängigkeit bestehen kann, sieht man nicht gleich jedes Abweichen von der Zufalls-Erwartung als einen Kausal-Zusammenhang. Sondern erst ab einer gewissen Größe der Abweichung, einer *signifikanten* Abweichung, ist man bereit, Kausalität zu unterstellen.

5.5.2 Kausal-Fehler 2

Zu *Fehler 2*: Auch wenn man bei einer empirischen Untersuchung ein Ergebnis bekommt, dass der *Zufalls-Erwartung* entspricht oder nahe an ihr dran ist, ist damit ein Zusammenhang noch nicht ausgeschlossen.

Einerseits kann es sich auch hier um *zufällige* Abweichungen handeln. Z. B. untersuche ich 50 Elemente einer Menge von 100 Elementen, davon haben 25 die Eigenschaft F und 25 nicht. Es liegt also exakt die Zufalls-Erwartung vor. Es könnte aber passieren, obwohl es nicht wahrscheinlich ist, dass wenn man weiter untersucht, alle weiteren 50 Elemente die Eigenschaft F haben, man also insgesamt ein Verhältnis von 75 zu 25 erhält, was sehr wohl für eine Abhängigkeit spricht, wenn auch eine *deterministische* Kausalität ausgeschlossen ist.

Außerdem können andere (Kausal-) Faktoren reinspielen, welche die Beziehung beeinflussen (Störgrößen).

5.5.3 Zusammenfassung

Ich fasse einige wesentliche Aspekte der Kausalanalyse noch einmal anhand eines klassischen Beispiels zusammen:

Man stellt zunächst eine *funktionale Abhängigkeit* zwischen folgenden Ereignissen fest:
X = es regnet, Y = die Strasse ist nass.

Der Zusammenhang ist: „Wenn es regnet, dann gilt: die Straße ist nass“. $X \rightarrow Y$.
Hier sind vor allem folgende Faktoren relevant:

- *Gesetzmäßigkeit*: der Zusammenhang muss normalerweise in allen Fällen auftreten, also immer und überall. Logisch: p(es regnet → die Strasse ist nass) = 1
- Die Ursache geht der Wirkung *zeitlich* voraus: *erst* regnet es, *dann* wird die Strasse nass (das führt über die Logik hinaus).
- Bei Kausalität in der materiellen Welt: Es erfolgt ein *Transport* von Materie, Energie oder Information, im Beispiel wird Wasser vom Himmel auf die Erde transportiert (das führt ebenfalls über die Logik hinaus).

Fehler begeht man grundsätzlich, wenn man die Korrelation nicht genügend absichert, wenn man den Zeitverlauf nicht beachtet und keine Erklärung für den Weg der Verursachung anbietet.

INHALTS DES EXKURSES „WAHRSCHEINLICHKEIT UND KAUSALITÄT“