Kapitel 2
Der wirtschaftliche Wert von Information

In diesem Kapitel wollen wir sehen, wie wir den Wert der Information einer Zufallsvariablen quantifizieren können. Dieser Wert hängt natürlich stark vom Umfeld ab und kann nur in Abhängigkeit davon betrachtet werden.

Wir interessieren uns hauptsächlich für Anwendungen im Portfoliomanagement, betrachten aber zum besseren Verständnis der Grundideen zunächst den mathematisch deutlich einfacheren Fall von Pferdewetten.

§ 1: Pferdewetten

Bei einem Rennen starten m Pferde, wobei das k-te mit einer Wahrscheinlichkeit p_k siegen wird. Vor dem Rennen können Wetten auf den Sieger abgeschlossen werden. Dazu legt der Wettanbieter für jedes der Pferde eine Quote o_k fest: Wer einen Einsatz e auf das k-te Pferd setzt, verliert diesen Einsatz, falls ein anderes Pferd gewinnt; wenn er richtig getippt hat, bekommt er das o_k-fache seines Einsatzes ausbezahlt, hat also einen Gewinn von $(o_k - 1)e$.

Um nicht gegebenenfalls alles zu verlieren, wird ein Spieler seinen Einsatz oft auf mehrere Pferde verteilen; diese Verteilung beschreiben wir durch ein *Portfolio*, d.h. einen Vektor $(b_1, \dots, b_m)$ mit $b_k \geq 0$ für alle k und $\sum_{k=1}^m b_k = 1$; dabei soll b_k festlegen, daß vom Gesamteinsatz e der Teil $b_k e$ auf das k-te Pferd gesetzt wird.

Das Rennen selbst beschreiben wir durch eine Zufallsvariable X, die den Wert k annimmt, wenn das k-te Pferd siegt. Die weitere Zufallsvariable $S(X)$ gibt an, womit der Einsatz am Ende des Rennens multipliziert wird; für $X = k$ ist das $S(k) = b_k o_k$, denn nur der auf das k-te Pferd gesetzte Teil des Einsatzes wird mit o_k multipliziert; der Rest verfällt. Der Erwartungswert für $S(X)$ ist somit

$$\mathbb{E}\big(S(X)\big) = \sum_{k=1}^{m} p_k b_k o_k \,.$$

Eine möglicher Strategie zur Optimierung des Portfolios ist die Maximierung dieses Erwartungswerts; da sowohl die Zielfunktion als auch die Nebenbedingungen an die b_k linear sind, handelt es sich hier um ein Problem der linearen Optimierung, d.h. das Maximum wird in einem Eckpunkt angenommen.

Die Eckpunkte des zulässigen Bereichs sind die Punkte, in denen ein $b_k = 1$ ist, während der Rest verschwindet; für diese Portfolios ist der Erwartungswert von $S(X)$ gleich $p_k o_k$. Bei dieser Strategie setzt man also den gesamten Einsatz auf dasjenige Pferd, bei dem das Produkt $p_k o_k$ maximal wird. Für jemand, der jede Woche eine kleinere Summe setzt, deren Totalverlust er verschmerzen kann, ist das durchaus eine sinnvolle Strategie; man muß aber bedenken, daß die Wahrscheinlichkeit eines Totalverlusts mit $1 - p_k$ sehr groß sein kann: Das maximale Produkt könnte beispielsweise einem krassen Außenseiter entsprechen, für den, da niemand mit seinem Sieg rechnet, eine große Quote o_k angeboten wird.

Wir betrachten Pferderennen hier als Einstieg in das deutlich komplexere Gebiet der Kapitalanlage; hier ist eine Strategie mit sehr wahrscheinlichem Totalverlust natürlich nicht akzeptabel, und wir müssen unser Optimierungsproblem anders stellen.

Der Ansatz, den wir hier verfolgen, geht zurück auf eine Arbeit von J.L. KELLY JR. von den (im Rahmen der Öffnung des amerikanischen Telephonmarkts inzwischen aufgeteilten) *Bell Telephone Laboratories*, die dieser im Juli 1956 in deren Forschungszeitschrift unter dem Titel *A New Interpretation of Information Rate* veröffentlichte. Er bringt Nachhaltigkeit dadurch ins Spiel, daß er von einer vielfachen Wiederholung des Rennens ausgeht. Anstelle eines einzigen Rennens, beschrieben durch eine Zufallsvariable X, betrachten wir also eine große Anzahl n von

Rennen, beschrieben durch unabhängige Zufallsvariablen $X_1, \ldots, X_n$, die allesamt dieselbe Verteilungsfunktion haben wie X. Beim ersten Rennen X_1 setzen wir den Einsatz gemäß dem gewählten Portfolio, bei jedem folgenden Rennen setzen wir die Auszahlung des vorigen Rennens gemäß diesem Portfolio. Nach n Rennen ist der ursprüngliche Einsatz dann multipliziert mit

$$S_n = \prod_{i=1}^{n} S(X_i)\,.$$

Eine nachhaltige Strategie könnte darauf basieren, daß S_n als Funktion von n zumindest im Mittel möglichst schnell wachsen sollte. Strategien mit hohem Risiko eines Totalverlusts sind dadurch praktisch ausgeschlossen, denn sobald ein $S(X_i)$ verschwindet, ist $S_n = 0$ für alle $n \geq i$.

JOHN LARRY KELLY JR. (1923–1965) war während des zweiten Weltkriegs vier Jahre lang Pilot bei der US Air Force; nach dem Krieg begann er ein Physikstudium an der University of Texas in Austin, das er 1953 mit der Promotion abschloß. Nach dem Studium arbeitete er bei den Bell Labs unter anderem auf dem Gebiet der Sprachsynthese. Bei der Anwendung der Informationstheorie auf die Spieltheorie arbeitete er eng mit SHANNON zusammen, der die Resultate im Gegensatz zu KELLY auch wirklich in Las Vegas anwendete.

Das Wachstumsverhalten von S_n kann am besten mittels des Erwartungswerts des Logarithmus von $S(X)$ beschrieben werden:

Definition: Die Verdoppelungsrate eines Rennens mit Portfolio b und Wahrscheinlichkeitsverteilung p ist

$$W(b,p) \underset{\text{def}}{=} \mathbb{E}\big(\log S(X)\big) = \sum_{k=1}^{m} p_k \log_2(b_k o_k)\,.$$

Lemma: S_n wächst als Funktion von n mit Wahrscheinlichkeit eins asymptotisch wie $2^{nW(b,p)}$.

Beweis: Da wir die hypothetischen Wiederholungen X_i als unabhängig annehmen, sind auch die Zufallsvariablen $\log_2 S(X_i)$ unabhängig, und sie haben allesamt dieselbe Verteilung wie $\log_2 S(X)$. Nach dem Gesetz der großen Zahl konvergiert

$$\frac{1}{n}\log_2 S(X) = \frac{1}{n}\sum_{i=1}^{n} \log_2 S(X_i)$$

daher mit Wahrscheinlichkeit eins gegen den Erwartungswert

$$\mathbb{E}\big(\log_2 S(X)\big) = W(b,p)\,,$$

der Logarithmus von S_n wächst also wie $nW(b,p)$. ■

Wenn wir ein möglichst schnelles Wachstum von S_n wollen, müssen wir also die Verdoppelungsrate $W(b,p)$ maximieren.

Definition: Ein Portfolio b^* heißt *log-optimal*, wenn $W(b,p)$ für $b = b^*$ seinen maximalen Wert annimmt.

(Da die Bedingungen $b_k \geq 0$ und $\sum b_k = 1$ eine kompakte Teilmenge des $\mathbb{R}^m$ definieren, ist klar, daß die stetige Funktion $W(b,p)$ dort ein Maximum annehmen muß.)

Zur Berechnung eines log-optimalen Portfolios arbeiten wir wieder mit LAGRANGE-Multiplikatoren; um die Ableitungen einfach zu halten, maximieren wir anstelle von $W(b,p)$ die dazu proportionale Funktion

$$f(b) = W(b,p)\cdot \ln 2 = \sum_{k=1}^{m} p_k \ln b_k o_k\,.$$

Ihre partielle Ableitung nach b_k ist

$$\frac{\partial f(b)}{\partial b_k} = \frac{p_k}{b_k}\,;$$

die Nebenbedingungsfunktion $g(b) = \sum b_k - 1$ hat partielle Ableitung eins. Im Optimum muß es daher ein $\lambda \in \mathbb{R}$ geben mit

$$\frac{p_k}{b_k} = \lambda \quad \text{oder} \quad p_k = \lambda b_k \quad \text{für alle } k\,.$$

Da sowohl die Summe aller p_k als auch die aller b_k eins sind, muß $\lambda = 1$ sein, d.h. $b_k = p_k$. In diesem Punkt $b^* = p$ ist

$$\begin{aligned} W(b^*, p) = W(p,p) &= \sum_{k=1}^{m} p_k \log_2 p_k o_k \\ &= \sum_{k=1}^{m} p_k \log_2 o_k + \sum_{k=1}^{m} p_k \log_2 p_k \\ &= \sum_{k=1}^{m} p_k \log_2 o_k - H(p)\,; \end{aligned}$$

wir müssen zeigen, daß dies der maximal mögliche Wert für $W(b,p)$ ist. Für jedes Portfolio b gilt

$$\begin{aligned} W(b,p) &= \sum_{k=1}^{m} p_k \log_2 b_k o_k = \sum_{k=1}^{m} p_k \log_2 \left(\frac{b_k}{p_k} p_k o_k\right) \\ &= \sum_{k=1}^{m} p_k \log_2 o_k + \sum_{k=1}^{m} \log_2 p_k + \sum_{k=1}^{m} p_k \log_2 \frac{b_k}{p_k} \\ &= \sum_{k=1}^{m} p_k \log_2 o_k - H(p) - D(p\|b) = W(b^*, p) - D(p\|b) \\ &\le W(b^*, p)\,, \end{aligned}$$

da die KULLBACK-LEIBLER-Distanz $D(p\|b)$ keine negativen Werte annimmt. Sie verschwindet genau dann, wenn $b = p = b^*$ ist, also liegt dort das einzige Maximum.

Das optimale Wachstum von S_n wird also erreicht, wenn wir auf jedes Pferd genau den Anteil des Einsatzes wetten, der seiner Gewinnwahrscheinlichkeit entspricht.

Bei der praktischen Umsetzung dieser wie auch fast jeder anderen Strategie haben wir das Problem, daß wir die Gewinnwahrscheinlichkeiten nicht wirklich kennen; wir können sie nur schätzen anhand von früheren Rennen der beteiligten Pferde und ähnlichen Informationen. Genau das gleiche Problem hat freilich auch der Wettanbieter bei der Festlegung seiner Quoten: Auch er kennt die p_k nicht.

Nehmen wir zunächst an, er wolle seine Quoten so bestimmen, daß jeder Wetter *im Mittel* gerade seinen Einsatz zurückbekommt, daß der Wettanbieter also *im Mittel* weder einen Gewinn noch einen Verlust macht. Für einen Wetter, der einen Einsatz e auf das k-te Pferd setzt, ist die erwartete Auszahlung gleich $p_k o_k e$; bei so einer *fairen* Wette sollte also $o_k = 1/p_k$ sein. Da der Wettanbieter die p_k nicht kennt, muß er stattdessen mit einer geschätzten Wahrscheinlichkeitsverteilung $(r_1, \ldots, r_m)$ arbeiten und $o_k = 1/r_k$ setzen; entsprechend nimmt der Wetter als Portfolio *seine* Schätzung $(b_1, \ldots, b_m)$ der Wahrscheinlichkeitsverteilung. Die Verdoppelungsrate bei dieser Strategie ist

$$\begin{aligned} W(b,p) &= \sum_{k=1}^{m} p_k \log_2 b_k o_k = \sum_{k=1}^{m} p_k \log_2 \left(\frac{b_k}{p_k}\frac{p_k}{r_k}\right) \\ &= \sum_{k=1}^{m} p_k \log_2 \frac{p_k}{r_k} - \sum_{k=1}^{m} p_k \log_2 \frac{p_k}{b_k} \\ &= D(p\|r) - D(p\|b)\,, \end{aligned}$$

d.h. die Verdoppelungsrate ist positiv, wenn der Wetter die Wahrscheinlichkeitsverteilung besser geschätzt hat, und negativ, falls der Buchmacher die bessere Schätzung hatte.

(Echte Sportwetten sind natürlich nie fair; hier sind die Quoten deutlich kleiner als $1/r_k$, denn sowohl der Wettanbieter als auch das Finanzamt und gegebenenfalls der Ausrichter des Rennens wollen ein sicheren Gewinn einstreichen.)

Wer nicht professionell auf Pferde wettet, dürfte im allgemeinen keine realistische Chance haben, die Gewinnwahrscheinlichkeiten besser zu schätzen als ein erfahrenes auf Pferdewetten spezialisiertes Wettbüro. Er könnte aber seine Chancen erhöhen, indem er den Rat von Fachleuten einholt, also zum Beispiel Spezialzeitschriften oder Rundbriefe abonniert. Da diese nicht umsonst verteilt werden, stellt sich natürlich die Frage, ob sich dieser Aufwand lohnt.

Abstrakt mathematisch betrachtet läßt sich der Wert solcher Zusatzinformation einfach berechnen: Wir kodieren die Information in einer Zufallsvariablen Y und können nun die gemeinsame Wahrscheinlichkeitsverteilung der beiden Zufallsvariablen X und Y betrachten. Ohne

Zusatzinformation würden wir mit einen Portfolio $b = (b_1, \ldots, b_m)$ arbeiten, das auf unserer Schätzung des Vektors der Wahrscheinlichkeiten beruht; mit Kenntnis von Y verwenden wir stattdessen ein Portfolio $b(y) = \big(b_1(y), \ldots, b_m(y)\big)$, das auf den bedingten Wahrscheinlichkeiten $p_k(y) = p(X = k|Y = y)$ unter der Nebenbedingung $Y = y$ beruht. Die optimale Verdoppelungsrate ohne Kenntnis von Y ist, wie wir oben gesehen haben,

$$W^*(X) = \sum_{k=1}^{m} p_k \log_2 o_k - H(X)\,;$$

mit Kenntnis von Y erreichen wir

$$W^*(X|Y) = \sum_{k=1}^{m} \sum_{y} p_k(y) \log_2 p_k(y) o_k = \sum_{k=1}^{m} p_k \log_2 o_k - H(X|Y)\,.$$

Die Differenz zwischen den beiden Werten ist

$$\Delta W = W^*(X|Y) - W^*(X) = H(X) - H(X|Y) = I(X;Y)\,;$$

der Zuwachs bei der optimalen Verdoppelungsrate ist also gerade die wechselseitige Information der beiden Zufallsvariablen.

§ 2: Portfolio Management

Wenn man Zeitungsberichte der letzten Jahre liest, bekommt man durchaus den Eindruck, als handelten selbst große Banken am Aktienmarkt ähnlich wie Zocker, die bei Pferderennen alles auf einen Außenseiter mit minimalen Gewinnchancen setzen. Trotzdem gibt es aus Sicht eines Mathematikers große Unterschiede zwischen einem Pferderennen und einer Börse: Während es beim Pferderennen immer nur einen Sieger gibt, kann es an der Börse durchaus vorkommen, daß an einem Tag alle gehandelten Aktien gewinnen und an einem anderen Tag alle verlieren. Dabei geht es, von eher seltenen Ausnahmen abgesehen, bei Gewinn und Verlust nicht um alles oder nichts, sondern an den meisten Börsentagen um eher moderate Kursveränderungen.

Entsprechend komplizierter muß auch unser mathematisches Modell sein: Eine einzige Zufallsvariable, die den Gewinner des Rennens beschreibt, genügt hier definitiv nicht mehr.

a) Das Modell

Wenn wir von einer Börse ausgehen, an der m Aktien gehandelt werden, brauchen wir für jede einzelne Aktie eine eigene Zufallsvariable, die deren Entwicklung beschreibt. Wir gehen der Einfachheit halber davon aus, daß wir uns für den Wert der einzelnen Aktien nur in gewissen diskreten Zeitintervallen interessieren; der Anschaulichkeit halber wird im folgenden meist von einem Börsentag die Rede sein, beim computergestützten Handel an heutigen Börsen kann es sich bei diesem Zeitintervall aber auch durchaus um eine Minute oder einen noch kleineren Zeitraum handeln.

Für jede der m Aktien betrachten wir eine Zufallsvariable X_k, die angibt, mit welchem Faktor der Kurs dieser Aktie im betrachteten Zeitraum multipliziert wird. Wir müssen realistischerweise davon ausgehen, daß wir über die Verteilungsfunktionen dieser Zufallsvariablen nur wenig wissen: Durch langfristige Beobachtung können wir zwar einige statistische Kennwerte der Verteilung schätzen; da wir aber nicht davon ausgehen können, daß die Verteilungsfunktion langfristig stabil bleibt, sind selbst diese Schätzungen von begrenztem Nutzen, und über die exakte Verteilungsfunktion werden wir nur in seltenen Fällen halbwegs zuverlässige Aussagen machen können. In diesem Paragraphen muß es daher zwangsläufig deutlich weniger konkret zugehen als im vorigen.

Trotzdem arbeiten wir im wesentlichen mit denselben Begriffen: Wir beschreiben das Geschehen an der Börse durch den Vektor $X = (X_1, \ldots, X_m)$ aus den Zufallsvariablen für die Wertentwicklung der einzelnen Aktien und wählen als Anlagestrategie wieder ein Portfolio $b = (b_1, \ldots, b_m)$, das angibt, welcher Teil des Anlagekapitals wir in welche Aktie investieren.

Vom gesamten Kapitaleinsatz K entfällt also der Teil $b_k K$ auf die k-te Aktie, und dieser wird am Ende des Börsentags mit dem Wert von X_k multipliziert. Der Wert der gesamten Anlage ist dann also

$$\sum_{k=1}^{m} b_k K X_k = K \cdot S(X) \quad \text{mit} \quad S(X) = \sum_{k=1}^{m} b_k X_k = \langle b, X\rangle\ .$$

Eine mögliche Anlagestrategie könnte darin bestehen, daß wir den Erwartungswert der Wertentwicklung $S(X)$ maximieren wollen; da

$$\mathbb{E}\big(S(X)\big) = \sum_{k=1}^{m} b_k \mathbb{E}(X_k)$$

linear in den b_k ist, würde dies bedeuten, daß wir alles Kapital in die Aktie(n) mit dem höchsten Erwartungswert investieren. Da die tatsächliche Entwicklung der Aktie nicht durch den Erwartungswert, sondern durch den Wert einer Zufallsvariable bestimmt wird, ist klar, daß so eine Strategie ihre Risiken hat.

b) log-optimale Portfolios

Eine mögliche Alternative, die von manchen Großinvestoren anscheinend auch wirklich angewendet wird, ist wieder der Ansatz von KELLY: Wir wählen die Strategie, bei der wir bei beliebig häufiger Wiederholung des Börsentags die größte Wachstumsrate erzielen. Wir betrachten also anstelle des einen Vektors X eine Folge von Vektoren $X^{(i)}$ von jeweils m Zufallsvariablen, wobei die $X^{(i)}$ voneinander unabhängig sein sollen, aber allesamt dieselbe Verteilungsfunktion F haben. Wir suchen ein Portfolio b, für das die Folge der

$$S_n = \prod_{i=1}^{n} S(X^{(i)})$$

zumindest im Mittel das größtmögliche Wachstum hat. Man beachte, daß wir bei diesem Ansatz zwar von einem festen Portfolio b ausgehen, daß dieses sich aber nicht auf die Anzahl, sondern auf den *Wert* der Aktien bezieht. Da sich dieser Wert ständig ändert, muß nach jeder „Wiederholung“ eines Börsentags der Aktienbestand neu ausbalanciert werden, damit wieder der Anteil b_k des vorhandenen Kapitals in Aktie k investiert ist.

Um in diesem Sinne optimale Portfolios zu charakterisieren, betrachten wir wieder die *Verdoppelungsrate* als Erwartungswert des Logarithmus von$S(X)$, d.h.

$$W(b,F) = \mathbb{E}\big(\log_2 S(X)\big) = \mathbb{E}\big(\log_2(\langle b,X\rangle\big) = \int \log_2 \langle b,X\rangle \; dF(x)\,.$$

Wie im Fall der Pferdewetten gilt

Satz: Mit Wahrscheinlichkeit eins wächst S_n wie $2^{nW(b,F)}$.

Auch der *Beweis* geht im wesentlichen wie dort: Nach dem Gesetz der großen Zahlen konvergiert

$$\frac{1}{n}\log_2 S_n = \frac{1}{n}\sum_{i=1}^{n} S(X^{(i)}) = \frac{1}{n}\sum_{i=1}^{n}\log_2 \langle b, X^{(i)}\rangle$$

mit Wahrscheinlichkeit eins gegen den Erwartungswert von $\log_2 \langle b, X\rangle$, also gegen $W(b,F)$.

■

Definition: *a)* Die optimale Wachstumsrate $W^*(F)$ ist das Maximum von $W(b,F)$, wobei b alle Portfolios durchläuft.
b) Ein Portfolio b heißt *log-optimal*, wenn $W(b,F) = W^*(F)$ ist.

Da die Menge aller Portfolios kompakt ist, wissen wir, daß $W^*(F)$ existiert und daß es log-optimale Portfolios gibt; da wir F nicht kennen, können wir allerdings weder die Verdoppelungsraten $W(b,F)$ noch deren Maximalwert $W^*(F)$ wirklich ausrechnen. Trotzdem können wir einige Aussagen über diese Funktionen machen:

Lemma: *a)* $W(b,F)$ ist linear in F und konkav in b.
b) $W^*(F)$ ist konvex in F.

Beweis: a) Die Linearität in F folgt sofort aus der Linearität der Integration. Da der Logarithmus eine konkave Funktion ist, gilt für alle $\lambda \in [0,\,1]$

$$\log_2 \langle (1-\lambda)b_1 + \lambda b_2, X\rangle \ge (1-\lambda)\log_2 \langle b_1, X\rangle + \lambda \log_2 \langle b_2, X\rangle\ ;$$

aus der Monotonie der Integration folgt damit auch die Konkavität von $W(b,F)$ in b.

b) F_1 und F_2 seien zwei Verteilungsfunktionen, und $b^*(F_1), b^*(F_2)$ seien log-optimale Portfolios dazu. Dann ist für jedes $\lambda \in [1,\,0]$ auch $(1-\lambda)F_1 + \lambda F_2$ eine Verteilungsfunktion; das log-optimale Portfolio dazu sei $b^*\big((1-\lambda)F_1 + \lambda F_2\big)$. Dann ist

$$W^*\big((1-\lambda)F_1 + \lambda F_2\big) = W\Big(b^*\big((1-\lambda)F_1 + \lambda F_2\big), (1-\lambda)F_1 + \lambda F_2\Big)$$

$$= (1-\lambda)W\Big(b^*\big((1-\lambda)F_1+\lambda F_2\big), F_1\Big)+\lambda W\Big(b^*\big((1-\lambda)F_1+\lambda F_2\big), F_2\Big)$$
$$\le (1-\lambda)W\big(b^*(F_1), F_1\big)+\lambda W\big((b^*(F_2), F_2\big) = (1-\lambda)W^*(F_1)+\lambda W^*(F_2)\,.$$
■

Im Gegensatz zur Situation bei den Pferdewetten gibt es hier natürlich keinen Grund für die Annahme, es gäbe nur ein log-optimales Portfolio; es kann eine ganze Reihe davon geben. Wir können allerdings zeigen

Lemma: Die Menge aller log-optimaler Portfolios zu einer festen Verteilungsfunktion f ist konvex.

Beweis: b und b' seien zwei log-optimale Portfolios. Wegen der Konkavität von $W(b,F)$ in b ist dann für jedes $\lambda \in [0,\ 1]$

$$\begin{aligned} W\big((1-\lambda)b_1+\lambda b_2\big) &\ge (1-\lambda)W(b_1,F)+\lambda W(b_2,F) \\ &= (1-\lambda)W^*(F)+\lambda W^*(F) = W^*(F)\,. \end{aligned}$$

Da W^* nach Definition die maximal erreichbare Wachstumsrate ist, muß also $W\big((1-\lambda)b_1+\lambda b_2\big) = W^*(F)$ sein, d.h. auch $(1-\lambda)b_1+\lambda b_2$ ist ein log-optimales Portfolio.
■

c) Eine erste Charakterisierung log-optimaler Portfolios

Da wir nur wenig über die Verteilungsfunktion F wissen, kann uns auch ein Ansatz via LAGRANGE-Multiplikationen kein konkretes Gleichungssystem liefern, anhand dessen wir ein log-optimales Portfolio berechnen könnten. Zusammen mit der Konkavität von $W(b,F)$ in b liefert er aber ein nützliches Kriterium zur Charakterisierung log-optimaler Portfolios:

Satz: *a)* Ein Portfolio $b^* = (b_1^*,\dots,b_m^*)$ ist genau dann log-optimal bezüglich der Verteilungsfunktion F, wenn für jedes andere Portfolio b gilt:

$$\mathbb{E}\left(\frac{\langle b,X\rangle}{\langle b^*,X\rangle}\right) \le 1\,.$$

b) Ein Portfolio $b^* = (b_1^*,\dots,b_m^*)$ ist genau dann log-optimal bezüglich der Verteilungsfunktion F, wenn

$$\mathbb{E}\left(\frac{X_k}{\langle b^*,X\rangle}\right) \le 1 \qquad \text{für alle } k\,.$$

c) Ist für ein log-optimales Portfolio b^* die k-te Komponente b_k^* von Null verschieden, so ist sogar

$$\mathbb{E}\left(\frac{X_k}{\langle b^*,X\rangle}\right) = 1\,.$$

Beweis: a) Wie immer wenn wir differenzieren müssen betrachten wir die Wachstumsrate bezüglich des natürlichen Logarithmus, also die Funktion

$$V(b) = W(b,F)\cdot\ln 2 = \mathbb{E}\big(\ln\langle b,X\rangle\big)\,.$$

Da auch diese Funktion konkav ist, hat sie genau dann ein Maximum in einem Punkt b^* des zulässigen Bereichs, wenn ihre Richtungsableitung entlang der Strecke von b^* zu irgendeinem anderen Portfolio b stets kleiner oder gleich Null ist. (Wegen der Konvexität der Menge aller Portfolios liegt diese Strecke im zulässigen Bereich.)

Die Verbindungsstrecke besteht aus den Punkten $b_\lambda = (1-\lambda)b^*+\lambda b$ mit $\lambda \in [0,\ 1]$ und

$$\begin{aligned} \left.\frac{dV(b_\lambda)}{d\lambda}\right|_{\lambda=0^+} &= \left.\frac{d}{d\lambda}\mathbb{E}\big(\ln\langle b_\lambda,X\rangle\big)\right|_{\lambda=0^+} \\ = \lim_{\lambda\searrow 0}\mathbb{E}\left(\frac{\ln\langle b_\lambda,X\rangle-\ln\langle b^*,X\rangle}{\lambda}\right) &= \lim_{\lambda\searrow 0}\mathbb{E}\left(\ln\frac{\langle b_\lambda,X\rangle}{\langle b^*,X\rangle}\right) \\ &= \lim_{\lambda\searrow 0}\frac{1}{\lambda}\mathbb{E}\left(\ln\frac{(1-\lambda)\langle b^*,X\rangle+\lambda\langle b,X\rangle}{\langle b^*,X\rangle}\right) \\ &= \mathbb{E}\left(\lim_{\lambda\searrow 0}\frac{1}{\lambda}\ln\left(1+\lambda\left(\frac{\langle b,X\rangle}{\langle b^*,X\rangle}-1\right)\right)\right) \end{aligned}$$

Aus der TAYLOR-Entwicklung von

$$\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots$$

oder auch aus dem Mittelwertsatz der Differentialrechnung folgt, daß

$$\lim_{\lambda\to 0}\frac{\ln(1+\lambda x)}{\lambda} = x$$

ist, also können wir die Ableitung weiter ausrechnen als

$$\left.\frac{dV(b_\lambda)}{d\lambda}\right|_{\lambda=0^+} = \mathbb{E}\left(\frac{\langle b,X\rangle}{\langle b^*,X\rangle}\right)-1\,.$$

b^* ist somit genau dann ein log-optimales Portfolio, wenn dies für alle Portfolios b kleiner oder gleich Null ist, und das wiederum ist äquivalent zur Ungleichung $\mathbb{E}\left(\frac{\langle b,X\rangle}{\langle b^*,X\rangle}\right) \le 1$.

b) Da alle $b_k \ge 0$ sind und

$$\mathbb{E}\left(\frac{\langle b,X\rangle}{\langle b^*,X\rangle}\right) - 1 = \sum_{k=1}^m b_k \mathbb{E}\left(\frac{X_k}{\langle b^*,X\rangle}\right) - 1 = \sum_{k=1}^m b_k \left(E\left(\frac{X_k}{\langle b^*,X\rangle}\right) - 1\right)$$

ist, gilt die Bedingung aus *a)* genau dann, wenn alle $\frac{E\left(X_k\right.}{\left.\langle b^*,X\rangle\right)\le 1}$ sind.

c) Falls b_k^* nicht verschwindet, betrachten wir jenes Portfolio b, das das gesamte Kapital in die k-te Aktie investiert. b^* ist dann ein innerer Punkt auf dem Durchschnitt der Geraden durch b^* und b mit der Menge aller Portfolios; daher hat $V(b)$ auf diesem Durchschnitt genau dann ein Maximum in b^*, wenn dort die Richtungsableitung sogar verschwindet, d.h. $\frac{\mathbb{E}\left(X_k\right.}{\left.\langle b^*,X\rangle\right)}$ muß gleich eins sein. ■

Die Bedingung aus *a)* sagt insbesondere, daß log-optimale Portfolios auch bezüglich des Erwartungswerts des Quotienten der Gewinnentwicklung optimal sind; sie erfüllen also auch ein kurzfristiges Optimalitätskriterium.

Nach der hier betrachteten Strategie wird nach jedem Börsentag das vorhandene Kapital so reinvestiert, daß der Anteil b_k auf die k-te Aktie entfällt. Welcher Teil des Kapitals am Ende des Börsentags in welcher Aktie steckt, hängt natürlich von der konkreten Entwicklung am jeweiligen Tag ab, aber für ein log-optimales Portfolio b^* können wir zumindest den Erwartungswert berechnen: Das gesamte Kapital nach Handelsschluß ist das $\langle b^*, X\rangle$-fache des Ausgangskapitals K, und aus dem Investment $b_k^* K$ in die k-te Aktie wurde ein Betrag von $b_k^* K X_k$. Der Anteil der k-ten Aktie ist daher $b_k^* X_k / \langle b^*, X\rangle$ mit Erwartungswert

$$\mathbb{E}\left(\frac{b_k^* X_k}{\langle b^*,X\rangle}\right) = b_k^* \mathbb{E}\left(\frac{X_k}{\langle b^*,X\rangle}\right) = b_k^*$$

unabhängig davon, ob b_k^* verschwindet oder nicht. Zumindest was die Erwartungswerte betrifft, bleiben die Anteile also stabil.

d) Asymptotische Optimalität

Wenn wir von einer Börse ausgehen, bei der die Verteilungsfunktion F über einen längeren Zeitraum fest bleibt, können wir die KELLY-Strategie statt auf hypothetische Wiederholungen eines Börsentags auch auf die Folge der Börsentage anwenden.

Da die Verteilungsfunktion konstant bleibt, ist auch das log-optimale Portfolio jeden Tag das gleiche; wir investieren also nach einem festen log-optimalen Portfolio b^*, wobei wir zu Beginn eines jedes Börsentags dafür sorgen, daß unabhängig von der Kursentwicklung des Vortags wieder der Anteil b_k^* des Anlagekapitals in Aktie k investiert wird.

Diese Strategie erscheint recht unflexibel; wenn wir stattdessen das Portfolio jeden Tag neu festlegen in Abhängigkeit von der bisherigen Kursentwicklung der Aktien, können wir möglicherweise ein besseres Ergebnis erzielen.

Definition: Eine *kausale Anlagestrategie* ist eine Folge von Abbildungen

$$b^{(i)}\colon \begin{cases} \mathbb{R}^{m(i-1)} \to \mathbb{R}^m \\ (x^{(1)},\dots,x^{(i-1)}) \mapsto b^{(i)}(x^{(1)},\dots,x^{(i-1)}) \end{cases},$$

die in Abhängigkeit von den Kurs Entwicklungen $X^{(j)} = x^{(j)}$ für $j < i$ das Portfolio für den i-ten Börsentag festlegt.

Eine einfache Abschätzung zeigt aber, daß man trotz der größeren Flexibilität mit einer solchen Strategie zumindest im Mittel keine besseren Ergebnisse erzielt als mit der log-optimalen Strategie:

Lemma: Bezeichnet $S_n = \prod_{i=1}^n S(X^{(i)})$ die Wertentwicklung nach n Tagen für eine kausale Anlagestrategie, so ist $\mathbb{E}(\log_2 S_n) \le nW^*$, wobei W^* die Verdoppelungsrate der log-optimalen Strategie ist.

Beweis: Der Erwartungswert von $\log_2 S_n$ ist höchstens gleich

$$\max_{b^{(1)},\dots,b^{(n)}} \mathbb{E}(\log S_n) = \max_{b^{(1)},\dots,b^{(n)}} \mathbb{E}\left(\sum_{i=1}^{n} \log_2 \left\langle b^{(i)}, X^{(i)} \right\rangle\right)$$
$$= \sum_{i=1}^{n} \max_{b^{(i)}} \mathbb{E}\big(\log_2 \left\langle b^{(i)}, X^{(i)} \right\rangle\big) = nW^* \,. \qquad \blacksquare$$

Nun wissen wir freilich, daß der Erwartungswert allein noch nicht viel aussagt; das gerade bewiesene Lemma schließt nicht aus, daß wir mit einer geeigneten kausalen Strategie vielleicht doch mit hoher Wahrscheinlichkeit besser fahren als mit der log-optimalen. Um zu sehen, daß auch das nicht der Fall ist, benötigen wir zunächst zwei Aussagen aus der Wahrscheinlichkeitstheorie, als erstes einen einfachen Spezialfall der Ungleichung von MARKOV:

Lemma: Ist X eine Zufallsvariable, die nichtnegative reelle Zahlen als Werte annimmt, so ist für jedes $a > 0$

$$p(X \ge a) \le \frac{\mathbb{E}(X)}{a} \,.$$

Beweis: $\chi_a(x)$ sei die charakteristische Funktion der Menge aller reeller Zahlen größer oder gleich a. Dann ist, wenn P das Wahrscheinlichkeitsmaß bezeichnet,

$$p(X \ge a) = \int \chi_a(x)\, dP \le \int \chi_a(x) \frac{x}{a}\, dP \le \int \frac{x}{a}\, dP = \frac{\mathbb{E}(X)}{a} \,. \qquad \blacksquare$$

Lemma von Borel und Cantelli: $E_1, E_2, \dots$ sei eine Folge von Ereignissen derart, daß $\sum_{i=1}^{\infty} p(E_i)$ konvergiert. Dann ist die Wahrscheinlichkeit für das Eintreten von unendlich vielen dieser Ereignisse gleich Null.

Beweis: Die Wahrscheinlichkeit dafür, daß irgendein Ereignis E_k mit $k \ge n$ eintritt, ist höchstens $p_n = \sum_{i=n}^{\infty} p(E_i)$; wegen der vorausgesetzten Konvergenz der Summe bilden die p_n eine Nullfolge. Falls unendlich viele der Ereignisse E_i auftreten, tritt insbesondere für jedes n mindestens eines mit $i \ge n$ ein; die Wahrscheinlichkeit für das

Eintreten unendlich vieler der Ereignisse ist somit für jedes n kleiner oder gleich p_n. Damit muß diese Wahrscheinlichkeit gleich Null sein. $\blacksquare$

EMILE BOREL (1871–1956) wurde in Saint-Affrique in den Pyrenäen als Sohn eines protestantischen Geistlichen geboren. Ab 1889 studierte er Mathematik an der Ecole Normale in Paris, wo er 1893 promovierte. Danach bekam er zunächst eine Stelle als *maître de conférence* an der Universität von Lille, drei Jahre später an der Ecole Normale Supérieure; von 1899 bis 1902 lehrte er am Collège de France. 1909 richtete die Sorbonne speziell für ihn einen Lehrstuhl ein, den er bis 1941 innehatte. Ab 1924 war er auch politisch aktiv als Abgeordneter der Nationalversammlung (1924–1936) und als Marineminister (1925–1940). Während des zweiten Weltkriegs kämpfte er für die Résistance. Seine mathematischen Arbeiten kommen aus fast allen Teilgebieten der Mathematik; besonders bekannt sind seine Beiträge zur Maß- und Wahrscheinlichkeitstheorie sowie zur Theorie reeller und komplexer Funktionen.

FRANCESCO PAOLO CANTELLI (1875–1966) wurde in Palermo geboren und studierte an der dortigen Universität Mathematik. Neben seinem Studium arbeitete er auch am dortigen Observatorium und veröffentlichte verschiedene Arbeiten über Astronomie. Von 1903 bis 1923 arbeitete er als Aktuar für eine Versicherung und beschäftigte sich mit Anwendungen der Wahrscheinlichkeitstheorie; danach lehrte er Finanz- und Versicherungsmathematik zunächst an der Universität von Catania, ab 1925 in Neapel und ab 1931 bis zu seiner Emeritierung 1951 in Rom. Neben vielen anderen Arbeiten zur Wahrscheinlichkeitstheorie bewies er unter anderem auch das starke Gesetz der großen Zahlen.

Außerdem sei noch an zwei Begriffe aus der Analysis erinnert:

Definition: *a)* Eine reelle Zahl x heißt *Häufungspunkt* der reellen Zahlenfolge $(x_n)_{n\in\mathbb{N}}$, wenn es zu jedem $\varepsilon > 0$ unendlich viele Folgenglieder x_n gibt mit $|x - x_n| < \varepsilon$.
b) Der kleinste Häufungspunkt einer Folge wird als *Limes inferior* $\varliminf_{n\to\infty} x_n$ bezeichnet, der größte als *Limes superior* $\varlimsup_{n\to\infty} x_n$.

Damit können wir nun beweisen, daß keine Strategie langfristig wesentlich besser sein kann als die log-optimale:

Satz: $X^{(1)}, X^{(2)}, \dots$ sei eine Folge unabhängiger Vektoren von Zufallsvariablen, die allesamt Verteilungsfunktion F haben. Das log-optimale Portfolio dazu sei b^*, und $b^{(1)}, b^{(2)}, \dots$ sei irgendeine kausale Anlagestrategie. Dann gilt für die Wertentwicklungen $S_n^* = \prod_{i=1}^n \langle b^*, X^{(i)} \rangle$ und $S_n = \prod_{i=1}^n \langle b^{(i)}, X^{(i)} \rangle$ mit Wahrscheinlichkeit eins, daß

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log_2 \frac{S_n}{S_n^*} \leq 0 .$$

Beweis: Wie wir aus Abschnitt *c)* wissen, ist der Erwartungswert von S_n/S_n^* kleiner oder gleich eins; nach der Ungleichung von MARKOV ist daher

$$p(S_n > n^2 S_n^*) = p\left(\frac{S_n}{S_n^*} > n^2\right) < \frac{1}{n^2} .$$

Da $\sum_{n=1}^\infty \frac{1}{n^2}$ konvergiert, sagt uns das Lemma von BOREL und CANTELLI, daß mit Wahrscheinlichkeit null unendlich viele der Ungleichungen

$$\frac{S_n}{S_n^*} > n^2 \quad \text{oder} \quad \frac{1}{n} \log_2 \frac{S_n}{S_n^*} > \frac{2 \log_2 n}{n}$$

erfüllt sind. Es gibt daher mit Wahrscheinlichkeit eins ein $N \in \mathbb{N}$, so daß für alle $n \geq N$ gilt

$$\frac{1}{n} \log \frac{S_n}{S_n^*} < \frac{2 \log_2 n}{n} .$$

Da rechts eine Nullfolge steht, kann der Limes superior nicht positiv sein. ■

e) Der Einfluß zusätzlicher Information

Nicht nur bei Pferderennen, sondern auch an der Börse wird viel mehr oder weniger nützliche Zusatzinformation angeboten. Um deren Wert abzuschätzen, betrachten wir zunächst, wie sich die Wachstumsrate eines log-optimalen Portfolios ändert, wenn wir von einer falschen Verteilungsfunktion ausgehen. Bei Pferderennen hatten wir gesehen, daß

dies mit der KULLBACK-LEIBLER-Distanz der Wahrscheinlichkeitsverteilungen zusammenhängt; da wir hier kontinuierliche Zufallsvariablen haben, müssen wir diese erst definieren:

Definition: *a)* Die KULLBACK-LEIBLER-Distanz zwischen zwei Wahrscheinlichkeitsdichten f und g ist

$$D(f\|g) = \int f(x) \log_2 \frac{f(x)}{g(x)}\, dx .$$

b) Die wechselseitige Information zweier Zufallsvariablen X und Y mit den Wahrscheinlichkeitsdichten f_X und f_Y sowie der gemeinsamen Wahrscheinlichkeitsdichte f ist

$$I(X;Y) = \int f(x,y) \log_2 \frac{f(x,y)}{f_X(x) f_Y(y)}\, dx\, dy .$$

Satz: $f(x_1, \dots, x_m)$ sei die Wahrscheinlichkeitsdichte zum Vektor X der Zufallsvariablen $X_1, \dots, X_m$, und b_f sei das log-optimale Portfolio dazu. $g(x_1, \dots .x_m)$ sei eine weitere Wahrscheinlichkeitsdichte, und b_g sei das log-optimale Portfolio zu g. Dann ist

$$\Delta W = W(b_f, F) - W(b_g, F) \leq D(f\|g) .$$

Beweis: Nach Definition ist

$$\begin{aligned}
\Delta W &= \int f(x) \log_2 \langle b_f, x \rangle\, dx - \int f(x) \log_2 \langle b_g, x \rangle\, dx \\
&= \int f(x) \log_2 \frac{\langle b_f, x \rangle}{\langle b_g, x \rangle}\, dx \\
&= \int f(x) \log_2 \frac{\langle b_f, x \rangle}{\langle b_g, x \rangle} \frac{g(x)}{f(x)} \frac{f(x)}{g(x)}\, dx \\
&= \int f(x) \log_2 \frac{\langle b_f, x \rangle}{\langle b_g, x \rangle} \frac{g(x)}{f(x)}\, dx + D(f\|g)
\end{aligned}$$

Da f eine Wahrscheinlichkeitsdichte ist, können wir die Ungleichung von JENSEN auf das letzte Integral anwenden und den Logarithmus nach