

## Vorwort Heft 3-2011

**Hans-Christoph Grunau**

Online publiziert: 4. August 2011

© Vieweg+Teubner und Deutsche Mathematiker-Vereinigung 2011

„Ricci-Fluss“ und „Bewertung von Finanzprodukten“ sind die gleichermaßen aktuellen wie unterschiedlichen Themenschwerpunkte des vorliegenden Heftes des Jahresberichts.

Gemäß einem klassischen Ergebnis von Heinz Hopf ist eine einfach zusammenhängende kompakte  $n$ -dimensionale Riemannsche Mannigfaltigkeit mit konstanter Schnittkrümmung 1 schon isometrisch zur Standardsphäre. Mit Blick auf dieses Ergebnis warf ebenfalls Heinz Hopf die Frage auf, ob solche Mannigfaltigkeiten mit Schnittkrümmungen im Intervall  $(1 - \varepsilon, 1]$  immerhin noch topologische Sphären seien. Weitergehend wurde diese Frage von Marcel Berger und Wilhelm Klingenberg 1960/61 positiv sogar für Schnittkrümmungen im Intervall  $(\frac{1}{4}, 1]$  beantwortet; dieses Intervall ist auch optimal. Damit drängen sich aber sofort weitere Fragen auf, etwa ob ein analoges bzw. geeignet modifiziertes Resultat auch mit einer Diffeomorphieausage, ohne die Voraussetzung einfachen Zusammenhangs oder bei nur punktwisen Schranken für das Verhältnis der Schnittkrümmungen zu beweisen sei. Simon Brendle diskutiert diesen Themenkomplex in seinem Übersichtsartikel „Der Sphärensatz in der Riemannschen Geometrie“ und ordnet seine hier – zum Teil gemeinsam mit Richard Schoen – erzielten Ergebnisse in den Kontext des 1982 von Richard Hamilton eingeführten Ricci-Flusses ein. Simon Brendle widmet seinen Beitrag dem am 14.10.2010 verstorbenen Wilhelm Klingenberg.

Die Finanzkrisen der letzten Jahre haben deutlich vor Augen geführt, wie wichtig eine fundierte Bewertung und das Risikomanagement des Portfolios von Versicherungsgesellschaften für deren wirtschaftlichen Erfolg – oder oft sogar nur für deren wirtschaftliches Überleben – sind. Für eine fundierte mathematische Modellierung

---

H.-Ch. Grunau (✉)

Institut für Analysis und Numerik, Fakultät für Mathematik, Otto-von-Guericke-Universität,  
Postfach 4120, 39016 Magdeburg, Deutschland

e-mail: [hans-christoph.grunau@ovgu.de](mailto:hans-christoph.grunau@ovgu.de)

und Theorie muss man finanzmathematische Konzepte in die klassische Versicherungsmathematik einführen. Dieses erläutern Thomas Knispel, Gerhard Stahl und Stefan Weber in ihrem Übersichtsartikel „From the equivalence principle to market consistent valuation“ und geben nach einer ausführlichen Erläuterung der Grundkonzepte und einer Einführung unter anderem in das Black-Scholes-Modell einen Überblick über moderne mathematische Instrumente zur Risikobewertung.

Die Buchbesprechungen diskutieren Neuerscheinungen zu Ruinwahrscheinlichkeiten sowie zu Gruppenwirkungen und Ergodentheorie.



## Der Sphärensatz in der Riemannschen Geometrie

Simon Brendle

Eingegangen: 2. Mai 2011 / Online publiziert: 18. Juni 2011  
© Vieweg+Teubner und Deutsche Mathematiker-Vereinigung 2011

*Dem Gedenken an Wilhelm P.A. Klingenberg gewidmet*

**Zusammenfassung** Das Zusammenspiel zwischen Krümmung und Topologie einer Riemannschen Mannigfaltigkeit ist eine der zentralen Fragen in der globalen Differentialgeometrie. Ein klassisches Ergebnis von Heinz Hopf besagt, daß jede kompakte, einfach zusammenhängende Mannigfaltigkeit mit konstanter Krümmung 1 isometrisch zur runden Sphäre ist. Inspiriert durch dieses Ergebnis stellte Hopf die Frage, ob eine kompakte, einfach zusammenhängende Mannigfaltigkeit, deren Schnittkrümmungen im Intervall  $[1 - \varepsilon, 1]$  liegen, topologisch eine Sphäre ist.

In ersten Teil dieser Arbeit beschreiben wir die Geschichte dieses Fragenkomplexes. Im zweiten Teil diskutieren wir den Differenzierbaren Sphärensatz, sowie einige verwandte Ergebnisse. Die Beweise verwenden eine Vielzahl unterschiedlicher Methoden, darunter Stabilitätsargumente für Geodäten und Minimalflächen, sowie die Ricci-Fluß-Methode von Richard Hamilton. Für die detaillierten Argumente sei auf die Monographie des Authors (Graduate Studies in Mathematics, Bd. 111, Amer. Math. Soc., 2010) verwiesen.

**Schlüsselwörter** Krümmung · Ricci-Fluß · Sphärensatz

**Mathematics Subject Classification (2000)** 53C20 · 53C21 · 53C44 · 35K55

---

S. Brendle (✉)  
Department of Mathematics, Stanford University, 450 Serra Mall, Bldg. 380, Stanford, CA 94305,  
USA  
e-mail: [brendle@math.stanford.edu](mailto:brendle@math.stanford.edu)

## 1 Der Begriff der Krümmung

Der Begriff der Krümmung einer glatten Mannigfaltigkeit spielt eine zentrale Rolle in der Differentialgeometrie. Die einfachsten Beispiele glatter Mannigfaltigkeiten sind zwei-dimensionale Flächen im Euklidischen Raum  $\mathbb{R}^3$ . Für jede geschlossene Fläche  $M \subset \mathbb{R}^3$  läßt sich eine Distanzfunktion  $d : M \times M \rightarrow \mathbb{R}$  definieren. Der Abstand von zwei Punkten  $p, q \in M$  ist dabei definiert als die Länge der kürzesten Kurve in  $M$ , die  $p$  und  $q$  verbindet. Die so definierte Distanzfunktion wird auch als intrinsische Distanzfunktion bezeichnet. Ist beispielsweise  $M$  die Oberfläche der Einheitskugel in  $\mathbb{R}^3$ , so ist  $d(p, q)$  die Länge eines Großkreisbogens von  $p$  nach  $q$ . Die intrinsische Distanzfunktion kann durch den metrischen Tensor von  $M$  beschrieben werden. Der metrische Tensor von  $M$  liefert für jeden Punkt  $p \in M$  ein inneres Produkt auf der Tangentialebene  $T_p M \subset \mathbb{R}^3$ . Dieses innere Produkt  $g : T_p M \times T_p M \rightarrow \mathbb{R}$  erhält man durch Einschränkung des gewöhnlichen Skalarprodukts für Vektoren in  $\mathbb{R}^3$ . Der metrische Tensor erlaubt die Berechnung der Länge von Tangentialvektoren und von Kurven in  $M$ . Die intrinsische Distanz erhält man dann durch Minimierung der Länge von Kurven in  $M$ .

Für jeden Punkt  $p \in M$  kann man eine quadratische Form auf der Tangentialebene  $T_p M$  definieren. Diese quadratische Form mißt die infinitesimale Veränderung des Einheitsnormalenvektors zu  $M$  und wird als zweite Fundamentalform von  $M$  bezeichnet. Die Hauptkrümmungen von  $M$  am Punkt  $p$  sind definiert als die Eigenwerte der zweiten Fundamentalform von  $M$ . Ist  $M$  eine Ebene in  $\mathbb{R}^3$ , so ist der Einheitsnormalenvektor konstant, und die Hauptkrümmungen verschwinden. Im Fall der Oberfläche der Einheitskugel in  $\mathbb{R}^3$  sind die Hauptkrümmungen konstant gleich 1. Die Hauptkrümmungen einer konvexen Fläche haben stets dasselbe Vorzeichen; ist  $M$  dagegen eine Sattelfläche, so haben die Hauptkrümmungen verschiedene Vorzeichen.

Die Gaußsche Krümmung  $K$  ist definiert als das Produkt der Hauptkrümmungen von  $M$  am Punkt  $p$ . Das Theorema Egregium von Gauß besagt, daß die Gaußsche Krümmung von  $M$  eine intrinsische Größe ist. Dies bedeutet, daß zur Berechnung der Gaußsche Krümmung nur die Kenntnis des metrischen Tensors (bzw. der intrinsischen Distanzfunktion) erforderlich ist. Die Gaußsche Krümmung verschwindet genau dann, wenn  $M$  lokal isometrisch zu  $\mathbb{R}^2$  ist. Als Beispiel betrachten wir die Einheitssphäre in  $\mathbb{R}^3$ . In diesem Fall ist die Gaußsche Krümmung gleich 1; es ist daher unmöglich, eine längentreue Abbildung von einem Gebiet auf der Einheitssphäre auf ein Gebiet in  $\mathbb{R}^2$  zu finden. Die Oberfläche eines Zylinders oder Kegels hingegen hat Gaußsche Krümmung 0 und kann isometrisch auf ein Gebiet in der Ebene abgebildet werden.

Das Theorema Egregium ermöglicht es, einen Begriff der Krümmung für abstrakte zwei-dimensionale Mannigfaltigkeiten zu definieren. Mit anderen Worten, man betrachtet eine abstrakte zwei-dimensionale Mannigfaltigkeit  $M$ , die nicht notwendigerweise in  $\mathbb{R}^3$  eingebettet ist. Wir nehmen ferner an, daß eine Metrik  $g$  auf  $M$  gegeben ist. Dem Paar  $(M, g)$  läßt sich dann eine wohldefinierte Gaußsche Krümmung  $K$  zuordnen. Das Integral der Gaußschen Krümmung steht in enger Beziehung zur Euler-Charakteristik von  $M$ :

**Satz von Gauß-Bonnet** Sei  $M$  eine abstrakte zwei-dimensionale Mannigfaltigkeit mit einer Metrik  $g$ . Dann gilt

$$\int_M K \, d\text{vol} = 2\pi \chi(M),$$

wobei  $K$  die Gaußsche Krümmung und  $d\text{vol}$  das Flächenelement bezeichnet.

Die Euler-Charakteristik  $\chi(M)$  ist eine topologische Invariante von  $M$ . Für eine orientierbare Fläche gilt  $\chi(M) = 2 - 2k$ , wobei  $k$  das Geschlecht von  $M$  bezeichnet. Eine zwei-dimensionale Sphäre beispielsweise hat Geschlecht 0 und Euler-Charakteristik 2, während ein zwei-dimensionaler Torus Geschlecht 1 und Euler-Charakteristik 0 hat.

Die auf Gauß zurückgehende Theorie der Flächen wurde von Riemann auf höhere Dimensionen verallgemeinert. Wir betrachten im folgenden eine glatte Mannigfaltigkeit  $M$  der Dimension  $n \geq 2$ . Das bedeutet, daß jeder Punkt  $p \in M$  eine Umgebung besitzt, die homöomorph zu einem Gebiet in  $\mathbb{R}^n$  ist, und daß die Kartenwechsel glatte Funktionen sind. Für jeden Punkt  $p \in M$  kann man einen abstrakten Tangentialraum  $T_p M$  definieren. Wir nehmen, daß für jeden Punkt  $p \in M$  ein inneres Produkt  $g : T_p M \times T_p M \rightarrow \mathbb{R}$  gegeben ist, welches glatt von  $p$  abhängt. Man bezeichnet dann  $g$  als Riemannsche Metrik, und das Paar  $(M, g)$  als Riemannsche Mannigfaltigkeit.

Riemann definierte außerdem einen Begriff der Krümmung für Riemannsche Mannigfaltigkeiten. Genauer läßt sich jedem Punkt  $p \in M$  und jeder zwei-dimensionalen Ebene  $\pi \subset T_p M$  die sogenannte Schnittkrümmung  $K(\pi)$  zuordnen. Im folgenden skizzieren wir die geometrische Interpretation der Schnittkrümmung. Dazu betrachten wir kurze geodätische Segmente durch den Punkt  $p$ , deren Tangentialvektoren in der Ebene  $\pi$  liegen. Die Vereinigung dieser geodätischen Segmente definiert eine zwei-dimensionale Untermannigfaltigkeit  $\Sigma$  von  $M$ . Die Tangentialebene an  $\Sigma$  im Punkt  $p$  kann mit der Ebene  $\pi$  identifiziert werden. Die Schnittkrümmung  $K(\pi)$  ist dann genau die Gaußsche Krümmung der Untermannigfaltigkeit  $\Sigma$  am Punkt  $p$ .

Im folgenden betrachten wir einige wichtige Beispiele Riemannscher Mannigfaltigkeiten:

- Der Euklidische Raum  $\mathbb{R}^n$  hat konstante Schnittkrümmung  $K = 0$ .
- Die Einheitskugel  $S^n \subset \mathbb{R}^{n+1}$  besitzt eine kanonische Metrik mit konstanter Schnittkrümmung  $K = 1$ .
- Die Metrik  $g_{ij} = \frac{4}{(1-x^2)^2} \delta_{ij}$  auf der Einheitskugel  $B^n = \{x \in \mathbb{R}^n : |x| < 1\}$  hat konstante Schnittkrümmung  $K = -1$ . Die Riemannsche Mannigfaltigkeit  $(B^n, g)$  wird als der  $n$ -dimensionale hyperbolische Raum bezeichnet.

Durch Quotientenbildung erhält man weitere Beispiele von Mannigfaltigkeiten mit konstanter Schnittkrümmung. Die Quotienten von  $S^n$  wurden in [61] klassifiziert, und werden als sphärische Raumformen bezeichnet. Das einfachste Beispiel ist der reell-projektive Raum  $\mathbb{R}P^n$ . Diese Mannigfaltigkeit ist definiert als der Quotient von  $S^n$  unter der Spiegelung  $x \mapsto -x$ . Eine weitere Klasse von Beispielen bilden die Linsenräume. Diese Räume wurden bereits 1908 von Tietze [60] studiert. Um die Linsenräume zu definieren, betrachten wir zwei teilerfremde ganze Zahlen  $k, l$ . Wir bezeichnen mit  $\Gamma = \{\alpha \in \mathbb{C} : \alpha^k = 1\}$  die Gruppe der  $k$ -ten Einheitswurzeln. Die

Gruppe  $\Gamma$  wirkt auf der Sphäre  $S^3 \subset \mathbb{C}^2$  durch  $\alpha.(z_1, z_2) = (\alpha z_1, \alpha' z_2)$ . Die Wirkung von  $\Gamma$  auf  $S^3$  ist frei von Fixpunkten, und der Quotient  $S^3/\Gamma$  ist daher eine glatte Mannigfaltigkeit.

Wir diskutieren schließlich noch einige wichtige Beispiele von Mannigfaltigkeiten mit nicht-konstanter Schnittkrümmung:

- Der komplex-projektive Raum  $\mathbb{C}\mathbb{P}^m$  ist definiert als der Raum der komplexen Geraden in  $\mathbb{C}^{m+1}$ . ( $\dim \mathbb{C}\mathbb{P}^m = 2m$ )
- Der quaternionische-projektive Raum  $\mathbb{H}\mathbb{P}^m$  ist definiert als der Raum der quaternionischen Geraden in  $\mathbb{H}^{m+1}$ . ( $\dim \mathbb{H}\mathbb{P}^m = 4m$ )
- Die projektive Ebene über den Cayleyschen Oktonionen. ( $\dim \mathbb{O}\mathbb{P}^2 = 16$ )

Jeder der Räume  $\mathbb{C}\mathbb{P}^m$ ,  $\mathbb{H}\mathbb{P}^m$ , und  $\mathbb{O}\mathbb{P}^2$  besitzt eine kanonische Riemannsche Metrik, deren Schnittkrümmungen im Intervall  $[1, 4]$  liegen. Diese Räume werden als kompakte symmetrische Räume vom Rang 1 bezeichnet.

Neben den sphärischen Raumformen und den kompakten symmetrischen Räumen vom Rang 1 sind weitere Beispiele von Mannigfaltigkeiten mit positiver Schnittkrümmung bekannt. Für Details verweisen wir auf [1, 2], und [6].

Der Differenzierbare Sphärensatz liefert eine vollständige Klassifikation aller Mannigfaltigkeiten  $(M, g)$ , deren Schnittkrümmungen im Intervall  $[1, 4]$  liegen. Ist  $(M, g)$  eine Riemannsche Mannigfaltigkeit, die die Ungleichung  $1 \leq K \leq 4$  erfüllt, so ist entweder  $M$  diffeomorph zu einer sphärischen Raumform, oder  $M$  ist lokal isometrisch zu einem kompakten symmetrischen Raum vom Rang 1. In den folgenden Sektionen werden wir einige der zentralen Ideen skizzieren, die dem Beweis zugrundeliegen (siehe [12, 17, 18]).

## 2 Der Topologische Sphärensatz

Der Zusammenhang zwischen Krümmung und Topologie einer Riemannschen Mannigfaltigkeit wurde bereits 1926 in einer Arbeit von Hopf untersucht. Dort bewies Hopf folgenden Satz:

**Theorem 2.1** (H. Hopf [37, 38]) *Sei  $(M, g)$  eine kompakte, einfach zusammenhängende Mannigfaltigkeit mit konstanter Schnittkrümmung  $K = 1$ . Dann ist  $(M, g)$  isometrisch zur Einheitskugel  $S^n$ .*

Im Lichte dieses Ergebnisses stellte Hopf die Frage, ob eine kompakte, einfach zusammenhängende Riemannsche Mannigfaltigkeit, deren Schnittkrümmungen im Intervall  $[1 - \varepsilon, 1]$  liegen, immer noch topologisch eine Sphäre sein muß (siehe [7], Seite 545). Diese Frage wurde 1951 in einer grundlegenden Arbeit von Rauch [55] aufgegriffen. Rauch zeigte, daß jede kompakte, einfach zusammenhängende Mannigfaltigkeit, deren Schnittkrümmungen im Intervall  $[\delta, 1]$  liegen, homöomorph zu  $S^n$  ist. Die Konstante  $\delta$  ist explizit berechenbar ( $\delta \approx 0.75$ ), aber nicht optimal. Um 1960 gelang Berger und Klingenberg eine entscheidende Verbesserung der ‘‘Pinching-Konstante’’  $\delta$ :

**Theorem 2.2** (M. Berger [4]; W. Klingenberg [42]) *Sei  $(M, g)$  eine kompakte, einfach zusammenhängende Riemannsche Mannigfaltigkeit, deren Schnittkrümmungen im Intervall  $(1, 4]$  liegen. Dann ist  $M$  homöomorph zu  $S^n$ .*

Die Pinching-Konstante  $1/4$  ist optimal in Anbetracht der kompakten symmetrischen Räume vom Rang 1, die im vorangehenden Abschnitt besprochen wurden. Der Grenzfall von Mannigfaltigkeiten, deren Schnittkrümmungen im Intervall  $[1, 4]$  liegen, wurde von Berger [5] analysiert.

Theorem 2.2 ist bekannt als der Topologische Sphärensatz und stellt einen Meilenstein in der modernen Differentialgeometrie dar. Einige zentrale Fragen blieben jedoch unbeantwortet:

- Kann die Aussage, daß  $M$  homöomorph zu  $S^n$  ist, durch die stärkere Aussage ersetzt werden, daß  $M$  diffeomorph zu  $S^n$  ist?
- Können die globalen Krümmungsschranken in Theorem 2.2 durch eine punktweise Krümmungsbedingung ersetzt werden?
- Kann Theorem 2.2 auf Mannigfaltigkeiten verallgemeinert werden, die nicht einfach zusammenhängend sind?

Die Frage nach dem Diffeomorphietyp von  $M$  ist von besonderer Bedeutung aufgrund der Existenz von exotischen Sphären. 1957 konstruierte Milnor [47] Beispiele von Mannigfaltigkeiten, die zwar homöomorph, aber nicht diffeomorph zu  $S^7$  sind. Kervaire und Milnor [41] zeigten, daß die 7-dimensionale Sphäre  $S^7$  genau 28 verschiedene differenzierbare Strukturen zuläßt. Mannigfaltigkeiten, die homöomorph, aber nicht diffeomorph zu  $S^n$  sind, werden als exotische Sphären bezeichnet. Die folgende geometrische Interpretation der exotischen 7-Sphären wurde von Brieskorn gegeben (siehe [20, 21]): für  $k \in \{1, \dots, 28\}$  definiert man

$$\Sigma_k = \{(z_1, z_2, z_3, z_4, z_5) \in \mathbb{C}^5 : z_1^2 + z_2^2 + z_3^2 + z_4^3 + z_5^{6k-1} = 0\}$$

$$\cap \{(z_1, z_2, z_3, z_4, z_5) \in \mathbb{C}^5 : |z_1|^2 + |z_2|^2 + |z_3|^2 + |z_4|^2 + |z_5|^2 = 1\}.$$

Brieskorn zeigte, daß die Menge  $\Sigma_k$  eine glatte Mannigfaltigkeit ist, die homöomorph zu  $S^7$  ist. Die Mannigfaltigkeiten  $\Sigma_1, \dots, \Sigma_{28}$  realisieren alle 28 differenzierbaren Strukturen auf  $S^7$ .

Die oben erwähnten Fragen waren Gegenstand einer Vielzahl von Untersuchungen, wobei eine Reihe von Teilergebnissen erreicht wurden. Wir geben im folgenden eine Übersicht über diese Entwicklungen (siehe auch [19]). Gromoll [25] zeigte, daß jede kompakte, einfach zusammenhängende Mannigfaltigkeit, deren Schnittkrümmungen im Intervall  $[\delta(n), 1]$  liegen, diffeomorph zu  $S^n$  ist. Dasselbe Ergebnis wurde unabhängig von Calabi bewiesen. Die Pinching-Konstante  $\delta(n)$  hängt ausschließlich von der Dimension ab. Für  $n \rightarrow \infty$  konvergiert  $\delta(n)$  gegen 1. Sugimoto, Shiohama, und Karcher [59] gelang es, ein entsprechendes Ergebnis mit einer von  $n$  unabhängigen Pinching-Konstante ( $\delta \approx 0.87$ ) zu beweisen. Die Pinching-Konstante wurde von verschiedenen Autoren graduell verbessert (siehe [28, 56]). H. Im Hof und E. Ruh bewiesen schließlich eine äquivalente Version des Sphärensatzes (siehe auch [27]).

**Theorem 2.3** (H. Im Hof, E. Ruh [40]) *Es existiert eine monoton fallende Folge reeller Zahlen  $\delta(n) \rightarrow 0.68$  mit der folgenden Eigenschaft: ist  $M$  eine kompakte Rie-*

*mannsche Mannigfaltigkeit der Dimension  $n$ , deren Schnittkrümmungen im Intervall  $[\delta(n), 1]$  liegen, so ist  $M$  diffeomorph zu einer sphärischen Raumform.*

Wir bezeichnen mit  $\text{diam}(M, g) = \sup\{d(p, q) : p, q \in M\}$  den Durchmesser einer Riemannschen Mannigfaltigkeit  $(M, g)$ . Im folgenden Theorem wird die obere Schranke an die Schnittkrümmung durch eine untere Schranke an den Durchmesser von  $(M, g)$  ersetzt:

**Theorem 2.4** (K. Grove, K. Shiohama [29]) *Sei  $(M, g)$  eine kompakte Riemannsche Mannigfaltigkeit mit Schnittkrümmung  $K \geq 1$  und Durchmesser  $\text{diam}(M, g) > \frac{\pi}{2}$ . Dann ist  $M$  homöomorph to  $S^n$ .*

Die Beweise von Theorems 2.2, 2.3, und 2.4 verwenden klassische Methoden der Vergleichsgeometrie. Im Folgenden skizzieren wir einen Beweis von Theorem 2.4. Dieses Argument geht auf Berger zurück; ein detailliertes Argument findet man in [15], §1.3.

Der Einfachheit halber betrachten wir nur den Fall  $n \geq 4$ . Aufgrund von Arbeiten von Freedman [24] und Smale [58] genügt es zu zeigen, daß  $M$  eine Homotopiesphäre ist. Angenommen,  $M$  sei keine Homotopiesphäre. Dann existiert ein Index  $k \in \{1, \dots, n-1\}$  mit der Eigenschaft, daß die  $k$ -te Homotopiegruppe von  $M$  nicht verschwindet. Wir betrachten nun zwei Punkte  $p, q \in M$  mit maximalem Abstand, so daß  $d(p, q) = \text{diam}(M, g) > \frac{\pi}{2}$  ist. Wir bezeichnen mit  $\Omega$  den Raum aller Kurven  $\gamma : [0, 1] \rightarrow M$  mit  $\gamma(0) = \gamma(1) = p$ , und mit  $L : \Omega \rightarrow \mathbb{R}$  das Längenfunktional. Die erste Variation des Längenfunktionals an einem Punkt  $\gamma \in \Omega$  verschwindet genau dann, wenn  $\gamma$  eine Geodäte ist. Die zweite Variation des Längenfunktionals am Punkt  $\gamma$  hat endlich viele negative Eigenwerte. Die Anzahl dieser negativen Eigenwerte wird als der Morse-Index von  $\gamma$  bezeichnet.

Nach Voraussetzung gilt  $\pi_k(M) \neq 0$ , und daher  $\pi_{k-1}(\Omega) \neq 0$ . Mittels Morse-Theorie kann man zeigen, daß es eine Geodäte  $\gamma \in \Omega$  gibt, deren Morse-Index kleiner als  $k$  ist. Der Vergleichssatz von Toponogov impliziert, daß die Länge von  $\gamma$  größer als  $\pi$  ist. Eine Geodäte der Länge größer als  $\pi$  in einer Mannigfaltigkeit mit Schnittkrümmung  $K > 1$  ist jedoch stets instabil. Um dies zu sehen, betrachten wir ein paralleles Vektorfeld  $X(s)$  längs  $\gamma$ , das orthogonal zum Geschwindigkeitsvektor  $\gamma'(s)$  ist. Das Vektorfeld  $V(s) = \sin(\pi s) X(s)$  liefert dann eine Variation von  $\gamma$ , die die Endpunkte fixiert und die Bogenlänge reduziert. Auf diese Weise erhält man einen  $(n-1)$ -dimensionalen Raum von Variationsvektorfeldern, für die die zweite Variation der Bogenlänge negativ ist. Der Morse-Index von  $\gamma$  kann daher nicht kleiner als  $n-1$  sein. Dies widerspricht der Wahl von  $\gamma$ .

Am Ende dieses Abschnitts beschreiben wir eine Verallgemeinerung des Topologischen Sphärensatzes, die 1988 von Micallef und Moore [46] bewiesen wurde. Dazu ist es nützlich, anstelle der Schnittkrümmung den Riemannschen Krümmungstensor von  $(M, g)$  zu betrachten. Für jeden Punkt  $p \in M$  kann der Riemannsche Krümmungstensor als multilineare Abbildung  $R : T_p M \times T_p M \times T_p M \times T_p M \rightarrow \mathbb{R}$  aufgefaßt werden. Diese multilineare Abbildung erfüllt die algebraischen Identitäten

$$R(X, Y, Z, W) = -R(Y, X, Z, W) = R(Z, W, X, Y) \quad (1)$$



und

$$R(X, Y, Z, W) + R(Y, Z, X, W) + R(Z, X, Y, W) = 0 \quad (2)$$

für alle Vektoren  $X, Y, Z, W \in T_p M$ .

Der Riemannsche Krümmungstensor enthält dieselbe Information wie die Schnittkrümmung. Die Schnittkrümmung einer zwei-dimensionalen Ebene  $\pi$  kann in der Form

$$K(\pi) = \frac{R(X, Y, X, Y)}{|X|^2 |Y|^2 - \langle X, Y \rangle^2} \quad (3)$$

geschrieben werden, wobei  $\{X, Y\}$  eine beliebige Basis von  $\pi$  ist und  $R$  den Riemannschen Krümmungstensor bezeichnet. Umgekehrt ist der Riemannsche Krümmungstensor durch die Schnittkrümmung bereits eindeutig bestimmt.

In der Arbeit [46] führten Micallef und Moore den Begriff der isotropen Krümmung ein:

**Definition 2.5** Sei  $(M, g)$  eine Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$ . Die Mannigfaltigkeit  $(M, g)$  hat *nichtnegative isotrope Krümmung*, wenn für jedes orthonormale Vierbein  $\{e_1, e_2, e_3, e_4\} \subset T_p M$  die Ungleichung

$$\begin{aligned} &R(e_1, e_3, e_1, e_3) + R(e_1, e_4, e_1, e_4) \\ &+ R(e_2, e_3, e_2, e_3) + R(e_2, e_4, e_2, e_4) \\ &- 2R(e_1, e_2, e_3, e_4) \geq 0 \end{aligned}$$

erfüllt ist. Falls die strikte Ungleichung gilt, sagen wir, daß  $(M, g)$  *positive isotrope Krümmung* hat.

Micallef und Moore bewiesen den folgenden Sphärensatz für Mannigfaltigkeiten mit positiver isotroper Krümmung:

**Theorem 2.6** (M. Micallef, J.D. Moore [46]) *Sei  $(M, g)$  eine kompakte, einfach zusammenhängende Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$  mit positiver isotroper Krümmung. Dann ist  $M$  homöomorph zu  $S^n$ .*

Theorem 2.6 ist eine Verallgemeinerung des Topologischen Sphärensatzes. Um dies zu sehen, betrachten wir eine Mannigfaltigkeit  $(M, g)$ , deren Schnittkrümmungen im Intervall  $(1, 4]$  liegen. In diesem Fall gilt

$$|R(e_1, e_2, e_3, e_4)| \leq 2$$

für jedes orthonormale Vierbein  $\{e_1, e_2, e_3, e_4\} \subset T_p M$ . Daraus folgt

$$\begin{aligned} &R(e_1, e_3, e_1, e_3) + R(e_1, e_4, e_1, e_4) \\ &+ R(e_2, e_3, e_2, e_3) + R(e_2, e_4, e_2, e_4) \\ &- 2R(e_1, e_2, e_3, e_4) > 0. \end{aligned}$$

Jede solche Mannigfaltigkeit  $(M, g)$  hat daher positive isotrope Krümmung.

Der Begriff der positiven isotropen Krümmung ist motiviert durch Stabilitätsfragen für zwei-dimensionale Minimalflächen. Während im Beweis von Theorem 2.4 das Längenfunktional auf dem Raum der Kurven eine Schlüsselrolle spielt, beruht der Beweis von Theorem 2.6 auf Stabilitätsüberlegungen für minimale 2-Sphären.

### 3 Der Ricci-Fluß und der Differenzierbare Sphärensatz

Wir beschreiben nun eine andere Methode, die sich im Zusammenhang mit dem Sphärensatz als nützlich erwiesen hat. Diese Methode beruht auf einer Deformation der Riemannschen Metrik und wurde 1982 von Richard Hamilton eingeführt. Wir beschreiben zunächst den Begriff der Ricci-Krümmung. Dazu betrachten wir eine Riemannsche Mannigfaltigkeit  $(M, g)$  sowie einen Punkt  $p \in M$ . Man definiert dann eine symmetrische Bilinearform  $\text{Ric} : T_p M \times T_p M \rightarrow \mathbb{R}$  durch

$$\text{Ric}(X, Y) = \sum_{k=1}^n R(X, e_k, Y, e_k),$$

wobei  $R$  den Riemannschen Krümmungstensor bezeichnet und  $\{e_1, \dots, e_n\}$  eine Orthonormalbasis des Tangentialraums  $T_p M$  ist. Die Bilinearform  $\text{Ric} : T_p M \times T_p M \rightarrow \mathbb{R}$  wird als der Ricci-Tensor von  $(M, g)$  bezeichnet. Die Ricci-Krümmung hat eine geometrische Interpretation als ein Mittelwert von Schnittkrümmungen. Dazu betrachten wir einen Vektor  $0 \neq X \in T_p M$ . Der Quotient  $\frac{\text{Ric}(X, X)}{(n-1)|X|^2}$  kann dann interpretiert werden als der Mittelwert aller Schnittkrümmungen  $K(\pi)$  über alle zwei-dimensionalen Ebenen  $\pi \subset T_p M$  mit  $X \in T_p M$ .

Wir betrachten nun eine Mannigfaltigkeit  $M$  und eine Anfangsmetrik  $g_0$  auf  $M$ . Wir deformieren dann die Riemannsche Metrik durch die Evolutionsgleichung

$$\frac{\partial}{\partial t} g(t) = -2 \text{Ric}_{g(t)}, \quad (4)$$

wobei  $\text{Ric}_{g(t)}$  den Ricci-Tensor von  $g(t)$  bezeichnet (siehe [31, 32]). Die Evolutionsgleichung (4) kann als ein dynamisches System auf dem Raum der Riemannschen Metriken aufgefaßt werden. Vom analytischen Standpunkt aus betrachtet verhält sich (4) wie eine nichtlineare Wärmeleitungsgleichung. Um dies zu sehen, betrachten wir die Evolution des Krümmungstensors von  $(M, g(t))$ . Es wurde von Hamilton gezeigt, daß der Krümmungstensor eine Evolutionsgleichung der Form

$$D_t R = \Delta R + Q(R)$$

erfüllt, wobei  $\Delta$  den Laplace-Operator bezüglich der Metrik  $g(t)$  bezeichnet und  $Q(R)$  eine quadratische Funktion des Krümmungstensors ist.

Als einfaches Beispiel betrachten wir den Fall, daß  $g_0$  die Standardmetrik auf der  $n$ -dimensionalen Sphäre  $S^n$  ist. Die Lösung des Ricci-Flusses für diese Anfangsmetrik  $g_0$  ist gegeben durch  $g(t) = (1 - 2(n-1)t)g_0$  für  $t \in [0, \frac{1}{2(n-1)})$ . Mit anderen Worten,  $(M, g(t))$  ist isometrisch zu einer runden Sphäre vom Radius

$\sqrt{1 - 2(n-1)t}$ . Für  $t \rightarrow \frac{1}{2(n-1)}$  konvergiert der Radius gegen 0, so daß die Mannigfaltigkeit  $(M, g(t))$  auf einen Punkt kollabiert.

Für allgemeine Anfangsmetriken zeigte Hamilton, daß der Ricci-Fluß stets eine Lösung für kurze Zeit besitzt.

**Theorem 3.1** (R. Hamilton [31]) *Sei  $(M, g_0)$  eine kompakte Riemannsche Mannigfaltigkeit. Dann existiert eine reelle Zahl  $T > 0$  und eine Lösung  $g(t)$  des Ricci-Flusses, die auf dem Zeitintervall  $[0, T)$  definiert ist und die Anfangsbedingung  $g(0) = g_0$  erfüllt. Die Lösung  $g(t)$  ist durch die Anfangsmetrik eindeutig bestimmt.*

Die Frage der Kurzzeitexistenz ist nicht-trivial, da der Ricci-Fluß nur schwach parabolisch ist. Hamilton überwand dieses Hindernis mit Hilfe des Satzes über inverse Funktionen von Nash und Moser. DeTurck [22] gelang eine Vereinfachung des Beweises.

Wir diskutieren nun die globalen Eigenschaften des Ricci-Flusses. Dazu sei  $\{g(t) : t \in [0, T)\}$  eine maximale Lösung des Ricci-Flusses auf einer kompakten Mannigfaltigkeit  $M$ . Mit anderen Worten, die Lösung  $g(t)$  kann nicht über den Zeitpunkt  $T$  hinaus fortgesetzt werden. Nach einem Theorem von Hamilton [31] gilt dann  $T = \infty$  oder  $\sup_{M \times [0, T)} |R_{g(t)}| = \infty$ . Falls eine Singularität in endlicher Zeit auftritt ( $T < \infty$ ), muß der Krümmungstensor von  $(M, g(t))$  daher unbeschränkt sein. Šešum zeigte, daß im Fall einer Singularität in endlicher Zeit auch der Ricci-Tensor von  $(M, g(t))$  unbeschränkt sein muß (siehe [57]).

Von besonderem Interesse ist das asymptotische Verhalten des Ricci-Flusses für  $t \rightarrow T$ . In Dimension 2 bewies Hamilton das folgende Theorem:

**Theorem 3.2** (R. Hamilton [33]) *Sei  $M$  eine kompakte zwei-dimensionale Mannigfaltigkeit mit einer Riemannschen Metrik  $g_0$ . Weiter sei  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetrik  $g_0$ .*

- *Ist die Euler-Charakteristik von  $M$  negativ, dann ist  $T = \infty$  und die reskalierten Metriken  $\frac{1}{2t} g(t)$  konvergieren für  $t \rightarrow \infty$  gegen eine Metrik mit konstanter Gaußscher Krümmung  $-1$ .*
- *Ist die Euler-Charakteristik von  $M$  gleich 0, dann ist  $T = \infty$  und die Metriken  $g(t)$  konvergieren für  $t \rightarrow \infty$  gegen eine Metrik mit konstanter Gaußscher Krümmung 0.*
- *Hat  $g_0$  positive Gaußsche Krümmung, dann gilt*

$$T = \frac{\text{vol}(M, g_0)}{4\pi \chi(M)}.$$

*In diesem Fall konvergieren die reskalierten Metriken  $\frac{1}{2(T-t)} g(t)$  für  $t \rightarrow T$  gegen eine Metrik mit konstanter Gaußscher Krümmung 1.*

In Dimension 3 analysierte Hamilton [31] das asymptotische Verhalten des Ricci-Flusses für Anfangsmetriken mit positiver Ricci-Krümmung. Mit Hilfe des Maximumprinzips konnte Hamilton zeigen, daß positive Ricci-Krümmung unter dem Ricci-Fluß erhalten ist. Mit anderen Worten, ist  $\{g(t) : t \in [0, T)\}$  eine Lösung

des Ricci-Flusses auf einer drei-dimensionalen Mannigfaltigkeit  $M$  mit der Eigenschaft, daß  $(M, g_0)$  positive Ricci-Krümmung hat, dann hat  $(M, g(t))$  positive Ricci-Krümmung für alle  $t \in [0, T)$ .

**Theorem 3.3** (R. Hamilton [31]) *Sei  $(M, g_0)$  eine kompakte drei-dimensionale Mannigfaltigkeit mit positiver Ricci-Krümmung. Weiter sei  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetrik  $g_0$ . Dann gilt  $T < \infty$ , und die reskalierten Metriken  $\frac{1}{4(T-t)} g(t)$  konvergieren für  $t \rightarrow T$  gegen eine Metrik mit konstanter Schnittkrümmung 1.*

Jede drei-dimensionale Mannigfaltigkeit mit positiver Ricci-Krümmung besitzt also eine andere Metrik mit konstanter Schnittkrümmung 1. Damit erhält man das folgende topologische Ergebnis:

**Korollar 3.4** (R. Hamilton [31]) *Sei  $(M, g_0)$  eine kompakte drei-dimensionale Mannigfaltigkeit mit positiver Ricci-Krümmung. Dann ist  $M$  diffeomorph zu einer sphärischen Raumform  $S^3/\Gamma$ .*

Die Ricci-Fluß-Methode von Richard Hamilton ist von fundamentaler Bedeutung für viele neuere Entwicklungen in der Differentialgeometrie. In diesem Zusammenhang sind insbesondere zwei Themenkomplexe erwähnenswert:

Eine naheliegende Frage betrifft das globale Verhalten des Ricci-Flusses in Dimension 3 für beliebige Anfangsmetriken (d.h. ohne die Annahme der Positivität der Ricci-Krümmung). Dies erfordert ein genaues Verständnis der dabei auftretenden Singularitäten. Hamilton entwickelte eine Reihe neuer Techniken, um die Frage der Singularitätenbildung in Dimension 3 zu studieren (siehe [34–36]). In drei spektakulären Arbeiten gelang es Perelman, Lösungen des Ricci-Flusses über Singularitäten hinaus fortsetzen und auf diese Weise die Poincaré-Vermutung zu beweisen (siehe [51–53]). Wir gehen auf diese Entwicklungen an dieser Stelle nicht näher ein und verweisen stattdessen auf die Übersichtsartikel von Besson [9], Ecker [23] und Leeb [43], sowie auf die Monographie [48].

Eine andere interessante Frage betrifft die Konvergenztheorie in Dimension  $n \geq 4$ . Dazu nehmen wir an, daß die Anfangsmetrik  $g_0$  eine geeignete Krümmungsbedingung erfüllt. Das Ziel ist zu zeigen, daß der Ricci-Fluß nach geeigneter Reskalierung gegen eine Metrik mit konstanter Schnittkrümmung konvergiert. Eine Schwierigkeit liegt darin, daß der Ricci-Fluß in Dimension  $n \geq 4$  die Positivität der Ricci-Krümmung nicht erhält (siehe [45]). Es ist daher erforderlich, andere Krümmungsbedingungen zu finden, die unter dem Ricci-Fluß erhalten bleiben. Als Beispiel betrachten wir den Begriff der Positivität des Krümmungsoperators. Dazu sei  $(M, g)$  eine Riemannsche Mannigfaltigkeit, und  $p$  ein Punkt in  $M$ . Der Riemannsche Krümmungstensor ist dann eine Multilinearform  $R : T_p M \times T_p M \times T_p M \times T_p M \rightarrow \mathbb{R}$ . Aufgrund der Identität (1) kann der Riemannsche Krümmungstensor als symmetrische Bilinearform auf dem Raum der 2-Formen aufgefaßt werden. Dazu definiert man

$$R(\varphi, \psi) = \sum_{i,j,k,l=1}^n R_{ijkl} \varphi^{ij} \psi^{kl}$$

für  $\varphi, \psi \in \wedge^2 T_p M$ .

**Definition 3.5** Eine Riemannsche Mannigfaltigkeit  $(M, g)$  hat *nichtnegativen Krümmungsoperator*, falls für jede 2-Form  $0 \neq \varphi \in \wedge^2 T_p M$  die Ungleichung  $R(\varphi, \varphi) \geq 0$  gilt. Wenn die strikte Ungleichung gilt, sagen wir, daß  $(M, g)$  *positiven Krümmungsoperator* hat.

Hamilton zeigte in [32], daß die Positivität des Krümmungsoperators unter Ricci-Fluß erhalten ist. Hamilton bewies außerdem das folgende Konvergenztheorem in Dimension 4:

**Theorem 3.6** (R. Hamilton [32]) *Sei  $(M, g_0)$  eine kompakte vier-dimensionale Mannigfaltigkeit mit positivem Krümmungsoperator. Ferner sei  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetrik  $g_0$ . Dann ist  $T < \infty$ , und die reskalierten Metriken  $\frac{1}{6(T-t)} g(t)$  konvergieren für  $t \rightarrow T$  gegen eine Metrik mit konstanter Schnittkrümmung 1.*

Theorem 3.6 impliziert, daß eine kompakte vier-dimensionale Mannigfaltigkeit mit positivem Krümmungsoperator diffeomorph zu  $S^4$  oder  $\mathbb{R}P^4$  ist.

Das erste Konvergenzresultat in beliebiger Dimension wurde 1985 von Huisken [39] bewiesen (siehe auch [44, 50]). Der folgende Satz ist eine direkte Konsequenz von Huiskens Theorem:

**Theorem 3.7** (G. Huisken [39]) *Für jede ganze Zahl  $n \geq 4$  existiert eine reelle Zahl  $\delta(n) \in (0, 1)$ , so daß folgendes gilt: Sei  $(M, g_0)$  eine kompakte Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$  mit der Eigenschaft, daß für alle zwei-dimensionalen Ebenen  $\pi_1, \pi_2 \subset T_p M$  die Ungleichung*

$$0 < \delta(n) K(\pi_1) < K(\pi_2)$$

*gilt. Wir bezeichnen ferner mit  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetrik  $g_0$ . Dann ist  $T < \infty$ , und die reskalierten Metriken  $\frac{1}{2(n-1)(T-t)} g(t)$  konvergieren für  $t \rightarrow T$  gegen eine Metrik mit konstanter Schnittkrümmung 1.*

Theorem 3.7 erfordert keine globalen Krümmungsschranken. Es genügt, daß der Quotient von zwei beliebigen Schnittkrümmungen am Punkt  $p$  größer als  $\delta(n)$  ist. Eine solche Bedingung wird als punktweise Pinching-Bedingung bezeichnet.

Das folgende Ergebnis ist eine Verallgemeinerung von Theorem 3.6:

**Theorem 3.8** (C. Böhm, B. Wilking [10]) *Sei  $(M, g_0)$  eine kompakte Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$  mit positivem Krümmungsoperator. Weiter sei  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetrik  $g_0$ . Dann gilt  $T < \infty$ , und die reskalierten Metriken  $\frac{1}{2(n-1)(T-t)} g(t)$  konvergieren für  $t \rightarrow T$  gegen eine Metrik mit konstanter Schnittkrümmung 1.*

Die Bedingung, daß die Anfangsmetrik  $g_0$  Krümmungsoperator hat, ist in hohen Dimensionen sehr restriktiv.

Im Folgenden beschreiben wir einige Konvergenzergebnisse für den Ricci-Fluß, die wesentlich schwächere Krümmungsannahmen erfordern. Eine Schlüsselrolle kommt dabei dem Begriff der positiven isotropen Krümmung zu, der im vorangehenden Abschnitt diskutiert wurde. In [17] wurde gezeigt, daß positive isotrope Krümmung stets unter dem Ricci-Fluß erhalten ist (siehe auch [49]):

**Theorem 3.9** (S. Brendle, R. Schoen [17]) *Sei  $M$  eine kompakte Mannigfaltigkeit der Dimension  $n \geq 4$ , und sei  $\{g(t) : t \in [0, T)\}$  eine Lösung des Ricci-Flusses auf  $M$ . Wenn  $(M, g(0))$  nicht-negative isotrope Krümmung hat, dann hat  $(M, g(t))$  nicht-negative isotrope Krümmung für alle  $t \in [0, T)$ .*

Der Beweis von Theorem 3.9 verwendet das Maximumprinzip. Dazu betrachtet man eine Situation, in der das Minimum der isotropen Krümmung gleich 0 ist. Unter Ausnutzung spezieller algebraischer Identitäten kann man zeigen, daß der Ricci-Fluß das Minimum der isotropen Krümmung vergrößert.

Das folgende Theorem liefert eine Reihe weiterer Krümmungsbedingungen, die ebenfalls unter Ricci-Fluß erhalten sind (siehe [12, 17]).

**Theorem 3.10** *Sei  $M$  eine kompakte Mannigfaltigkeit der Dimension  $n \geq 4$ . Weiter sei  $\{g(t) : t \in [0, T)\}$  eine Lösung des Ricci-Flusses auf  $M$ . Schließlich sei  $S^2(1)$  eine zwei-dimensionale Sphäre mit konstanter Krümmung 1. Dann gilt:*

- Wenn  $(M, g(0)) \times \mathbb{R}$  nicht-negative isotrope Krümmung hat, dann hat  $(M, g(t)) \times \mathbb{R}$  nicht-negative isotrope Krümmung für alle  $t \in [0, T)$ .
- Wenn  $(M, g(0)) \times \mathbb{R}^2$  nicht-negative isotrope Krümmung hat, dann hat  $(M, g(t)) \times \mathbb{R}^2$  nicht-negative isotrope Krümmung für alle  $t \in [0, T)$ .
- Wenn  $(M, g(0)) \times S^2(1)$  nicht-negative isotrope Krümmung hat, dann hat das Produkt  $(M, g(t)) \times S^2(1)$  nicht-negative isotrope Krümmung für alle  $t \in [0, T)$ .

Mit Hilfe von Theorem 3.10 läßt sich das folgende Konvergenztheorem beweisen:

**Theorem 3.11** (S. Brendle [12]) *Sei  $(M, g_0)$  eine kompakte Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$  mit der Eigenschaft, daß das Produkt  $(M, g_0) \times \mathbb{R}$  positive isotrope Krümmung hat. Weiter sei  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetrik  $g_0$ . Dann gilt  $T < \infty$ , und die reskalierten Metriken  $\frac{1}{2(n-1)(T-t)} g(t)$  konvergieren für  $t \rightarrow T$  gegen eine Metrik mit konstanter Schnittkrümmung 1.*

Die Bedingung, daß  $(M, g_0) \times \mathbb{R}$  nicht-negative isotrope Krümmung hat, ist äquivalent zu der Aussage, daß für jedes orthonormale Vierbein  $\{e_1, e_2, e_3, e_4\} \subset T_p M$  und jede reelle Zahl  $\lambda \in [0, 1]$  die Ungleichung

$$\begin{aligned} & R(e_1, e_3, e_1, e_3) + \lambda^2 R(e_1, e_4, e_1, e_4) \\ & + R(e_2, e_3, e_2, e_3) + \lambda^2 R(e_2, e_4, e_2, e_4) \\ & - 2\lambda R(e_1, e_2, e_3, e_4) \geq 0 \end{aligned}$$

gilt.

Theorem 3.7 und Theorem 3.8 sind Spezialfälle von Theorem 3.11. Das folgende Ergebnis wurde zuerst in [17] bewiesen, und ist ebenfalls ein Spezialfall von Theorem 3.11:

**Korollar 3.12** (S. Brendle, R. Schoen [17]) *Sei  $(M, g_0)$  eine kompakte Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$ . Wir nehmen an, daß für alle zwei-dimensionalen Ebenen  $\pi_1, \pi_2 \subset T_p M$  die Ungleichung*

$$0 < K(\pi_1) < 4 K(\pi_2)$$

*gilt. Wir bezeichnen mit  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit der Anfangsmetric  $g_0$ . Dann ist  $T < \infty$ , und die reskalierten Metriken  $\frac{1}{2(n-1)(T-t)} g(t)$  konvergieren für  $t \rightarrow T$  gegen eine Metrik mit konstanter Schnittkrümmung 1.*

Eine unmittelbare Konsequenz von Korollar 3.12 ist der Differenzierbare Sphärensatz:

**Theorem 3.13** (S. Brendle, R. Schoen [17]) *Sei  $M$  eine kompakte Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$ . Wir nehmen an, daß für alle zwei-dimensionalen Ebenen  $\pi_1, \pi_2 \subset T_p M$  die Ungleichung*

$$0 < K(\pi_1) < 4 K(\pi_2)$$

*gilt. Dann ist  $M$  diffeomorph zu einer sphärischen Raumform  $S^n / \Gamma$ .*

Wir bemerken, daß Theorem 3.13 kein Analogon in negativer Krümmung hat. Gromov und Thurston [26] haben Beispiele kompakter Mannigfaltigkeiten konstruiert, deren Schnittkrümmungen im Intervall  $[-1, -1 + \varepsilon]$  liegen, die aber keine Metrik mit konstanter Schnittkrümmung  $-1$  besitzen. (Dabei ist  $\varepsilon$  eine positive reelle Zahl, die beliebig klein gewählt werden kann.)

Wir diskutieren schließlich noch zwei Starrheitssätze. Das folgende Ergebnis liefert eine Klassifikation aller Riemannschen Mannigfaltigkeiten mit der Eigenschaft, daß  $M \times \mathbb{R}$  nicht-negative isotrope Krümmung hat.

**Theorem 3.14** (S. Brendle [15]) *Sei  $(M, g_0)$  eine kompakte, einfache zusammenhängende Riemannsche Mannigfaltigkeit der Dimension  $n \geq 4$  mit der Eigenschaft, daß  $(M, g_0) \times \mathbb{R}$  nicht-negative isotrope Krümmung hat. Weiter sei  $\{g(t) : t \in [0, T)\}$  die maximale Lösung des Ricci-Flusses mit Anfangsmetric  $g_0$ . Dann gilt eine der folgenden Aussagen:*

- *$T < \infty$ , und die reskalierten Metriken  $\frac{1}{2(n-1)(T-t)} g(t)$  konvergieren gegen eine Metrik mit konstanter Schnittkrümmung 1.*
- *$(M, g_0)$  ist eine Kähler-Mannigfaltigkeit.*
- *$(M, g_0)$  ist isometrisch zu einem symmetrischen Raum.*
- *$(M, g_0)$  ist isometrisch zu einem Produkt von zwei Mannigfaltigkeiten niedrigerer Dimension.*

Der Beweis von Theorem 3.14 benutzt eine auf Bony [11] zurückgehende Version des strikten Maximumprinzips. Für jede reelle Zahl  $\tau \in (0, T)$  kann man zeigen, daß das Produkt  $(M, g(\tau)) \times \mathbb{R}$  positive isotrope Krümmung oder die Mannigfaltigkeit  $(M, g(\tau))$  nicht-generische Holonomiegruppe hat. Die möglichen Holonomiegruppen einfach zusammenhängender Riemannscher Mannigfaltigkeiten wurden 1955 von Berger klassifiziert (siehe [3] oder [8], Corollary 10.92).

Das folgende Ergebnis ist ein Spezialfall von Theorem 3.14:

**Korollar 3.15** (S. Brendle, R. Schoen [18]) *Sei  $(M, g_0)$  eine kompakte, einfach zusammenhängende Mannigfaltigkeit der Dimension  $n \geq 4$ . Wir nehmen an, daß für alle zwei-dimensionalen Ebenen  $\pi_1, \pi_2 \subset T_p M$  die Ungleichung*

$$0 \leq K(\pi_1) \leq 4 K(\pi_2)$$

*gilt. Wir bezeichnen mit  $\{g(t) : t \in [0, T]\}$  die maximale Lösung des Ricci-Flusses mit Anfangsmetrik  $g_0$ . Dann gilt eine der folgenden Aussagen:*

- *$T < \infty$ , und die reskalierten Metriken  $\frac{1}{2(n-1)(T-t)} g(t)$  konvergieren gegen eine Metrik mit konstanter Schnittkrümmung 1.*
- *$(M, g_0)$  ist isometrisch zu einem symmetrischen Raum.*

Die in diesem Abschnitt beschriebenen Methoden haben sich auch in anderem Zusammenhang als nützlich erwiesen. Unter anderem erhält man ein Starrheitsergebnis für Einstein-Mannigfaltigkeiten mit nicht-negativer isotroper Krümmung (siehe [13]), sowie eine Verallgemeinerung der wichtigen Harnack-Ungleichung von Richard Hamilton (siehe [14]). Andere Anwendungen findet man in [16, 30, 54] und [62].

## Literatur

1. Aloff, S., Wallach, N.: An infinite family of distinct 7-manifolds admitting positively curved Riemannian structures. *Bull. Am. Math. Soc.* **81**, 93–97 (1975)
2. Bérard-Bergery, L.: Les variétés riemanniennes homogènes simplement connexes de dimension impaire à courbure strictement positive. *J. Math. Pures Appl.* **55**, 47–67 (1976)
3. Berger, M.: Sur les groupes d’holonomie homogènes des variétés à connexion affine et des variétés riemanniennes. *Bull. Soc. Math. Fr.* **283**, 279–330 (1955)
4. Berger, M.: Les variétés Riemanniennes 1/4-pincées. *Ann. Sc. Norm. Super. Pisa, Cl. Sci.* **14**, 161–170 (1960)
5. Berger, M.: Sur quelques variétés riemanniennes suffisamment pincées. *Bull. Soc. Math. Fr.* **88**, 57–71 (1960)
6. Berger, M.: Les variétés riemanniennes homogènes normales simplement connexes à courbure strictement positive. *Ann. Sc. Norm. Super. Pisa, Sci. Fis. Mat.* **15**, 179–246 (1961)
7. Berger, M.: *A Panoramic View of Riemannian Geometry*. Springer, Berlin (2003)
8. Besse, A.: *Einstein Manifolds, Classics in Mathematics*. Springer, Berlin (2008)
9. Besson, G.: Preuve de la conjecture de Poincaré en déformant la métrique par la courbure de Ricci (d’après G. Perelman). *Astérisque* **307**, 309–347 (2006), Séminaire Bourbaki, Bd. 2004/2005
10. Böhm, C., Wilking, B.: Manifolds with positive curvature operator are space forms. *Ann. of Math.* **167**, 1079–1097 (2008)
11. Bony, J.M.: Principe du maximum, inégalité de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés. *Ann. Inst. Fourier* **19**, 277–304 (1969)



12. Brendle, S.: A general convergence result for the Ricci flow. *Duke Math. J.* **145**, 585–601 (2008)
13. Brendle, S.: A generalization of Hamilton's differential Harnack inequality for the Ricci flow. *J. Differ. Geom.* **82**, 207–227 (2009)
14. Brendle, S.: Einstein manifolds with nonnegative isotropic curvature are locally symmetric. *Duke Math. J.* **151**, 1–21 (2010)
15. Brendle, S.: *Ricci Flow and the Sphere Theorem*. Graduate Studies in Mathematics, Bd. 111. American Mathematical Society, Providence (2010)
16. Brendle, S., Huisken, G., Sinestrari, C.: Ancient solutions to the Ricci flow with pinched curvature. *Duke Math. J.* (im Erscheinen)
17. Brendle, S., Schoen, R.: Manifolds with  $1/4$ -pinched curvature are space forms. *J. Am. Math. Soc.* **22**, 287–307 (2009)
18. Brendle, S., Schoen, R.: Classification of manifolds with weakly  $1/4$ -pinched curvatures. *Acta Math.* **200**, 1–13 (2008)
19. Brendle, S., Schoen, R.: Curvature, sphere theorems, and the Ricci flow. *Bull. Am. Math. Soc.* **48**, 1–32 (2011)
20. Brieskorn, E.: Examples of singular normal complex spaces which are topological manifolds. *Proc. Natl. Acad. Sci. USA* **55**, 1395–1397 (1966)
21. Brieskorn, E.: Beispiele zur Differentialtopologie von Singularitäten. *Invent. Math.* **2**, 1–14 (1966)
22. DeTurck, D.: Deforming metrics in the direction of their Ricci tensors. *J. Differ. Geom.* **18**, 157–162 (1983)
23. Ecker, K.: Heat equations in geometry and topology. *Jahresber. Dtsch. Math.-Ver.* **110**, 117–141 (2008)
24. Freedman, M.H.: The topology of four-dimensional manifolds. *J. Differ. Geom.* **17**, 357–453 (1982)
25. Gromoll, D.: Differenzierbare Strukturen und Metriken positiver Krümmung auf Sphären. *Math. Ann.* **164**, 353–371 (1966)
26. Gromov, M., Thurston, W.: Pinching constants for hyperbolic manifolds. *Invent. Math.* **89**, 1–12 (1987)
27. Grove, K., Karcher, H., Ruh, E.: Group actions and curvature. *Invent. Math.* **23**, 31–48 (1974)
28. Grove, K., Karcher, H., Ruh, E.: Jacobi fields and Finsler metrics on compact Lie groups with an application to differentiable pinching problems. *Math. Ann.* **211**, 7–21 (1974)
29. Grove, K., Shiohama, K.: A generalized sphere theorem. *Ann. of Math.* **106**, 201–211 (1977)
30. Gu, H.L.: A new proof of Mok's generalized Frankel conjecture theorem. *Proc. Am. Math. Soc.* **137**, 1063–1068 (2009)
31. Hamilton, R.: Three-manifolds with positive Ricci curvature. *J. Differ. Geom.* **17**, 255–306 (1982)
32. Hamilton, R.: Four-manifolds with positive curvature operator. *J. Differ. Geom.* **24**, 153–179 (1986)
33. Hamilton, R.: The Ricci flow on surfaces. *Contemp. Math.* **71**, 237–262 (1988)
34. Hamilton, R.: The Harnack estimate for the Ricci flow. *J. Differ. Geom.* **37**, 225–243 (1993)
35. Hamilton, R.: Eternal solutions to the Ricci flow. *J. Differ. Geom.* **38**, 1–11 (1993)
36. Hamilton, R.: *The Formation of Singularities in the Ricci Flow*. Surveys in Differential Geometry, Bd. II, S. 7–136. International Press, Somerville (1995)
37. Hopf, H.: Zum Clifford-Kleinschen Raumproblem. *Math. Ann.* **95**, 313–339 (1926)
38. Hopf, H.: Differentialgeometrie und topologische Gestalt. *Jahresber. Dtsch. Math.-Ver.* **41**, 209–229 (1932)
39. Huisken, G.: Ricci deformation of the metric on a Riemannian manifold. *J. Differ. Geom.* **21**, 47–62 (1985)
40. Im Hof, H., Ruh, E.: An equivariant pinching theorem. *Comment. Math. Helv.* **50**, 389–401 (1975)
41. Kervaire, M., Milnor, J.: Groups of homotopy spheres, I. *Ann. of Math.* **77**, 504–537 (1963)
42. Klingenberg, W.: Über Riemannsche Mannigfaltigkeiten mit positiver Krümmung. *Comment. Math. Helv.* **35**, 47–54 (1961)
43. Leeb, B.: Geometrization of 3-dimensional manifolds and Ricci flow: on Perelman's proof of the conjectures of Poincaré and Thurston. *Boll. Unione Mat. Ital.* **1**(9), 41–55 (2008)
44. Margerin, C.: Pointwise pinched manifolds are space forms. In: *Geometric Measure Theory and the Calculus of Variations* (Arcata 1984). Proc. Sympos. Pure Math., Bd. 44, S. 307–328. Amer. Math. Soc., Providence (1986)
45. Máximo, D.: Non-negative Ricci curvature on closed manifolds under Ricci flow. *Proc. Am. Math. Soc.* **139**, 675–685 (2011)
46. Micallef, M., Moore, J.D.: Minimal two-spheres and the topology of manifolds with positive curvature on totally isotropic two-planes. *Ann. of Math.* **127**, 199–227 (1988)

47. Milnor, J.: On manifolds homeomorphic to the 7-sphere. *Ann. of Math.* **64**, 399–405 (1956)
48. Morgan, J., Tian, G.: Ricci flow and the Poincaré conjecture, Clay Mathematics Monographs, Bd. 3. Amer. Math. Soc., Providence (2007)
49. Nguyen, H.: Isotropic curvature and the Ricci flow. *Int. Math. Res. Not.* **3**, 536–558 (2010)
50. Nishikawa, S.: Deformation of Riemannian metrics and manifolds with bounded curvature ratios. In: *Geometric Measure Theory and the Calculus of Variations* (Arcata 1984). Proc. Sympos. Pure Math., Bd. 44, S. 343–352. Amer. Math. Soc., Providence (1986)
51. Perelman, G.: The entropy formula for the Ricci flow and its geometric applications. [arXiv:0211159](https://arxiv.org/abs/0211159)
52. Perelman, G.: Ricci flow with surgery on three-manifolds. [arXiv:0303109](https://arxiv.org/abs/0303109)
53. Perelman, G.: Finite extinction time for solutions to the Ricci flow on certain three-manifolds. [arXiv:0307245](https://arxiv.org/abs/0307245)
54. Petersen, P., Tao, T.: Classification of almost quarter-pinched manifolds. *Proc. Am. Math. Soc.* **137**, 2437–2440 (2009)
55. Rauch, H.E.: A contribution to differential geometry in the large. *Ann. of Math.* **54**, 38–55 (1951)
56. Ruh, E.: Krümmung und differenzierbare Struktur auf Sphären, II. *Math. Ann.* **205**, 113–129 (1973)
57. Šešum, N.: Curvature tensor under the Ricci flow. *Am. J. Math.* **127**, 1315–1324 (2005)
58. Smale, S.: Generalized Poincaré’s conjecture in dimensions greater than four. *Ann. of Math.* **74**, 391–406 (1961)
59. Sugimoto, M., Shiohama, K., Karcher, H.: On the differentiable pinching problem. *Math. Ann.* **195**, 1–16 (1971)
60. Tietze, H.: Über die topologischen Invarianten mehrdimensionaler Mannigfaltigkeiten. *Monatshefte Math. Phys.* **19**, 1–118 (1908)
61. Wolf, J.: *Spaces of Constant Curvature*. Publish or Perish, Houston (1984)
62. Xu, H.-W., Gu, J.-R.: An optimal differentiable sphere theorem for complete manifolds. *Math. Res. Lett.* **17**, 1111–1124 (2010)



**Simon Brendle** ist Professor für Mathematik an der Stanford Universität in Palo Alto. Er war eingeladener Redner auf dem Internationalen Mathematiker-Kongress in Madrid (2006) und hatte Gastprofessuren an der Eidgenössischen Technischen Hochschule in Zürich (2008) und der Universität Princeton (2010) inne. Sein wissenschaftliches Interesse gilt der Differentialgeometrie und der Theorie der partiellen Differentialgleichungen, sowie den Querverbindungen zwischen beiden Gebieten.



# From the Equivalence Principle to Market Consistent Valuation

Thomas Knispel · Gerhard Stahl · Stefan Weber

Received: 18 February 2011 / Published online: 17 May 2011  
© Vieweg+Teubner und Deutsche Mathematiker-Vereinigung 2011

**Abstract** Insurance companies are exposed to many different types of risk, in particular actuarial as well as financial risks. As a consequence, the classical actuarial principle of pooling does not provide a sufficient basis for the valuation and risk management of the total portfolio of an insurance company. Instead, the methodology needs to be complemented by modern financial mathematics that enables a market consistent valuation. The current article provides an introduction to the fundamental principles of financial mathematics that were originally developed by Fischer Black, Robert Merton and Myron Scholes in the beginning of the 1970s. We will discuss the relevance of these concepts for insurance firms in the context of internal models and the computation of the *market consistent embedded value* (MCEV).

**Keywords** Actuarial equivalence principle · Fundamental theorems of asset pricing · Market consistent valuation · Risk measures

**AMS Subject Classification** 62P05 · 91B24 · 91B25 · 91B30

## 1 Background

Modern financial mathematics—also known as *financial engineering*—is today an indispensable tool for the risk management of banks and insurance companies. A thor-

---

T. Knispel (✉) · S. Weber  
Institut für Mathematische Stochastik, Leibniz Universität Hannover, Welfengarten 1,  
30167 Hannover, Germany  
e-mail: [knispel@stochastik.uni-hannover.de](mailto:knispel@stochastik.uni-hannover.de)

S. Weber  
e-mail: [sweber@stochastik.uni-hannover.de](mailto:sweber@stochastik.uni-hannover.de)

G. Stahl  
Talanx AG, Quantitative Risk Management, Riethorst 2, 30659 Hannover, Germany  
e-mail: [gerhard.stahl@talax.com](mailto:gerhard.stahl@talax.com)

ough quantitative analysis is a necessary prerequisite for the valuation of financial instruments and portfolios, the construction of optimal investment strategies and the design of products. Insurance companies and banks deal with uncertain future events and cash flows. Actuarial and financial mathematics, thus, provide mathematical models for structures and patterns on insurance and financial markets that are driven by randomness.

An important link between economic questions and mathematical techniques was developed at MIT by Fischer Black (\*1938–†1995), Myron S. Scholes (\*1941) and Robert C. Merton (\*1944) in the beginning of the 1970s. Their universal ideas how financial options and derivatives must be priced are not limited to the classical *Black-Scholes model*, but can also be applied to more general models of financial markets. In 1997 Merton and Scholes received the Nobel prize for their work. The seminal papers [5, 35] triggered new quantitative techniques and products, and provided the foundation for modern financial mathematics, an applied field that combines methods from stochastic analysis, differential equations, functional analysis and numerics.

In contrast to financial mathematics, traditional actuarial mathematics does not have the reputation of being the most exciting field of applied mathematics. This dichotomy is not only apparent in academia, but also in industry practice. While *high-tech quants*—also known as *rocket scientists*—have been playing a key role in investment banking since the middle of the 1990s, insurance companies relied on a more conservative and traditional methodology. However, the insurance business has been experiencing significant changes for the last ten years.

Insurance companies are not only exposed to traditional actuarial risks that can be handled quite easily, but struggle at the same time with systematic risks that need to be priced and hedged. In particular, insurance companies are confronted with financial risk that cannot be understood on the basis of actuarial techniques. Modern insurance mathematics needs to integrate financial mathematics. A prime example provides the notion of *market consistent embedded value* that has been discussed in the context of *Solvency II*<sup>1</sup> and that can only be understood within the framework of ‘Black-Scholes’.

*High-tech quants* have been joining the insurance industry employing techniques from financial engineering. These enable two main applications. Firstly, the total portfolios of insurance companies can be priced consistently with market data. Secondly, complex insurance products with exposure to financial risk and embedded options can be designed, valued and hedged. Like financial derivatives, these structures provide investment tools that can be used for hedging or speculation. Examples include catastrophe bonds (‘CAT-bonds’) and variable annuities. New products offer new opportunities for companies and investors: insurance firms can mitigate risks to financial market participants which contributes to their risk management strategies; investors, on the other hand, profit from a broader spectrum of instruments which can be used to better diversify and further optimize their portfolios.

In view of the recent financial crisis, the conditions under which these positive effects do indeed materialize must, of course, be characterized very carefully.

---

<sup>1</sup>Solvency II is the updated regulatory framework for the insurance industry in the European Union. Its key aim is to ensure the liquidity and solvency of insurance companies.

Complex products did indeed contribute to the crisis, but are they primarily responsible?

This question has frequently been discussed recently, but the answer is all but simple. Among the key words that were mentioned in this context are, for example, insufficient due diligence, high leverage, inadequate bonus systems without liability, Ponzi schemes a la Madoff, ignorance, fraud, short-sighted interventions from politics and central banks, etc. Math was sometimes also blamed: Markets + Math = Mayhem?<sup>2</sup> However, the truth is that a complex world cannot be understood without adequate models. The random nature of market movements forces us to employ probabilistic models, if we want to partially understand the reality of the banking and insurance business. Without math, there will be no understanding: Markets – Math = Mayhem!

The current article reviews the key concepts of financial mathematics. We will explain how economic questions can be translated into proper mathematics and how financial engineering and insurance mathematics are linked to each other.

## 2 Principles of Financial Engineering

### 2.1 Classical Actuarial Mathematics vs. Financial Mathematics

Classical actuarial mathematics relies on the key principle of the insurance business—the principle of pooling. The exposure of individuals to idiosyncratic risks is transferred to a community of insured members that bear all risks together; as a consequence, individuals can be covered under insurance policies for fixed premiums that are small compared to possible individual claim sizes. The total claim should, however, be covered by the sum of the premia.

A particularly simple situation occurs if insurance claims are independent and no systematic risk components are involved. In this case, classical valuation methods can be applied very easily. Consider, for example, a random payment stream  $C_0, C_1, \dots, C_n$  at dates  $0, 1, \dots, n$  associated with a pension or a pure life insurance. With deterministic interest rate  $r \geq 0$ , its value at time 0 is given as the expected present value  $PV_0$ , defined as the expectation of the sum of the discounted payoffs  $(1+r)^{-t}C_t, t = 0, 1, \dots, n$ , i.e.,

$$PV_0(C_0, C_1, \dots, C_n) = E \left[ \sum_{t=0}^n \frac{1}{(1+r)^t} C_t \right].$$

This valuation principle provides the conceptual basis for the calculation of premiums in classical life insurance, as postulated by the *equivalence principle*:

<b>expected present value of premiums = expected present value of benefits</b>
--

<sup>2</sup>Financial Times, March 21, 2009.

A key feature of the classical insurance valuation principle consists in the computation of all expectations under the *statistical probability measure* that models the relative frequency with which events occur. In the context of life insurance mathematics, quantities like death and survival probabilities are, for example, estimated by the fraction of the dead and the survivors in the total population of each age group. In case of perfect *pooling of risks*, a valuation based on statistical means is mathematically justified by the *law of large numbers*.

The following example outlines the computation of single net premiums for pure life insurance policies. For a comprehensive introduction to life insurance mathematics we refer to the textbooks by Gerber [22], Milbrodt and Helbig [36] and Koller [31].

*Example 2.1* Consider a pure life insurance contract of a male insured of age  $x$  with a term of ten years that promises the death benefit of 100,000 € to be paid retrospectively at the end of the year of the insuree's death. Denoting by  $T_x$  the random remaining lifetime of the man, the insurance benefits are described by the payment stream

$$C_0 = 0 \quad \text{and} \quad C_t = \begin{cases} 100,000 & \text{if } t-1 < T_x \leq t, \\ 0 & \text{otherwise,} \end{cases} \quad t = 1, \dots, 10.$$

The expected present value of the benefits is thus given by

$$PV_0(C_0, C_1, \dots, C_{10}) = 100,000 \cdot \sum_{t=1}^{10} \frac{1}{(1+r)^t} P[t-1 < T_x \leq t], \quad (1)$$

where  $P[t-1 < T_x \leq t]$  signifies the probability of dying between  $t-1$  and  $t$ . Under a stationarity hypothesis these expressions can be rewritten in terms of the one-year death probabilities  $q_z$  and survival probabilities  $p_z$  of male policy holders aged  $z$ :

$$P[t-1 < T_x \leq t] = q_{x+t-1} \prod_{j=0}^{t-2} p_{x+j}.$$

The one-year statistical probabilities are listed in mortality tables, and for a given technical interest rate  $r$  the expected present value of the benefits  $PV_0$  can easily be computed from (1). The insurance premium is implied by the equivalence principle which states that the present value of the benefits equals the single net premium.

The principle of pooling is applicable to the valuation of products, if loss events occur independently. Systematic risks, however, alter the picture and require a more complex analysis that cannot solely be based on the equivalence principle. Both modern insurance products like variable annuities and the total portfolio of insurance firms are both exposed to financial as well as *systematic* actuarial risk components.

In the case of systematic risks, the classical actuarial equivalence principle needs to be replaced by the *principle of risk-neutral valuation*. Risk-neutral valuation

is not merely an ad-hoc method, but a mathematical consequence of specific axiomatic assumptions in the context of market models. The key hypothesis states that efficient markets do not admit any risk-free profits ('no free lunch') excluding arbitrage strategies, i.e., trading strategies that yield profits without any downside risk. Conceptually, a rationale to the assumption of the absence of arbitrage is provided by the market forces of supply and demand. A 'free lunch' would trigger a high demand for the profit opportunity on markets with profit-orientated participants and affect prices adversely. As a consequence, the arbitrage would vanish very quickly.

Formally, *risk-neutral valuation* resembles the equivalence principle. Prices are computed as expectation of future payments, i.e., *as if* market participants were risk-neutral. This *formal* observation is the reason why the valuation procedure is called the principle of *risk-neutral valuation*. However, risk-neutral and actuarial pricing are substantially different concepts. Risk-neutral valuation weights scenarios differently when expectations are computed. The statistical measure (of actuarial mathematics) must be replaced by a *technical* probability measure which is called *risk-neutral measure, martingale measure, or pricing measure*.

*Risk-neutrality* refers to the form of the valuation procedure only, but does not imply that economic agents are assumed to be risk-neutral. In fact, the risk-neutral measure encodes the risk-aversion of the market which is reflected by the modified weights that are assigned to scenarios. Market consistent valuation of modern insurance products and total portfolios of firms requires an integrated approach that combines techniques from financial engineering as well as actuarial mathematics.

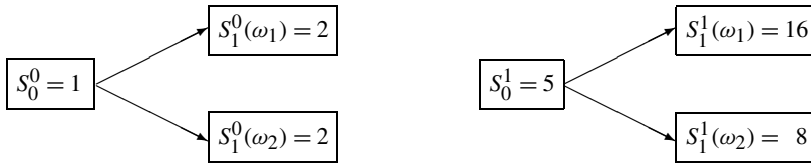
## 2.2 A Simple One-period Model with Two Scenarios

What are the basic concepts of financial mathematics? And how can the valuation principle be derived from basic axioms? A simple exposition to the answers can be provided in the context of elementary one-period models. The key ideas, however, generalize to advanced multi-period models.

In a one-period model, time is discrete with two dates, say  $t = 0, 1$ . Time 0 is interpreted as today, time  $t = 1$  as the future. Today's market prices are observable, but typically not future prices which are modeled as random. In the simplest case, the world is described by only two scenarios labeled  $\omega_1$  and  $\omega_2$  with strictly positive probabilities  $P[\{\omega_1\}] > 0$  and  $P[\{\omega_2\}] > 0$  under the statistical measure; scenarios with probability 0 would be irrelevant. The set of all scenarios is typically denoted by  $\Omega = \{\omega_1, \omega_2\}$ . A probability space with two scenarios can be interpreted as a model of a random experiment in which a (not necessarily fair) coin is tossed: scenario  $\omega_1$  can e.g. be identified with the event 'head', scenario  $\omega_2$  with the event 'tail'.

In the simplest case, the market consists of two primary products, a savings account and a stock, whose initial prices are observable in the market. The asset prices at time 1 may depend on the realized scenario which is not revealed before time 1. Formally, we label the savings account with the security identification number '0', the stock with number '1'. With this convention,  $S_t^0, S_t^1$  correspond to the prices of the savings account and the stock at times  $t = 0, 1$ , respectively.

A specific one-period market model with two scenarios is provided by the following example:



In this case, the performance of the savings account does not depend on the future state of the world. However, it is not difficult to incorporate stochastic interest rates in the model.

Agents can buy or sell shares of the primary products. Positive amounts are called *long positions*, negative amounts *short positions*. A short position in the savings account corresponds to debt.

**Definition 2.2** (Trading strategy) In a one-period market model, a *trading strategy* is a vector  $\vartheta = (\vartheta^0, \vartheta^1) \in \mathbb{R}^2$  whose components correspond to the number of shares invested into the savings account and the stock, respectively. The initial value  $V_0^\vartheta$  of the strategy is given by

$$V_0^\vartheta := \vartheta^0 S_0^0 + \vartheta^1 S_0^1, \quad (2)$$

and its terminal value  $V_1^\vartheta$  by

$$V_1^\vartheta := \vartheta^0 S_1^0 + \vartheta^1 S_1^1. \quad (3)$$

The analysis of the model is particularly simple under the following standard assumptions: Products are liquidly tradable, all assets can be sold and bought for the same price ('no bid-ask spread'), agents can hold short and non integer positions, taxes and transaction costs are neglected. In practice, these aspects are, of course, highly relevant, and can adequately be taken into account in more complex models.

The key assumption of financial engineering is the **absence of arbitrage opportunities**:

**Definition 2.3** (Arbitrage) An *arbitrage opportunity* is a trading strategy  $\vartheta$  with the following properties:

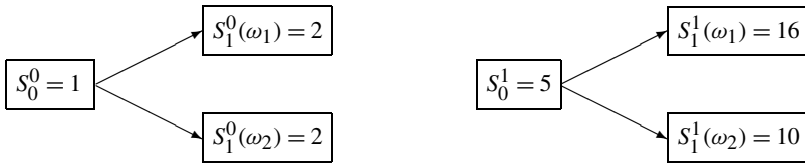
1. The cost of implementing  $\vartheta$  is zero, i.e.,  $V_0^\vartheta = 0$ .
2. Losses at time 1 are impossible, i.e.,  $V_1^\vartheta(\omega) \geq 0$  for all  $\omega \in \Omega$ .
3. The strategy  $\vartheta$  generates a profit with positive probability, i.e.,  $P[V_1^\vartheta > 0] > 0$ .

A financial market model is said to be *free of arbitrage*, if it does not admit arbitrage opportunities.

Financial market models are not always arbitrage-free. Whether or not a model is arbitrage-free, needs to be investigated. This is illustrated by the following example of a one-period model that permits arbitrage opportunities.



*Example 2.4* Suppose that the price dynamics of the primary products is specified as follows:



The stock systematically outperforms the savings account, since the terminal stock price is always at least twice its initial price. An agent with trading strategy  $\vartheta = (-5, 1)$  finances the purchase of one share of stock by debt. The strategy does not cost anything at time  $t = 0$ , but yields terminal wealth  $V_1^{\vartheta}(\omega_1) = 6 > 0$  resp.  $V_1^{\vartheta}(\omega_2) = 0$ . This is an arbitrage opportunity! The strategy—not involving any downside risk—is scalable and thus offers the chance of a profit of arbitrary size.

Financial engineering provides appropriate methods for the valuation and (partial) hedging of financial products, commonly called *contingent claims*. Contingent claims are contracts between two or more parties which specify the size of future payments to be exchanged between the parties conditional or contingent on future events. These describe both financial derivatives as well as insurance payments. In the first case, the size of payments may e.g. depend on the performance of reference products like stocks, bonds, currencies, commodities, and indexes. In the second case, payments are determined by the occurrence of loss events and the extent of insurance liabilities.

**Definition 2.5** (Contingent Claim) A *European contingent claim* with maturity 1 is a contract specifying the payments  $C_1(\omega)$  due at time 1 for all scenarios  $\omega \in \Omega$ . Mathematically, a contingent claim is described by a measurable mapping  $C_1 : \Omega \rightarrow [0, \infty)$ .

The basic examples are call and put options.

- A *European call option* on the stock with *maturity* 1 and *strike price*  $K$  is a financial contract that gives the contract holder the right to buy one share of the stock at time 1 for the fixed price  $K$  (without any obligation). This right will be exercised by a rational investor if and only if the stock price  $S_1^1$  exceeds  $K$ ; otherwise the strike price is higher than the spot price, and it is cheaper to buy the stock on the spot market. The random payoff of the call option at time 1 is given by

$$C_1 = \max\{S_1^1 - K, 0\}.$$

In practice, option contracts are frequently not settled physically, but in cash.

- A *European put option* on the stock with *maturity* 1 and *strike price*  $K$  entitles its owner to sell the stock at time 1 at the price of  $K$ , corresponding to a payoff of the form

$$C_1 = \max\{K - S_1^1, 0\}.$$

Two questions are immediate. What is the fair price  $C_0$  at time 0 of a contingent claim  $C_1$ ? How can the issuer of a claim manage and hedge the risks associated with the random payment  $C_1$ ? The answers to both questions are closely related and can be derived from the ‘no arbitrage’ hypothesis.

**Assumption 2.6** The price  $C_0$  at time  $t = 0$  of a contingent claim  $C_1$  maturing at time  $t = 1$  is fair if the extended market model specified by the savings account, the stock, and the contingent claim with price process  $(C_0, C_1)$  is free of arbitrage.

The notion of *fair* price is closely related to the *cost of perfect replication*.

**Definition 2.7** (Replicable Claim) A contingent claim  $C_1$  is *replicable* if there exists a trading strategy  $\vartheta = (\vartheta^0, \vartheta^1)$  such that its terminal value  $V_1^\vartheta$  coincides with the payoff  $C_1$  for all scenarios. In this case,  $\vartheta$  is called a *replicating trading strategy*, and the initial investment  $V_0^\vartheta$  is the cost of replication.

*Example 2.8* Let us consider a contingent claim with payoffs  $C_1(\omega_1) = 12$ ,  $C_1(\omega_2) = 4$ . A replicating strategy  $\vartheta$  has to satisfy  $V_1^\vartheta(\omega) = C_1(\omega)$  for each scenario  $\omega \in \Omega$ , thus (3) yields the following system of linear equations for  $\vartheta$ :

$$\vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega_1) = C_1(\omega_1), \quad (4a)$$

$$\vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega_2) = C_1(\omega_2). \quad (4b)$$

Inserting the specific asset prices of the model specified on page 144, the system takes the form

$$\vartheta^0 2 + \vartheta^1 16 = 12,$$

$$\vartheta^0 2 + \vartheta^1 8 = 4,$$

admitting the unique solution  $\vartheta^0 = -2$ ,  $\vartheta^1 = 1$ . The replicating strategy thus consists in buying one share of the stock and borrowing 2. The strategy  $\vartheta$  requires the initial capital

$$V_0^\vartheta = -2S_0^0 + 1S_0^1 = -2 \cdot 1 + 1 \cdot 5 = \underline{3},$$

and this amount is equal to the unique fair (i.e. arbitrage-free) price  $C_0$  at time  $t = 0$  of the contingent claim  $C_1$ .

How can this result be derived from the absence of arbitrage? Suppose that  $C_0 > V_0^\vartheta$ . In this case, an agent could sell at time 0 one share of the contingent claim  $C_1$  at a price of  $C_0$ , implement the replicating strategy  $\vartheta$  with cost  $V_0^\vartheta$ , and invest the remaining capital  $C_0 - V_0^\vartheta > 0$  into the savings account. This trading strategy does not cost anything at time 0. At maturity 1, the agent must pay  $C_1$ , the replicating strategy  $\vartheta$  generates the terminal wealth  $V_1^\vartheta = C_1$ , and the investment into the savings account yields  $S_1^0(C_0 - V_0^\vartheta)$ . The strategy thus provides the payoff  $S_1^0(C_0 - V_0^\vartheta) > 0$  at no cost. A price  $C_0 > V_0^\vartheta$  is, thus, not consistent with the absence of arbitrage. Analogously, an arbitrage opportunity can be constructed if the price  $C_0$  is strictly less than  $V_0^\vartheta$ .

The example illustrates a general valuation principle that applies to any *replicable* contingent claim  $C_1$ :

**arbitrage-free price  $C_0 = \text{cost of perfect replication!}$**

In Example 2.8, a replicating strategy is characterized by a system (4) of two linear equations and two unknowns. This system of equations admits a unique solution for any contingent claim  $C_1(\omega_1) = c_1$ ,  $C_1(\omega_2) = c_2$ , i.e., any contingent claim is replicable in the simple one-period model with two possible scenarios.

**Definition 2.9** (Complete Market Model) A financial market model is called *complete*, if all contingent claims can be replicated.

In more complex multi-period models the computation of a replicating strategy becomes more involved and requires, for example, the recursive solution of various systems of equations. The computations can be simplified by the application of martingale methods that are particularly powerful in the context of financial market models in continuous time, see, e.g., Musiela and Rutkowski [38].

The pricing of contingent claims can be separated from the computation of a replicating strategy via the principle of *risk-neutral valuation*. In one-period models this alternative approach does, of course, not provide any advantages; but it provides substantial simplifications in more complex market models. The key ideas can easily be illustrated in the context of one-period models.

Risk-neutral pricing resembles *formally* the classical actuarial pricing principle: the fair price is a (conditional) expectation of all future discounted payments. However, there are important differences. In contrast to classical actuarial mathematics, expectations have to be computed with respect to a *risk-neutral measure* rather than the statistical measure. At the same time, payments are discounted with respect to a reference product, the *numéraire*. This encodes that from an economic point of view only relative prices are meaningful, but not prices measured in absolute monetary units (e.g. € or DM). In the one-period model above, the savings account is taken as numéraire.<sup>3</sup>

**Definition 2.10** (Risk-Neutral Measure) A *risk-neutral measure* with respect to the numéraire  $S^0$  is a technical probability measure  $Q$  such that

1.  $Q[\{\omega\}] > 0$  for all  $\omega \in \Omega$ , and
2. the discounted price processes of all primary assets satisfy

$$\frac{S_0^i}{S_0^0} = E_Q \left[ \frac{S_1^i}{S_1^0} \right] := Q[\{\omega_1\}] \frac{S_1^i}{S_1^0}(\omega_1) + Q[\{\omega_2\}] \frac{S_1^i}{S_1^0}(\omega_2), \quad i = 0, 1. \quad (5)$$

<sup>3</sup>The choice of the numéraire is flexible. The savings account (or, more generally, money market account) constitutes only one possible alternative. The choice of the price process of zero coupon bonds as numéraire and a corresponding martingale measure, a *forward measure*, facilitates the efficient valuation of interest rate derivatives; cf., e.g., Filipović [15], Chap. 7.

Property 1 formalizes that each scenario  $\omega \in \Omega$  that possibly occurs according to the real-world measure  $P$  is also incorporated into the pricing (mathematically, this property is the equivalence of the probability measures  $P$  and  $Q$ ). Property 2 states that for all primary products—in the present setting these are the savings account and the stock—the discounted price at time 0 is given by the discounted prices at time 1 weighted by the  $Q$ -probabilities.<sup>4</sup> In the language of probability theory the discounted price processes are *martingales*<sup>5</sup> with respect to the measure  $Q$ , and for this reason  $Q$  is also called *equivalent martingale measure*.

In the one-period model above, (5) takes an explicit form

$$5 = Q[\{\omega_1\}]8 + Q[\{\omega_2\}]4.$$

Since  $Q[\{\omega_1\}] + Q[\{\omega_2\}] = 1$ , it follows that

$$Q[\{\omega_1\}] = \frac{1}{4} \quad \text{and} \quad Q[\{\omega_2\}] = \frac{3}{4}, \quad (6)$$

i.e., the arbitrage-free and complete simple market model admits a unique risk-neutral measure. This observation holds, in fact, more generally, and also the converse implication is true (see Sect. 2.4 and Sect. 3). Arbitrage-free prices of contingent claim can be computed using the risk-neutral measure  $Q$ . A replication problem does not need to be solved.

**Theorem 2.11** (Risk-neutral Valuation Formula, cf., e.g., [19], Corollary 1.35 and Theorem 1.32) *In the context of the complete financial market model above, the unique arbitrage-free price  $C_0$  at time 0 of a contingent claim with terminal payoff  $C_1$  is given by risk-neutral valuation with respect to the risk-neutral measure  $Q$ , i.e.,*

$$C_0 = S_0^0 E_Q \left[ \frac{C_1}{S_1^0} \right].$$

*In particular, the price computed with this formula is equal to the cost of perfect replication.*

The following example illustrates that the risk-neutral valuation formula is indeed consistent with the valuation via perfect replication.

**Example 2.12** Example 2.8 shows that the contingent claim  $C(\omega_1) = 12$ ,  $C(\omega_2) = 4$  can be replicated with initial investment  $V_0^\vartheta = 3$  and that the cost of replication corresponds to the unique arbitrage-free price  $C_0$ . Alternatively risk-neutral valuation

<sup>4</sup>In view of  $S_0^0/S_0^0 = S_1^0/S_1^0 = 1$  condition (5) is automatically satisfied for the numéraire product, the savings account. It thus remains to compute the weights  $Q[\{\omega_1\}]$  and  $Q[\{\omega_2\}]$  based on the discounted stock prices.

<sup>5</sup>Martingales are a synonym for ‘fair games’. Here ‘fair’ is used in the sense that the expected future value of the game equals the present value; in average there are no gains, but also no losses.

with the technical probabilities (6) yields

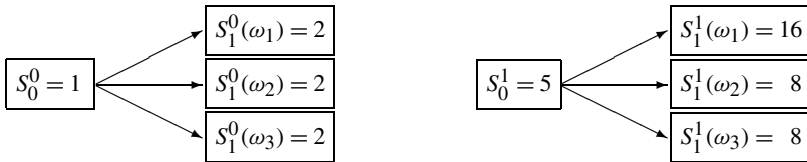
$$\begin{aligned}
 C_0 &= S_0^0 \cdot E_Q \left[ \frac{C}{S_1^0} \right] = S_0^0 \cdot Q[\{\omega_1\}] \frac{C_1(\omega_1)}{S_1^0} + Q[\{\omega_2\}] \frac{C_1(\omega_2)}{S_1^0} \\
 &= 1 \cdot \left( \frac{1}{4} \cdot \frac{12}{2} + \frac{3}{4} \cdot \frac{4}{2} \right) = 3.
 \end{aligned}$$

### 2.3 A Simple One-period Model with Three Scenarios

Reality is, however, more complex than the simple model above. Markets are typically incomplete, i.e., not every contingent claim can be replicated in terms of liquidly traded primary products. This fact is captured by incomplete market models. Pricing in incomplete markets is very different than in complete market models. The absence of arbitrage does not imply a unique price for contingent claims that cannot be replicated. Instead, ‘no free lunch’ is consistent with an interval of prices.

The main features of incomplete markets can easily be illustrated in the context of one-period models. For this purpose, we consider again a one-period model with the primary products ‘savings account’  $S^0$  and ‘stock’  $S^1$ . However, now we include three possible scenarios  $\omega_1, \omega_2, \omega_3$  that occur with strictly positive probability under the statistical measure  $P$ . Formally, this corresponds to the set of scenarios  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ .

The price dynamics of the primary products is modeled as follows:



The performance of the savings account in this example does not depend on the scenario  $\omega$  that is actually realized. The future stock price  $S_1^1$  is affected by the state of the world, but takes the same value if either scenario  $\omega_2$  or scenario  $\omega_3$  occurs. At first glance, the incorporation of a third scenario might seem like a minor modification only; however, it has serious consequences for the properties of the market model. The model with three scenarios is free of arbitrage, but not all contingent claims can be replicated.

*Example 2.13* The contingent claim with payoffs  $C_1(\omega_1) = 24$ ,  $C_1(\omega_2) = 16$ , and  $C_1(\omega_3) = 8$  is not replicable, i.e., there exists no trading strategy  $\vartheta = (\vartheta^0, \vartheta^1)$  with terminal wealth  $V_1^\vartheta = C_1$ . Mathematically, this simply means that the system

$$\begin{aligned}
 \vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega_1) &= C_1(\omega_1) \\
 \vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega_2) &= C_1(\omega_2) \\
 \vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega_3) &= C_1(\omega_3)
 \end{aligned}$$

of three linear equations with two unknowns does not have a solution  $(\vartheta^0, \vartheta^1)$ .

**Definition 2.14** (Incomplete Market Model) A financial market model that contains at least one non-replicable contingent claim is said to be *incomplete*.

A distinction between replicable and non-replicable contingent claims emerges immediately from Example 2.13. Replicable contingent claims possess terminal pay-offs of the form

$$C_1(\omega) = \vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega), \quad \omega \in \Omega,$$

and their unique fair price  $C_0$  coincides with the cost of perfect replication

$$V_0^\vartheta = \vartheta^0 S_0^0 + \vartheta^1 S_0^1. \quad (7)$$

Any other contingent claim is not replicable and cannot be priced on the basis of perfect replication. The principle of risk-neutral valuation, however, is still applicable, but its interpretation needs to be modified.

Let us first determine all risk-neutral measures in the sense of Definition 2.10: these are the technical probability measures  $Q$  such that

$$\frac{S_0^1}{S_0^0} = E_Q \left[ \frac{S_1^1}{S_1^0} \right] := Q[\{\omega_1\}] \frac{S_1^1}{S_1^0}(\omega_1) + Q[\{\omega_2\}] \frac{S_1^1}{S_1^0}(\omega_2) + Q[\{\omega_3\}] \frac{S_1^1}{S_1^0}(\omega_3). \quad (8)$$

Since  $Q[\{\omega_1\}] + Q[\{\omega_2\}] + Q[\{\omega_3\}] = 1$ , we obtain in the specific model that

$$\begin{aligned} 5 &= Q[\{\omega_1\}] \cdot \frac{16}{2} + Q[\{\omega_2\}] \cdot \frac{8}{2} + Q[\{\omega_3\}] \cdot \frac{8}{2} \\ &= Q[\{\omega_1\}] \cdot \frac{16}{2} + (1 - Q[\{\omega_1\}]) \cdot \frac{8}{2}. \end{aligned}$$

This implies  $Q[\{\omega_1\}] = 1/4$  and  $Q[\{\omega_2\}] + Q[\{\omega_3\}] = 3/4$ . Any choice  $Q[\{\omega_2\}] = \frac{3}{4}\alpha$ ,  $Q[\{\omega_3\}] = \frac{3}{4}(1 - \alpha)$  with  $\alpha \in (0, 1)$  defines a risk-neutral measure. The class of risk-neutral measures  $\mathcal{Q}$  consists of a continuum of elements:

$$\mathcal{Q} = \left\{ Q_\alpha : \begin{aligned} Q_\alpha[\{\omega_1\}] &= \frac{1}{4}, \quad Q_\alpha[\{\omega_2\}] = \frac{3}{4}\alpha, \quad Q_\alpha[\{\omega_3\}] = \frac{3}{4}(1 - \alpha) \\ &\text{with } \alpha \in (0, 1) \end{aligned} \right\}.$$

The existence of infinitely many risk-neutral measures characterizes incomplete market models.

It can be proven that for any given contingent claim with payoff  $C_1$  any risk-neutral measure  $Q_\alpha$  produces a price

$$C_{0,\alpha} := S_0^0 E_{Q_\alpha} \left[ \frac{C_1}{S_1^0} \right] \quad (9)$$

that is consistent with the absence of arbitrage. Conversely, any arbitrage-free price for the contingent claim can be written as in (9) for some  $Q_\alpha \in \mathcal{Q}$ . For a contingent claim the model thus admits a set of fair (i.e. arbitrage-free) prices, namely

$$\mathcal{C}_0 := \left\{ S_0^0 E_{Q_\alpha} \left[ \frac{C_1}{S_1^0} \right] \mid Q_\alpha \in \mathcal{Q} \right\}.$$

The lower and the upper value

$$C_0^{\text{inf}} := \inf_{Q \in \mathcal{Q}} S_0^0 E_Q \left[ \frac{C_1}{S_1^0} \right] \quad \text{and} \quad C_0^{\text{sup}} := \sup_{Q \in \mathcal{Q}} S_0^0 E_Q \left[ \frac{C_1}{S_1^0} \right]$$

are arbitrage bounds for the contingent claim with payoff  $C_1$ . The set  $\mathcal{C}_0$  is always either a singleton  $C_0^{\text{inf}} = C_0^{\text{sup}}$  or the open interval  $(C_0^{\text{inf}}, C_0^{\text{sup}})$ .

*Example 2.15 (Replicable Contingent Claim)* The contingent claim with terminal payoffs  $C_1(\omega_1) = 12$ ,  $C_1(\omega_2) = 4$ , and  $C_1(\omega_3) = 4$  can be replicated by buying one share of the stock and borrowing the amount of 2, corresponding to trading strategy  $\vartheta = (-2, 1)$ . According to (7), the cost of replication is  $V_0^\vartheta = 3$ . Alternatively, risk-neutral valuation with respect to an arbitrary risk-neutral measure yields

$$\begin{aligned} C_{0,\alpha} &= S_0^0 \left( Q_\alpha[\{\omega_1\}] \frac{C_1(\omega_1)}{S_1^0} + Q_\alpha[\{\omega_2\}] \frac{C_1(\omega_2)}{S_1^0} + Q_\alpha[\{\omega_3\}] \frac{C_1(\omega_3)}{S_1^0} \right) \\ &= 1 \left( \frac{1}{4} \cdot \frac{12}{2} + \frac{3}{4} \alpha \cdot \frac{4}{2} + \frac{3}{4} (1 - \alpha) \cdot \frac{4}{2} \right) = \underline{3}, \end{aligned}$$

and this price does not depend on the specific choice of a risk-neutral measure  $Q_\alpha$ ,  $\alpha \in (0, 1)$ . In other words, the class of arbitrage-free prices consists of exactly one element, namely the cost of perfect replication.

*Example 2.16 (Non-replicable Contingent Claim)* For the non-replicable contingent claim  $C_1(\omega_1) = 24$ ,  $C_1(\omega_2) = 16$ ,  $C_1(\omega_3) = 8$  introduced in Example 2.13 the arbitrage-free price computed in terms of the risk-neutral measure  $Q_\alpha$ ,  $\alpha \in (0, 1)$ , equals

$$\begin{aligned} C_{0,\alpha} &= S_0^0 \left( Q_\alpha[\{\omega_1\}] \frac{C_1(\omega_1)}{S_1^0} + Q_\alpha[\{\omega_2\}] \frac{C_1(\omega_2)}{S_1^0} + Q_\alpha[\{\omega_3\}] \frac{C_1(\omega_3)}{S_1^0} \right) \\ &= 1 \left( \frac{1}{4} \cdot \frac{24}{2} + \frac{3}{4} \alpha \cdot \frac{16}{2} + \frac{3}{4} (1 - \alpha) \cdot \frac{8}{2} \right) \\ &= \underline{6 + 3\alpha}. \end{aligned}$$

This price depends on the parameter  $\alpha$  resp. the measure  $Q_\alpha$  explicitly. In particular, the non-replicable claim  $C_1$  admits a whole class of arbitrage-free prices, namely the open interval  $\mathcal{C}_0 = (6, 9)$ .

Examples 2.15 and 2.16 illustrate the following dichotomy which generalizes to more complex incomplete financial market models.

**Theorem 2.17** (Arbitrage-free Prices, cf., e.g., [19], Corollary 1.35) *Let  $C_1$  be a contingent claim.*

1. *The contingent claim  $C_1$  can be replicated if and only if it admits a unique arbitrage-free price. In particular, this price is given by the cost of perfect replication.*
2. *If  $C_1$  is not replicable, then  $C_0^{\text{inf}} < C_0^{\text{sup}}$ , and the set of arbitrage-free price is the open interval*

$$C_0 = (C_0^{\text{inf}}, C_0^{\text{sup}}) = \left( \inf_{Q \in \mathcal{Q}} S_0^0 E_Q \left[ \frac{C_1}{S_1^0} \right], \sup_{Q \in \mathcal{Q}} S_0^0 E_Q \left[ \frac{C_1}{S_1^0} \right] \right).$$

Incomplete market models are characterized by the existence of contingent claims that cannot perfectly be replicated. While replicable claims can, by definition, be hedged, risk management of non-replicable claims is more sophisticated.

An extreme approach is *superhedging*.<sup>6</sup> A superhedge is a trading strategy  $\vartheta$  with minimal initial capital whose terminal value  $V_1^\vartheta$  dominates the payoff of the contingent claim for any possible scenario, i.e.,

$$V_1^\vartheta(\omega) = \vartheta^0 S_1^0 + \vartheta^1 S_1^1(\omega) \geq C_1(\omega) \quad \text{for all } \omega \in \Omega. \quad (10)$$

Superhedging excludes any downside risk at maturity. Superhedging strategies do indeed always exist. The minimal capital to set up the superhedging portfolio is given by the upper arbitrage bound  $C_0^{\text{sup}}$ , commonly labeled the *superhedging price*.

*Example 2.18* The contingent claim  $C_1(\omega_1) = 24$ ,  $C_1(\omega_2) = 16$ ,  $C_1(\omega_3) = 8$  in Example 2.13 is not replicable. According to Example 2.16, its superhedging price is  $C_0^{\text{sup}} = 9$ , i.e., the initial cost of a superhedging strategy  $\vartheta$  is given by

$$9 = V_0^\vartheta = \vartheta^0 S_0^0 + \vartheta^1 S_0^1.$$

This yields the relation  $\vartheta^0 = 9 - 5\vartheta^1$ , and the superhedging condition (10) provides a system of inequalities for the number of shares  $\vartheta_1$ :

$$(9 - 5\vartheta^1) \cdot 2 + \vartheta^1 \cdot 16 \geq 24,$$

$$(9 - 5\vartheta^1) \cdot 2 + \vartheta^1 \cdot 8 \geq 16,$$

$$(9 - 5\vartheta^1) \cdot 2 + \vartheta^1 \cdot 8 \geq 8.$$

The first and the second inequality imply  $\vartheta^1 \geq 1$  and  $\vartheta^1 \leq 1$ , and the third inequality is equivalent to  $\vartheta^1 \leq 5$ . Thus, the superhedging strategy consists in buying  $\vartheta^1 = 1$  shares of the stock and investing the amount  $\vartheta^0 = 9 - 5\vartheta^1 = 4$  into the savings account.

---

<sup>6</sup>For dynamic market models, seminal papers on superhedging are El Karoui and Quenez [12] and Kramkov [32].



In practice, superhedging strategies are usually not acceptable, since they are typically very expensive. Even in complete markets, agents might not follow replication strategies, but accept a certain amount of risk in return for potential profits. Partial hedges that require less capital and reduce risk at the same time provide sensible alternatives. One possible approach, the *quantile hedging* method, has been proposed by several authors in the 90s. Quantile hedging tolerates a positive probability (with respect to the statistical measure  $P$ ) that the payoff of the contingent claim is not covered at maturity, i.e., a probability of failure. Quantile hedging strategies can be constructed either by maximizing the probability of a successful hedge among all trading strategies, given a constraint on the required cost; or by minimizing the cost of the strategy, given a constraint on the probability of failure. A detailed discussion of quantile hedging for both complete and incomplete market models in continuous time can be found in Föllmer and Leukert [16].

Quantile hedging controls the probability of the hedging error, but it does not take into account the economically relevant size of the shortfall  $(C_1 - V_1^\vartheta)^+$  if it occurs. This deficiency motivates the construction of alternative methods like *efficient hedging* suggested by Föllmer and Leukert [17]. *Efficient hedging* measures the hedging error in terms of a loss function  $l$  that weights losses of different size. An economic agent can either minimize his expectation of the weighted shortfall

$$E_P[l((C_1 - V_1^\vartheta)^+)]$$

among all trading strategies  $\vartheta$ , given a constraint on his available capital; or, he might seek the trading strategy with minimal cost that does not exceed a fixed upper bound for the expected shortfall.

Extensions of quantile and efficient hedging incorporating the aspect of model uncertainty are discussed in Kirch [28]. Further references are, among others, Nakano [39], Cvitanic and Karatzas [9], Kirch and Runggaldier [29], Favero [13], Favero and Runggaldier [14], Schied [46, 47], Rudloff [43, 44], Sekine [48], and Klöppel and Schweizer [30]. The solution of the partial hedging problem relies in continuous time models, for example, on duality methods, stochastic control techniques, or backward stochastic differential equations (BSDE). A survey on this field and the related topic of robust portfolio optimization can be found in Föllmer, Schied and Weber [20].

#### 2.4 Risk-Neutral Valuation—The General Principles

The toy models above are, of course, too simple to capture the complex reality of financial markets. In practice, more realistic models are needed that include various risk factors, multiple primary products, and many trading times at which portfolio proportions can be readjusted. Such models are available both in discrete and continuous time. The classical Black-Scholes model in continuous time will be discussed in Sect. 3.

In discrete time models, the key results of asset pricing theory are summarized in the current section. We denote the random prices of  $d + 1$  primary financial products at times  $t = 0, 1, \dots, T$  by  $S_t^0, S_t^1, \dots, S_t^d$ . Product ‘0’ is typically an asset with strictly positive prices (e.g., a savings or money market account), playing the role of the numéraire.

Trading decisions at different times must be based on facts that are known at the specific date. Symbolically, the information that is available at time  $t$  is often denoted by  $\mathcal{F}_t$ ,  $t = 0, 1, \dots, T$ . The information sets  $\mathcal{F}_t$  are sigma-algebras that are interpreted as the classes of events (subsets of  $\Omega$ ) for which it is known at time  $t$  whether they have occurred or not. The increasing family of information sets  $\mathcal{F}_t$ ,  $t = 0, 1, \dots, T$ , is called information filtration.

In the multi-period case, a risk-neutral measure  $Q$  is determined (as extension of (5)) by the condition

$$\frac{S_t^i}{S_t^0} = E_Q \left[ \frac{S_{t+1}^i}{S_{t+1}^0} \middle| \mathcal{F}_t \right], \quad t = 0, 1, \dots, T-1, \quad i = 0, 1, \dots, d,$$

where  $E_Q[\cdot | \mathcal{F}_t]$  denotes the expectation under  $Q$  conditioned on the available information  $\mathcal{F}_t$ . This condition specifies that the discounted price processes of all primary products have to be martingales with respect to the risk-neutral resp. martingale measure.

Absence of arbitrage and market completeness can be described in terms of risk-neutral measures. These characterizations are provided by the *fundamental theorems of asset pricing*.

**Theorem 2.19** (1. Fundamental Theorem, cf., e.g., [19], Theorem 5.16)<sup>7</sup> *A market model is arbitrage-free if and only if there exists a risk-neutral measure.*

In the simple one-period models above, we verified that arbitrage free markets admit indeed risk-neutral measures. Conversely, the following example illustrates that for market models with arbitrage opportunities risk-neutral measures do not exist.

*Example 2.20* The market model in Example 2.4 admits arbitrage. Equation (5) has a unique solution  $Q[\{\omega_1\}] = 0$ ,  $Q[\{\omega_2\}] = 1$ , i.e., the measure  $Q$  assigns probability 0 to scenario  $\omega_1$ . This contradicts property 1 in Definition 2.10, thus  $Q$  does not define a risk-neutral measure.

As illustrated by the toy models above, there exist always either one or infinitely many risk-neutral measures. These two cases are in a one-to-one correspondence with the completeness and incompleteness of the underlying market model.

**Theorem 2.21** (2. Fundamental Theorem, cf., e.g., [19], Theorem 5.37) *The completeness of an arbitrage-free market model is equivalent to the uniqueness of the risk-neutral measure.*

---

<sup>7</sup>In continuous time market models the formulation of the fundamental theorems involves the general notion of equivalent local martingale measures (resp. equivalent sigma-martingale measures in the case of price processes that are not locally bounded). The existence of such a measure immediately implies the absence of arbitrage opportunities in the market model. In general, however, the converse does not hold. The existence of an equivalent local martingale measure (resp. sigma-martingale measures) requires a stronger condition, namely ‘no free lunch with vanishing risk’, cf. Delbaen and Schachermayer [10].

An arbitrage-free valuation of contingent claims can always be provided by risk-neutral measures.

**Theorem 2.22** (Risk-neutral Valuation, cf., e.g., [19], Theorem 5.32) *Let  $C_{T^*}$  be the payoff of a contingent claim with maturity  $T^* \leq T$ . Any risk-neutral measure  $Q$  defines an arbitrage-free price process for the contingent claim by*

$$C_t := S_t^0 E_Q \left[ \frac{C_{T^*}}{S_{T^*}^0} \middle| \mathcal{F}_t \right], \quad t \leq T^*.$$

In analogy to Theorem 2.17 the contingent claim  $C_{T^*}$  will be replicable if and only if there is exactly one arbitrage-free price  $C_0$  at time 0. This price corresponds to the cost of setting up a perfect hedge at time 0. Conversely, the contingent claim  $C_{T^*}$  admits no perfect hedge if the set of arbitrage-free prices forms a non-empty open interval.

An excellent survey of the *asset pricing theory* in discrete time is provided by the monograph of Föllmer and Schied [19]. A general discussion of the *fundamental theorems of asset pricing theory* that covers both the discrete time and the continuous time case with general semimartingale price processes can be found in the monograph of Delbaen and Schachermayer [10]. Applications in the context of life-insurance mathematics are presented in Møller and Steffensen [37].

### 3 The Black-Scholes Model

The fundamental concepts of asset pricing and hedging were originally discovered by Fischer Black, Robert C. Merton and Myron S. Scholes in the context of the famous Black-Scholes model in the beginning of the 1970s, see [5, 35]. Black-Scholes is a model in continuous time, but the same basic ideas that provide solutions in discrete time models can successfully be applied in the Black-Scholes model and other continuous time models. Black, Merton and Scholes developed their techniques within an asset model that was suggested by the MIT-economist Paul A. Samuelson<sup>8</sup> (\*1915–†2009) in 1965 as an extension of the Bachelier model, see [45]. The Bachelier model dates back to the Ph.D. dissertation [3] of Louis Bachelier from 1900.

Black-Scholes is a financial market model with two primary financial products: agents can either invest at a fixed interest rate into a money market account or into a stock. Time is continuous and corresponds to an interval  $[0, T]$  for some  $T > 0$ . As discussed on page 145 in a different context, the Black-Scholes model relies on certain simplifying assumptions that are also standard for many more complex models.

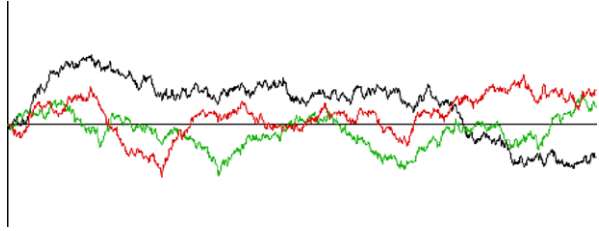
The constant interest rate of the money market account is  $r > 0$ . Thus, the dynamics of an investment of 1 is described by

$$S_t^0 = \exp(rt), \quad 0 \leq t \leq T.$$

---

<sup>8</sup>Nobel prize in economics, 1970.

**Fig. 1** Paths  $t \mapsto W_t(\omega)$  of a Brownian motion for different scenarios  $\omega$



Here, the upper index signifies the ‘security identification number’ which is equal to ‘0’ in the case of the money market account. The stock is the investment product with security identification number ‘1’ and price process  $(S_t^1)_{0 \leq t \leq T}$ , modeled by a *geometric Brownian motion*, a strictly positive stochastic process defined in terms of a standard Brownian motion, see (11).

Brownian motion is a key building block of the theory of stochastic processes in continuous time. It constitutes the driver of all random fluctuations in the model. The model itself specifies how the random innovations influence the movement of the objects in the model, e.g., the price processes. As illustrated in Fig. 1, Brownian motion is a specific stochastic process with rough paths that resemble the local behavior of price movements of stock indices.

Brownian motion has a very simple structure and is mathematically easily tractable, but it does not already provide a good model for asset price processes. Paths will, for example, take negative values with positive probability (= negative prices). In contrast to Samuelson in 1965, about sixty years earlier Bachelier used Brownian motion without any proper modification as a model for prices. In the Black-Scholes model Brownian motion is only a building block.

Brownian motion can be interpreted as a time-continuous random walk. A random walk  $(S_n)_{n=0,1,\dots}$  is a simple process in discrete time that is generated by the summation of independent, identically distributed random variables  $X_n$ ,  $n \in \mathbb{N}$ :

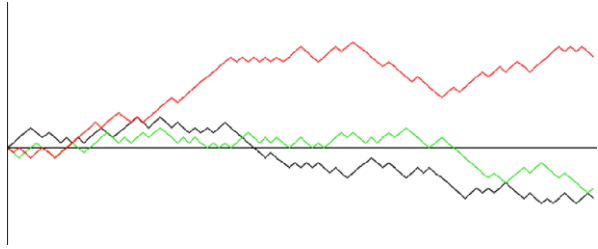
$$S_n := X_1 + \dots + X_n, \quad n = 0, 1, \dots$$

Like Brownian motion in continuous time, a random walk can be used as a basic building block for the construction of realistic financial market models in discrete time. Paths of a random walk are displayed in Fig. 2. For example, a symmetric random walk  $(S_n)_{n=0,1,\dots}$  on a probability space  $(\Omega, \mathcal{F}, P)$  with  $P[X_n = 1] = P[X_n = -1] = \frac{1}{2}$  can be generated by repeated coin tossing. At each point in time, the result of a coin toss determines whether the path of the process increases or decreases by 1. These two options correspond to the events ‘head’ and ‘tail’.

The relation between random walks and Brownian motion is described by the invariance principle [11] of Monroe David Donsker (\*1925–†1991). It provides a rationale why Brownian motion can be interpreted as a symmetric random walk in continuous time.

**Theorem 3.1** (Donsker’s Invariance Principle, Functional Central Limit Theorem, cf., e.g., [27], Theorem 14.9) *Suppose that  $(X_n)_{n \in \mathbb{N}}$  is a sequence of independent, identically distributed random variables on a probability space  $(\Omega, \mathcal{F}, P)$  with mean*

**Fig. 2** Paths  $n \mapsto S_n(\omega)$  of a random walk for different scenarios  $\omega$



$E[X_k] = 0$  and variance  $\text{Var}[X_k] = 1$ . Letting  $(S_n)_{n=0,1,\dots}$  be the corresponding random walk,

$$B_t^n := \frac{1}{\sqrt{n}}(S_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)X_{\lfloor nt \rfloor + 1}), \quad t \in [0, 1],$$

defines a sequence of stochastic processes in continuous time with continuous paths. The sequence  $B^n$  of random elements of  $C([0, 1])$  converges in law to a Brownian motion  $(W_t)_{t \in [0,1]}$ .

The value of the random walk in Fig. 2 changes by 1 on each unit time interval. Randomness with a frequency of 1 determines the movement of the process. The frequency of the innovations can be increased by decreasing the length of the time intervals. At the same time, the variance for fixed time horizons can be held constant, if simultaneously the size of the jumps is reduced. This procedure will generate a random walk that fluctuates up and down more quickly. At very high frequencies the constructed object looks very similar to a Brownian motion. Donsker’s Theorem states this relationship precisely: in the limiting case, when coins are tossed with infinite frequency, the rescaled random walk becomes a Brownian motion. Mathematically, Brownian motion can be defined as follows.

**Definition 3.2** A *Brownian motion* is a stochastic process  $W = (W_t)_{t \geq 0}$  with the following properties:

1. For  $P$ -almost all scenarios  $\omega \in \Omega$  the Brownian motion starts at 0, i.e.,  $W_0(\omega) = 0$ .
2. For all  $t > 0$  we have  $W_t \sim \mathcal{N}(0, t)$ , i.e.,

$$P[W_t \leq x] = \Phi\left(\frac{x}{\sqrt{t}}\right),$$

where  $\Phi$  denotes the cumulative distribution function of the standard normal distribution.

3. The increments are independent and stationary:
  - For all  $t < s$  the distribution of the increments  $W_s - W_t$  is equal to the distribution of  $W_{s-t}$ .
  - For all  $t_1 < t_2 < \dots < t_n$  the family  $W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$  is independent.

4. The paths  $t \rightarrow W_t(\omega)$  are continuous for almost all  $\omega \in \Omega$ .

The stock price process in the Black-Scholes model, a *geometric Brownian motion*, is a stochastic process with continuous paths that is driven by a Brownian motion  $W$ :

$$S_t^1(\omega) = S_0 \exp\left(\sigma W_t(\omega) + \left(\mu - \frac{1}{2}\sigma^2\right)t\right), \quad \omega \in \Omega. \quad (11)$$

This equation determines also the distribution of the stock price process  $S^1$  under the statistical measure  $P$ . Returns are normally distributed, prices log-normal. The parameter  $\sigma$  is called *volatility*; the parameter  $\mu$  *drift*. The volatility determines the impact of the random fluctuations of Brownian motion on the stock price dynamics; the drift specifies the size of the trend in the exponent.<sup>9</sup>

The information filtration  $\mathbb{F} = (\mathcal{F}_t)$  is generated by the observations of the stock price process and, thus, given by the Brownian filtration. As in the case of discrete time models, investments are described by self-financing strategies. In contrast to the discrete time case, in continuous time models investors can adjust the composition of their portfolio at any point in time. The analysis of the Black-Scholes model relies on Itô calculus that was invented by the Japanese mathematician Kiyoshi Itô [25] in the early 1940s.<sup>10</sup>

On the basis of these techniques, it can be shown that any derivative can be replicated by a self-financing strategy in the Black-Scholes model—the model is *complete*. Any derivative possesses a unique price that is solely implied by the absence of arbitrage: this price equals the cost of perfect replication. On the technical side, the concept of martingale measures generalizes to continuous time models like the Black-Scholes model.

The typical choice for the numéraire in the Black-Scholes model is the money market account. The construction of the unique equivalent martingale measure relies on techniques from *Stochastic Analysis*. Itô's formula implies a stochastic differential equation (SDE) for the dynamics of the discounted stock price process  $\tilde{S}_t^1 := S_t^1/S_t^0 = e^{-rt}S_t^1$ , i.e.,

$$d\tilde{S}_t^1 = \tilde{S}_t^1(\sigma dW_t + (\mu - r)dt) \quad (12)$$

<sup>9</sup>The infinitesimal dynamics of geometric Brownian motion is described by the *stochastic differential equation* (SDE)

$$dS_t^1 = S_t^1(\sigma dW_t + \mu dt)$$

that can be derived from (11) by the Itô-Döblin-formula. In contrast to classical calculus the deterministic drift equals  $\mu$  and not  $\mu - 1/2 \cdot \sigma^2$ —which might naively be conjectured from (11).

<sup>10</sup>Wolfgang Döblin (\*1915–†1940) obtained similar results in the late 1930s. His notes on the topic were, however, not discovered before the year 2000—sixty years after his death. Wolfgang Döblin, known in France as Vincent Döblin, escaped with his family from Nazi Germany to Paris. In World War II he became a French soldier and took his own life in the village Housseras, when his company surrendered, facing immediate capture. Wolfgang Döblin was the son of the writer Alfred Döblin (\*1878–†1957), best known for his novel 'Berlin Alexanderplatz' (1929). A literary portrait of Wolfgang and Alfred Döblin is provided in the a book by Marc Petit [41].

resp. in integral form

$$\tilde{S}_t^1 = \tilde{S}_0^1 + \int_0^t \tilde{S}_u^1 \sigma dW_u + \int_0^t \tilde{S}_u^1 (\mu - r) du.$$

The integral with respect to the Brownian motion is not a classical integral à la Stieltjes, since the paths of Brownian motion are ‘rough and wild’ and are, in particular, not of finite variation. Instead a stochastic integral, also called *Itô-integral*, is used which admits in its general form semimartingale-integrators. Itô-integrals with respect to Brownian motion are (local) martingales, and Itô’s Theorem of martingale representation states that, conversely, any continuous local martingale can be represented as an Itô-integral with respect to Brownian motion. Due to the SDE (12) of the discounted price process, an equivalent measure  $Q$  is a martingale measure if the stochastic process  $W_t^* := W_t + \frac{\mu-r}{\sigma}t$ ,  $t \in [0, T]$ , is a Brownian motion with respect to  $Q$ . The construction of this martingale measure relies on Girsanov’s Theorem.<sup>11</sup>

**Corollary 3.3** (Corollary of Girsanov’s Theorem for Brownian Motion, cf., e.g., [38], Proposition A.15.1) *For  $b \in \mathbb{R}$ , let  $P^*$  be a probability measure on  $(\Omega, \mathcal{F}_T)$  defined by the Radon-Nikodym density*

$$\frac{dP^*}{dP} := \exp\left(bW_T - \frac{1}{2}b^2T\right),$$

*i.e., for any event  $A \in \mathcal{F}_T$ ,*

$$P^*[A] = \int_A \frac{dP^*}{dP} dP.$$

*Then the process  $W_t^* := W_t - bt$ ,  $t \in [0, T]$ , is a Brownian motion with respect to the measure  $P^*$ .*

Thus, a martingale measure in the Black-Scholes model is defined by the Radon-Nikodym density

$$\frac{dQ}{dP} = \exp\left(-\frac{\mu-r}{\sigma}W_T - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2T\right).$$

A more general version of Girsanov’s Theorem implies that the equivalent martingale measure is indeed unique. As a consequence of versions of the fundamental theorems of asset pricing for continuous-time financial market models, the Black-Scholes model is free of arbitrage and the financial market is complete. Any derivative possesses a unique price—equal to the cost of perfect replication—that can be computed by risk-neutral valuation.

---

<sup>11</sup>Igor V. Girsanov (\*1934–†1965) was a Russian mathematician. Only four years after completing his Ph.D. in Moscow he died in an avalanche in the Sayan Mountains in southern Siberia. Versions of Girsanov’s Theorem for general semimartingales—as e.g. discussed in Protter [42]—allow the analysis of general financial market models.

**Theorem 3.4** (Risk-neutral Valuation, cf., e.g., [38], Corollary 3.3.1) *In the Black-Scholes model, the price of a financial derivative  $C_T$  with maturity  $T$  is given by*

$$C_0 = S_0^0 E_Q \left[ \frac{C_T}{S_T^0} \right] = e^{-rT} \cdot E_P \left[ \exp \left( -\frac{\mu - r}{\sigma} W_T - \frac{1}{2} \left( \frac{\mu - r}{\sigma} \right)^2 T \right) \cdot C_T \right]$$

which equals the cost at time 0 of setting up a perfect replication strategy of the claim  $C_T$ .

The Black-Scholes model is the simplest model for equity derivatives (besides the Bachelier model which admits negative prices). A pricing formula like the one in Theorem 3.4 can also be evaluated in more complex models by the application of Monte Carlo techniques. In the Black-Scholes model, prices of products like Vanilla European call and put options can explicitly be computed by integration with respect to a Gaussian density. Another approach for European, non path-dependent options is provided by the theory of partial differential equations. Particularly tractable is also the perfect replication of derivatives: the number of shares in the replicating portfolio is the delta of the option, the sensitivity of the option price with respect to the price of the stock (delta hedging).

The classical example of option valuation that was already discussed in the seminal papers by Black and Scholes [5] and Merton [35] is the pricing of European call options. As discussed before, a call option on a stock with maturity  $T$  and strike  $K$  is the right (but not the obligation) to buy one share of the stock at time  $T$  for the strike price  $K$ . A rational economic agent will execute this right, if and only if the stock price dominates the strike price at maturity. In this case, his immediate payoff at maturity is the difference of the spot price of the stock and the strike price at maturity  $T$ . The European call option is replicable (as are all derivatives in the Black-Scholes model), and its price can be determined by risk-neutral valuation according to Theorem 3.4.

**Corollary 3.5** (Black-Scholes Formula, cf., e.g., [38], Theorem 3.1.1) *In the Black-Scholes model, the price of a European call option  $C_T = \max\{S_T^1 - K, 0\}$  on a stock with maturity  $T$  and strike  $K$  is given by*

$$E_Q[e^{-rT} \max\{S_T^1 - K, 0\}] = S_0^1 \Phi(d_+(S_0^1, T)) - e^{-rT} K \Phi(d_-(S_0^1, T))$$

with constants

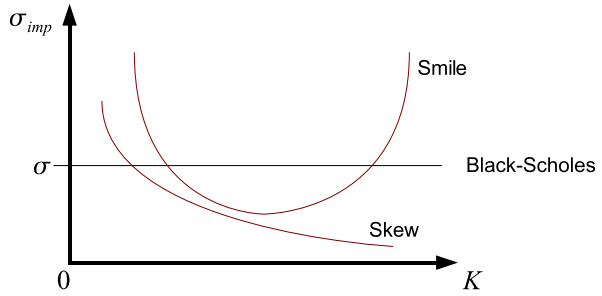
$$d_{\pm}(S_0^1, T) = \frac{\log(S_0^1/K) + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

Here  $\Phi$  denotes the distribution function of the standard normal distribution.

The Black-Scholes model is a very simple model that illustrates fundamental principles in the context of continuous time models. However, it does not provide a realistic description of market data.

Times series methods allow a detailed analysis of the statistical properties of financial market data. Returns are neither normally distributed nor independent as assumed by the Black-Scholes model. At the same time the volatility of the stock is not



**Fig. 3** Volatility Smile and Skew

constant, but fluctuates stochastically. The deficiencies of the Black-Scholes model are also apparent under the risk-neutral measure. Prices of Vanilla European call and put options are easily observable for many stocks, since these products are very liquidly traded. These options can easily be priced in the Black-Scholes model, if the five input parameters stock price, strike price, maturity, interest rate and stock volatility are known. If all input parameters but stock volatility are given, Black-Scholes option prices are described as a function of volatility. The *implied volatility*  $\sigma_{imp}$  of an option is defined as the volatility that produces a Black-Scholes price equal to the observed market price. Since various different options are traded, implied volatility can be computed for a collection of different strikes and maturities. As illustrated in Fig. 3, real market data show a dependence of implied volatility on maturity (*term structure*) and strike (*skew*). These observations contradict the assumptions of the Black-Scholes model with a constant stock volatility that determines the price of options of arbitrary strike and maturity. The implied volatility in a Black-Scholes world is a constant.

Over the last three decades various models for equity markets and many other asset classes have been developed in financial mathematics that provide a much better picture of reality than the Black-Scholes model. Extensions are, of course, often more complex and more difficult to handle. From a more abstract point of view, financial mathematics can be studied in the context of general semimartingale price processes. Since markets are usually incomplete, many products cannot be perfectly replicated and partial hedging strategies need to be constructed. For practical applications, specific models that can be estimated and calibrated have been designed: local and stochastic volatility models (Heston, Hull-White, SABR, . . .) for equity options; structural and intensity-based models for credit products; multi-asset-models with multiple correlated products; hybrid models for hybrid products, etc. (Musielà and Rutkowski [38], Brigo and Mercurio [7], Filipović [15], Bielecki and Rutkowski [4], Overhaus et al. [40]).

The more complex and more realistic models are still not a perfect image of the world, but at most a good approximation. At the same time, practitioners are faced with model risk. The choice of a deficient model might lead to misleading results and provide inadequate guidance in reality. A thoughtful application of financial mathematics in practice should always include a thorough analysis of model risk. A robust model approach that works well for specific applications must be the goal of any

quant in practice (cf. [6]):

**‘Essentially, all models are wrong, but some of them are useful.’**

#### 4 Market Consistent Valuation and Embedded Value

Financial mathematics a la Black & Scholes is not only highly relevant for banks and option traders, but also for insurance companies. New insurance products with exposure to financial risks have been developed that need to be priced and hedged. At the same time, on the level of the insurance companies balance sheet the valuation of assets and liabilities requires an integration of techniques from insurance and financial mathematics. The implementation of internal models under *Solvency II* and the computation of *market consistent embedded value* (MCEV) relies on the techniques that have been discussed in the previous sections.

How should a large portfolio be priced in a market consistent way? Letting  $(\Omega, \mathcal{F})$  be a measurable space that describes all relevant scenarios, all insurance and financial payments at a fixed time horizon (say  $t = 1$ ) are modeled as measurable functions on this space. The total value of the company’s portfolio of assets and liabilities is a measurable function  $X : \Omega \rightarrow \mathbb{R}$ . The portfolio  $X$  aggregates the company’s positions and provides an integrated view of assets and liabilities, capturing all financial and actuarial risks including embedded options and other derivative payments. In practice, scenarios and the corresponding portfolio values are usually produced by an *economic scenario generator*, a Monte Carlo simulation engine, that provides the backbone of market consistent valuation. The scenario generator together with suitable evaluation techniques is called *internal model*.

Valuation would be simple, if the real world was correctly described by Black-Scholes. The market of all systematic risks (equity, interest rates, credit, insurance risks) would be complete, a perfect model could easily be calibrated to market data. Unsystematic actuarial risks could be measured on the basis of the principle of risk pooling. The MCEV of the portfolio would be unique, and hedging as well as risk-management very simple.

Reality is, however, more complicated. Insurance companies are confronted with a complex world of incomplete markets, model uncertainty, and non-robust calibration algorithms. Perfect hedges are rarely available, and risk management of large portfolios is quite difficult. Since replicating strategies for portfolios are typically unavailable, the MCEV can usually not uniquely be determined. How can these problems be resolved by an insurance company that needs to value its portfolio and manage its risks?

A solution can be provided on the basis of risk measures which quantify unsystematic risks that cannot be perfectly hedged in the market. A detailed discussion of the concept of risk measures will be provided in Sect. 5. In incomplete markets, however, prices are not uniquely determined. Derivatives or portfolios possess a unique price in complete markets that equals the costs of perfect replication. The fair value of unsystematic insurance risks can approximately be obtained by the principle of pooling. A risk measure can be used to quantify the residual part.

We denote by  $\mathcal{X}$  the set of conceivable portfolios or subportfolios whose risk needs to be assessed. Mathematically, the family  $\mathcal{X}$  is a vector space of real-valued mappings  $X$  on  $(\Omega, \mathcal{F})$  that contains the constants. Negative values of  $X$  correspond to debt or losses. A risk measure  $\rho : \mathcal{X} \rightarrow \mathbb{R}$  quantifies the risk of the subportfolio that cannot be priced by the principle of pooling or perfect replication. In order to achieve consistency with the pricing principles implied by the absence of arbitrage, we impose the following requirement on the risk measure. If a position  $H$  can be perfectly replicated, its fair value  $H$  should be larger than the value  $-\rho(H)$  that is implied by its cost of capital. This insures that companies will compute the value of replicable positions by risk-free valuation.

An **algorithm for the computation of MCEV**<sup>12</sup> can be described as follows:

(a) Decompose  $X$  as  $X = H + Z$  where

- $H$  is replicable, and
- $Z$  denotes the residual.

(b) The MCEV is then computed by

$$\text{Value}(X) = H_0 - \rho(Z)$$

where

- $H_0$  signifies the cost of perfect replication of  $H$ , and
- $\rho(Z)$  is the cost of the risk capital for  $Z$ .

In the context of insurance portfolios,  $H$  is often called *valuation portfolio* and contains only risks that are liquidly traded. The position  $H$  must be constructed on the basis of the principle of pooling. A detailed introduction to valuation portfolios can be found in Bühlmann [8] and Wüthrich et al. [49].

From the point of view of the firm a large MCEV  $\text{Value}(X)$  is desirable. Equivalently, the costs of a hedge of  $X$  should be as small as possible, i.e., the sum of the cost  $-H_0$  of a perfect hedge of  $H$  and the capital costs  $\rho(Z)$  of  $Z$  should be minimized.

The **goal** of the decomposition of  $X$  into the summands  $(H, Z)$  is the maximization of the MCEV  $\text{Value}(X)$ , or, equivalently, the minimization of the cost of hedging. This optimization problem must be solved by the insurance company in cooperation with regulatory authorities and rating agencies. In practice, it might be reasonable to require that the solution to the problem satisfies additional constraints:

- (a) The value of the replicable part  $H$  might be required to be at least  $\alpha$ , i.e.  $H_0 \geq \alpha$ .
- (b) Alternative, one might require that the capital cost of the residual are at most  $\rho_{\max}$ , i.e.  $\rho(Z) \leq \rho_{\max}$ .

The optimization problem that determines the MCEV of a portfolio resembles the problem of robust *efficient hedging* ([9, 13, 14, 17, 28–30, 39, 43, 44, 46–48]) that was briefly discussed in Sect. 2.3. Implementing MCEV for large portfolios in practice is a highly complex problem.

---

<sup>12</sup>We would like to thank Stefan Jaschke for his inspiring talk in Hannover on the concept of MCEV (Common core in Swiss Solvency Test, Solvency II, IFRS phase II, CRO-Forum) [26].

**Table 1** Examples of the various risk categories

Replicable risks	Non replicable, quantifiable risk	Non quantifiable risks
Equity, Interest rates, FX Futures	Long-term interest rate risks	Structural breaks
Vanilla equity options	Long-term volatility risks	Business development
Swaps, Swaptions	Large insurance losses	Management
Variance swaps	Systematic insurance risks	Political developments
Predictable surrender		Longevity, long-term trends

The value of positions that can be replicated by simple strategies can directly be inferred from market data (*marking-to-market*). Complex structures, however, do frequently not admit a model-free construction of replication strategies. If perfect hedging strategies exist in reliable models, the risk-neutral value of the position can be computed from the model which needs to be calibrated to market data (*combination of marking-to-market and marking-to-model*). Non replicable risks that can quantitatively be reasonably modeled do not possess a unique risk-neutral value, but can be quantified using risk measures. A last category of non measurable risks requires a thorough qualitative analysis that must be conducted jointly by firms, regulatory authorities and rating agencies. The amount of such risk must already be constrained when portfolios are build up. At the same time, strategies must be devised how deal with extreme scenarios and how to contain potential damage. Examples of the various risk categories are provided in Table 1.

## 5 Risk Measures

The market consistent valuation of individual positions and portfolios whose value is not uniquely determined by replication or pooling is governed by the choice of a risk measure. The risk measure defines the cost of capital of non replicable and non diversifiable positions. Different choices will typically lead to different MCEVs and thus influence the size of the required equity capital of companies, an important determinant for solvency and stability. Regulatory authorities and rating agencies must therefore carefully evaluate the measures of the downside risk that are used in internal models.

The most commonly used risk measure in industry practice is *Value at Risk* (VaR):

**Definition 5.1** The *Value at Risk* at level  $\lambda \in (0, 1)$  of a financial position  $X$  is the smallest monetary amount that needs to be added to  $X$  such that the probability of a loss becomes smaller than  $\lambda$ :

$$\text{VaR}_\lambda(X) := \inf\{m \in \mathbb{R} : P[X + m < 0] \leq \lambda\}.$$

The parameter  $\lambda \in (0, 1)$  is typically a number close to 0.<sup>13</sup>

Value at Risk allows for a very simple interpretation and can at the same time be easily implemented in practice. Although popular among practitioners due to these reasons, Value at Risk has been heavily criticized in the academic literature as a risk measurement and management tool since the middle of the 1990s. VaR has two serious deficiencies: Firstly, VaR neglects extreme events that occur with small probability. Companies that measure downside risk by VaR thus fail to adequately prepare for extreme loss scenarios. In times of financial crisis, insufficient capital buffers of a large number of companies endanger the stability of the whole economy. Secondly, risk management tools should provide incentives for managers to spread risk and diversify the companies' portfolios. VaR, however, does not generally reward diversification, but charges a larger risk amount for a diversified position in many cases.

The discussion initially involved mainly academics, but spread to the financial service industry and the general public in the meantime. An article in the New York Times on January 2, 2009 entitled 'Risk Mismanagement' quotes the opinion of two practitioners:

"David Einhorn, who founded Greenlight Capital, a prominent hedge fund, wrote not long ago that VaR was '*relatively useless as a risk-management tool and potentially catastrophic when its use creates a false sense of security among senior managers and watchdogs. This is like an air bag that works all the time, except when you have a car accident.*' [...] Nicholas Taleb, the best-selling author of 'The Black Swan', has crusaded against VaR for more than a decade. He calls it, flatly, '*a fraud.*' "

VaR was also seriously criticized in the report of the British *Financial Service Authority* 2009, 'The Turner Review—A regulatory response to the global banking crisis' [1], named after their chairman Lord Adair Turner.

The academic literature investigates risk measures since the middle of the 1990s. An axiomatic theory of risk measures was initiated by a seminal paper of Artzner, Delbaen, Eber and Heath [2] that suggests replacing the *ad hoc* risk measure VaR by alternatives that are systematically constructed on the basis of sensible properties that should be incorporated in risk management procedures. An excellent overview can be found in a monograph by Föllmer and Schied [19].

A risk measure  $\rho : \mathcal{X} \rightarrow \mathbb{R}$  is a mapping that quantifies the risk of portfolios  $X : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$  in a family of portfolios  $\mathcal{X}$ . Mathematically,  $\mathcal{X}$  is a vector space of measurable functions that contains the constants.

**Definition 5.2** A mapping  $\rho : \mathcal{X} \rightarrow \mathbb{R}$  is a *monetary risk measure* if it satisfies the following properties:

1. **Monotonicity:** If  $X(\omega) \leq Y(\omega)$  for all scenarios  $\omega \in \Omega$ , then  $\rho(X) \geq \rho(Y)$ .
2. **Cash Invariance:** For any  $m \in \mathbb{R}$ ,  $\rho(X + m) = \rho(X) - m$ .

<sup>13</sup>Alternative notation is sometimes used for  $\text{VaR}_\lambda$ : instead of level  $\lambda$  the number  $1 - \lambda$  can be provided as the level;  $X$  is often replaced by  $-X$  in the formula. For example, *Value at Risk at level 99%* refers in some papers to the same quantity as *Value at Risk at level 1%* in the current paper.

Property 1 states that the risk of a position  $X$  is smaller than the risk of a position  $Y$ , if the future value of  $X$  is at least  $Y$  for any scenario. Property 2 states that risk is measured on a monetary scale: if  $m \in \mathbb{R}$  are added to  $X$ , then the risk of  $X$  is exactly reduced by this amount.

VaR is a risk measure according to the definition above. Other examples are provided by *Average Value at Risk*—also known as *Tail Value at Risk*, *Conditional Value at Risk* or *Expected Shortfall*—and *Utility-Based Shortfall Risk*.

*Example 5.3* Suppose that  $\mathcal{X}$  is the space  $L^\infty = L^\infty(\Omega, \mathcal{F}, P)$  of all essentially bounded random variables on the probability space  $(\Omega, \mathcal{F}, P)$ .

(a) *Average Value at Risk (AVaR)*:

For a financial position  $X$ , its AVaR $_\lambda$  at level  $\lambda \in (0, 1)$  is defined as

$$\text{AVaR}_\lambda(X) := \frac{1}{\lambda} \int_0^\lambda \text{VaR}_\alpha(X) d\alpha.$$

The risk of a financial position  $X$  measured by AVaR is always at least as large as risk measured by VaR, i.e.,

$$\text{AVaR}_\lambda(X) \geq \text{VaR}_\lambda(X).$$

More precisely, AVaR $_\lambda$  is the smallest distribution-based<sup>14</sup> convex<sup>15</sup> risk measure that is continuous from above and which dominates VaR $_\lambda$ .

(b) *Utility-based Shortfall Risk (UBSR)*:

UBSR with loss function  $l$  and loss level  $z$  is defined as

$$\text{UBSR}_z(X) := \inf\{m \in \mathbb{R} : E[l(-(X + m))] \leq z\}.$$

UBSR of a financial position  $X$  is the smallest monetary amount  $m$  that needs to be added to  $X$  such that the expected value of  $l(-(X + m))$  does not exceed the bound  $z$ .

All risk measures can be characterized by their acceptance sets: every risk measure is uniquely determined by the ‘acceptable positions’ with non-negative risk. The risk measure that corresponds to a given family of acceptable positions, the acceptance set, is equal to the smallest monetary amount that needs to be added to a given position to make it acceptable. In the case of VaR with level  $\lambda$ , a financial position is acceptable, if the probability of a loss is at most  $\lambda$ . The size of the losses and their distribution is neglected by VaR. These quantities are, however, highly relevant when the size of the loss scenario to the economy is appraised. The serious deficiencies of VaR are not shared by AVaR and UBSR. For example, UBSR defines those positions  $X$  as acceptable whose rescaled losses  $l(-X)$  do not exceed a level  $z$  in expectation. The size of the losses is thus explicitly measured.

<sup>14</sup>A risk measure  $\rho$  is *distribution-based*, if  $\rho(X)$  depends only on the distribution of  $X$ .

<sup>15</sup>*Convexity of risk measures* is defined in (13).

Another important aspect of risk management relates to incentives to properly diversify portfolios that should be provided by risk measurement procedures. Mathematically, this aspect can be formulated in terms of the (semi-)convexity of the risk measures. Suppose that  $X, Y \in \mathcal{X}$  are financial positions, then for any  $\alpha \in (0, 1)$  the convex combination  $\alpha X + (1 - \alpha)Y$  models a diversified position. A risk measure assesses diversification as positive, if the risk of the diversified position is at most as large as the maximum of the individual risks. Mathematically, this is described by the semiconvexity of  $\rho$ :

$$\rho(\alpha X + (1 - \alpha)Y) \leq \max\{\rho(X), \rho(Y)\}, \quad \alpha \in [0, 1].$$

The semiconvexity is equivalent to the convexity of the risk measure [18]:

$$\rho(\alpha X + (1 - \alpha)Y) \leq \alpha\rho(X) + (1 - \alpha)\rho(Y), \quad \alpha \in [0, 1]. \quad (13)$$

Convex risk measures can be characterized with the help of techniques from convex analysis via robust representation theorems that provide an interesting link to the aspect of model uncertainty (Föllmer and Schied [18], Föllmer, Schied and Weber [20]).

**Theorem 5.4** (Robust Representation, cf., e.g., [19], Theorem 4.33) *Letting  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\rho : L^\infty(\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$  be a risk measure, the following conditions are equivalent:*

- (a)  $\rho$  is convex and continuous from above.
- (b)  $\rho$  admits a robust representation in terms of all probability measures  $\mathcal{M}_1(P)$  on  $(\Omega, \mathcal{F})$  that are absolutely continuous with respect to  $P$ :

$$\rho(X) = \sup_{Q \in \mathcal{M}_1(P)} \{E_Q[-X] - \alpha(Q)\}, \quad X \in L^\infty. \quad (14)$$

Here,  $\alpha(Q) = \sup_{\{X \in L^\infty: \rho(X) \leq 0\}} E_Q[-X]$  ( $Q \in \mathcal{M}_1(P)$ ) denotes the minimal penalty function.

Condition (b) shows that the risk  $\rho$  of a position  $X$  is the *worst case* of the expected (and penalized) losses of  $X$  computed for all absolutely continuous probability measures; the function  $\alpha$  penalizes the probability measures according to their relevance (that is implied by the risk measure).

VaR is not a convex risk measure. Examples for convex risk measures that are also sensitive to extreme loss scenarios are *Average Value at Risk* and *Utility-based Shortfall Risk*.

#### Example 5.5

- (a) *Average Value at Risk (AVaR):*

The convex risk measure AVaR admits a simple robust representation:

$$\text{AVaR}_\lambda(X) = \max_{Q \in \mathcal{Q}_\lambda} E_Q[-X], \quad X \in L^\infty,$$

where the family  $\mathcal{Q}_\lambda$  contains all probability measures  $Q$  that are absolutely continuous with respect to  $P$  with Radon-Nikodym-densities that are  $P$ -a.s. bounded by  $1/\lambda$ , cf., e.g., [19], Theorem 4.52. The penalty function in (14) has the values  $\alpha(Q) = 0$  for  $Q \in \mathcal{Q}_\lambda$  and  $\alpha(Q) = \infty$  else, and the supremum is attained.

(a) *Utility-based Shortfall Risk (UBSR):*

The risk measure  $UBSR_z(X) = \inf\{m \in \mathbb{R} : E[l(-(X + m))] \leq z\}$  admits a robust representation (14) where the supremum is attained. The minimal penalty function can be computed as the solution of a one-dimensional minimization problem:

$$\alpha(Q) = \inf_{\lambda > 0} \frac{1}{\lambda} \left( z + E_P \left[ l^* \left( \lambda \frac{dQ}{dP} \right) \right] \right), \quad Q \in \mathcal{M}_1(P),$$

where  $l^*(u) = \sup_{x \in \mathbb{R}} (ux - l(x))$  denotes the Fenchel-Legendre-transform of  $l$ , cf., e.g., [19], Theorem 4.115. In the case of the entropic risk measure with exponential loss function  $l(x) = e^{\alpha x}$  the minimal penalty function can explicitly be computed and can be expressed in terms of the relative entropy with respect to  $P$ .

The risk measure AVaR is an important building stone in the family of distribution-based risk measures. The following representation theorem was discovered by Kusuoka [34] restricted to coherent convex risk measures and by Kunze [33] and Frittelli and Rosazza Gianin [21] in the general convex case.

**Theorem 5.6** (Robust Representation of Distribution-based Risk Measures, cf., e.g., [19], Theorem 4.62) *A convex risk measure  $\rho : L^\infty \rightarrow \mathbb{R}$  is distribution-based and continuous from above, if and only if*

$$\rho(X) = \sup_{\mu \in \mathcal{M}_1((0,1])} \left( \int_{(0,1]} AVaR_\lambda(X) \mu(d\lambda) - \beta(\mu) \right), \quad X \in L^\infty,$$

where  $\mathcal{M}_1((0, 1])$  denotes the set of probability measures on  $(0, 1]$  and

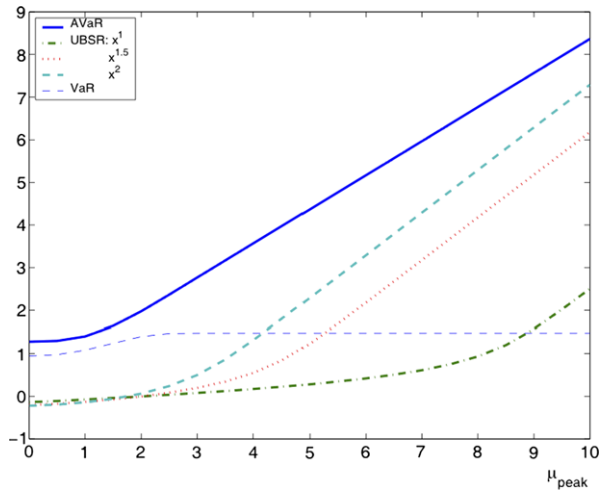
$$\beta(\mu) = \sup_{\{X \in L^\infty : \rho(X) \leq 0\}} \int_{(0,1]} AVaR_\lambda(X) \mu(d\lambda) \quad (\mu \in \mathcal{M}_1((0, 1])).$$

Apart from convexity, the sensitivity in detecting extreme loss scenarios is an important property of risk measures. Figure 4 illustrates the different sensitivities of VaR, AVaR and UBSR with polynomial loss functions  $l(x) = x^p$  with exponent  $p$ . A detailed analysis is provided by Giesecke, Schmidt and Weber [23]. Increasing the parameter  $\mu_{peak}$  increases the likelihood of extreme losses. The figure illustrates the low sensitivity of VaR and demonstrates that both AVaR as well as UBSR detect extreme risks very well.

From a theoretical point of view, AVaR and UBSR are superior to the classical risk measure VaR. The sensitivity of AVaR and UBSR to extreme losses entails, however, problems from a statistical point of view. AVaR and UBSR are distribution-based risk measures that depend on the distribution of  $X$  under the statistical measure. A robust



**Fig. 4** VaR<sub>0.05</sub>, AVaR<sub>0.05</sub> and UBSR with  $p \in \{1, \frac{3}{2}, 2\}$  and  $z = 0.3$  as a function of  $\mu$  for a mixture of a Student- $t$ -distribution (weight 0.96) and a normal distribution with mean  $\mu_{\text{peak}}$  (weight 0.04)



estimation of the tail of this distribution is usually not feasible.<sup>16</sup> This problem is inherited by any tail-sensitive risk measure. In practice, a purely data-driven analysis of extremal risks is impossible. Additional economic assumptions, a detailed analysis of the structure of portfolios and reasonable case studies can provide insights into the extremes. A thorough discussion with regulator authorities, transparency and a procedure that treats all market participants fairly is advisable. The choice of the risk measure should be subject to such a process. The MCEV can then be computed on a solid basis.

The modern world of financial and insurance companies is extremely complex. Interdisciplinary efforts are necessary to understand multiple aspects of the economy. Randomness is a key driver of the dynamics of prices and markets. A sensible analysis is nowadays impossible without sophisticated mathematical models. These provide in particular risk management strategies to minimize the likelihood of future financial crises.

**Acknowledgements** Useful remarks of two anonymous referees and the editor are gratefully acknowledged. A German version of the current paper will be published in the proceedings of the workshop *Die Folgen der Finanzkrise für Regulierung und Eigenkapital*, organized by Kompetenzzentrum Versicherungswissenschaften Hannover.

**References**

1. The Turner review—A regulatory response to the global banking crisis, published by the British Financial Services Authority, [www.fsa.gov.uk/pubs/other/turner\\_review.pdf](http://www.fsa.gov.uk/pubs/other/turner_review.pdf) (2009)
2. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Math. Finance* 9(3), 203–228 (1999)
3. Bachelier, L.: Théorie de la spéculation. Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics]. In: Sceaux, J.G. (ed.) *Théorie mathématique du jeu*. [Mathematical theory of games] (1995). Reprint of the 1900 original

<sup>16</sup>An interesting discussion of this problem and a suggestion for data-driven risk measures can be found in [24].

4. Bielecki, T.R., Rutkowski, M.: *Credit risk: modelling, valuation, and hedging*. Springer Finance. Springer, Berlin (2002)
5. Black, F., Scholes, M.S.: The pricing of options and corporate liabilities. *J. Polit. Econom.* **81**, 637–654 (1973)
6. Box, G.E.P., Draper, N.R.: *Empirical Model-building and Response Surfaces*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, New York (1987)
7. Brigo, D., Mercurio, F.: *Interest Rate Models—Theory and Practice*, 2nd edn. Springer Finance. Springer, Berlin (2006). With smile, inflation and credit
8. Bühlmann, H.: Multidimensional valuation. *Finance* **25**, 15–29 (2004)
9. Cvitanić, J., Karatzas, I.: Generalized Neyman-Pearson lemma via convex duality. *Bernoulli* **7**(1), 79–97 (2001)
10. Delbaen, F., Schachermayer, W.: *The Mathematics of Arbitrage*. Springer Finance. Springer, Berlin, (2006)
11. Donsker, M.D.: An invariance principle for certain probability limit theorems. *Mem. Am. Math. Soc.* **1951**(6), 12 (1951)
12. El Karoui, N., Quenez, M.-C.: Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM J. Control Optim.* **33**(1), 29–66 (1995)
13. Favero, G.: Shortfall risk minimization under model uncertainty in the binomial case: adaptive and robust approaches. *Math. Methods Oper. Res.* **53**(3), 493–503 (2001)
14. Favero, G., Runggaldier, W.J.: A robustness result for stochastic control. *Syst. Control Lett.* **46**(2), 91–97 (2002)
15. Filipović, D.: *Term-Structure Models*. Springer Finance. Springer, Berlin (2009). A graduate course
16. Föllmer, H., Leukert, P.: Quantile hedging. *Finance Stoch.* **3**(3), 251–273 (1999)
17. Föllmer, H., Leukert, P.: Efficient hedging: cost versus shortfall risk. *Finance Stoch.* **4**(2), 117–146 (2000)
18. Föllmer, H., Schied, A.: Convex measures of risk and trading constraints. *Finance Stoch.* **6**(4), 429–447 (2002)
19. Föllmer, H., Schied, A.: *Stochastic finance—an introduction in discrete time*, 3rd edn. Graduate Textbook Series. Walter de Gruyter, Berlin (2011)
20. Föllmer, H., Schied, A., Weber, S.: Robust preferences and robust portfolio choice. In: Bensoussan, A., Zhang (eds.) *Handbook of Numerical Analysis, Mathematical Modeling and Numerical Methods in Finance*, pp. 29–89 (2009)
21. Frittelli, M., Rosazza Gianin, E.: Law invariant convex risk measures. *Adv. Math. Econ.* **7**, 33–46 (2005)
22. Gerber, H.U.: *Life Insurance Mathematics*, 3rd edn. Springer, Berlin (1997)
23. Giesecke, K., Schmidt, T., Weber, S.: Measuring the risk of large losses. *J. Invest. Manag.* **6**(4), 1–15 (2008)
24. Heyde, C.C., Kou, S.G., Peng, X.H.: What is a good external risk measure: Bridging the gaps between robustness, subadditivity, and insurance risk measures. Preprint, Columbia University (2007)
25. Itô, K.: *Selected Papers*. Springer, New York (1987). Edited and with an introduction by S.R.S. Varadhan and D.W. Stroock
26. Jaschke, S.: Valuation and risk management of life reinsurance. In: *Talk, LUH-Kolloquium Versicherungs- und Finanzmathematik*, Hannover (2010)
27. Kallenberg, O.: *Foundations of modern probability*. Probability and its Applications (New York), 2nd edn. Springer, New York (2002)
28. Kirch, M.: Efficient hedging in incomplete markets under model uncertainty. PhD thesis, Humboldt-Universität zu Berlin. <http://edoc.hu-berlin.de/dissertationen/kirch-michael-2002-01-07/PDF/Kirch.pdf> (2001)
29. Kirch, M., Runggaldier, W.J.: Efficient hedging when asset prices follow a geometric Poisson process with unknown intensities. *SIAM J. Control Optim.* **43**(4), 1174–1195 (2004/05) (electronic)
30. Klöppel, S., Schweizer, M.: Dynamic indifference valuation via convex risk measures. *Math. Finance* **17**(4), 599–627 (2007)
31. Koller, M.: *Stochastische Modelle in der Lebensversicherung*, 2nd edn. Springer, Berlin (2010)
32. Kramkov, D.O.: Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probab. Theory Relat. Fields* **105**(4), 459–479 (1996)
33. Kunze, M.: Verteilungsinvariante konvexe Risikomaße. Diploma thesis, Humboldt-Universität zu Berlin (2003)
34. Kusuoka, S.: On law invariant coherent risk measures. *Adv. Math. Econ.* **3**, 83–95 (2001)

35. Merton, R.C.: Theory of rational option pricing. *Bell J. Econom. Manag. Sci.* **4**, 141–183 (1973)
36. Milbrodt, H., Helbig, M.: *Mathematische Methoden der Personenversicherung*. Walter de Gruyter, Berlin (1999)
37. Møller, T., Steffensen, M.: *Market-valuation methods in life and pension insurance*. International Series on Actuarial Science. Cambridge University Press, Cambridge (2007)
38. Musiela, M., Rutkowski, M.: *Martingale Methods in Financial Modelling*, 2nd edn. *Stochastic Modelling and Applied Probability*, vol. 36. Springer, Berlin (2005)
39. Nakano, Y.: Efficient hedging with coherent risk measures. *J. Math. Anal. Appl.* **293**(1), 345–354 (2004)
40. Overhaus, M., Bemudez, A., Bühler, H., Ferraris, A., Jordinson, C., Lamnour, A.: *Credit Hybrid Derivatives*. Wiley Finance. Wiley, New York (2007)
41. Petit, M.: *Die verlorene Gleichung-Auf den Spuren von Wolfgang und Alfred Döblin*. Eichborn Verlag (2005)
42. Protter, P.E.: *Stochastic integration and differential equations*, 2nd edn. *Stochastic Modelling and Applied Probability*, vol. 21. Springer, Berlin, (2005). Version 2.1, Corrected third printing
43. Rudloff, B.: Convex hedging in incomplete markets. *Appl. Math. Finance* **14**(5), 437–452 (2007)
44. Rudloff, B.: Coherent hedging in incomplete markets. *Quant. Finance* **9**(2), 197–206 (2009)
45. Samuelson, P.A.: Rational theory of warrant pricing. *Ind. Manage. Rev.* **6**(2), 13–39 (1951)
46. Schied, A.: On the Neyman-Pearson problem for law-invariant risk measures and robust utility functionals. *Ann. Appl. Probab.* **14**(3), 1398–1423 (2004)
47. Schied, A.: Risk measures and robust optimization problems. *Stoch. Models* **22**(4), 753–831 (2006)
48. Sekine, J.: Dynamic minimization of worst conditional expectation of shortfall. *Math. Finance* **14**(4), 605–618 (2004)
49. Wüthrich, M.V., Bühlmann, H., Furrer, H.: *Market-consistent actuarial valuation*. European Actuarial Academy (EAA) Series. Springer, Heidelberg (2010) (enlarged edition)



**Thomas Knispel** is a Senior Researcher at the Institute of Probability and Statistics at Leibniz Universität Hannover. He holds a Ph.D. in Mathematics from Humboldt-Universität zu Berlin. His current work focusses on portfolio optimization, model ambiguity, market-consistent actuarial valuation, and risk measurement.



**Gerhard Stahl** hat Mathematik an der TU Karlsruhe studiert. Im Anschluss hieran war er bei der Bundesanstalt für Finanzdienstleistungsaufsicht beschäftigt. Seit 2007 ist Gerhard Stahl bei der Talanx AG tätig und hat seit 2010 die Position des Chief Risk Officers inne. Im Jahr 2006 erhielt er die Ehrendoktorwürde der Universität Bamberg und wurde 2008 zum Honorarprofessor der Universität Ulm bestellt. Seit 2010 ist Gerhard Stahl Honorarprofessor an der Leibniz Universität Hannover. Seine Arbeits- und Forschungsschwerpunkte sind die Risikomessung bei Finanzinstitutionen und das Problem des Statistischen Modellrisikos.



**Stefan Weber** is a Professor of Insurance and Financial Mathematics at the Institute of Probability and Statistics at Leibniz Universität Hannover. He joined the institute in 2009. Prior to this, he was an Assistant Professor in Financial Engineering and Applied Probability in the School of Operations Research and Information Engineering at Cornell University. He holds a Ph.D. in Mathematics from Humboldt-Universität zu Berlin.

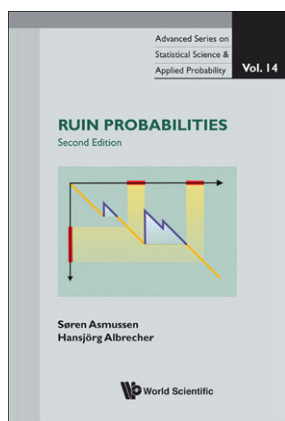
## Søren Asmussen und Hansjörg Albrecher: “Ruin Probabilities”

World Scientific, 620 pp.

Vicky Fasen

Online publiziert: 10. August 2011

© Vieweg+Teubner und Deutsche Mathematiker-Vereinigung 2011



Ruinwahrscheinlichkeiten und Risikotheorie sind klassische Bereiche der Versicherungsmathematik, deren theoretische Fundamente sich in den letzten Jahren immens weiterentwickelt haben. Zur Berechnung der Wahrscheinlichkeit, dass ein Versicherungsunternehmen ruiniert wird, muss dieses diverse Risiken berücksichtigen, die mit dem Verkauf von Versicherungspolicen entstehen. Auf der einen Seite sind die Prämieinnahmen der Versicherungspolicen einem Zufall unterworfen, da die Zahl der Kunden nicht konstant ist. Auf der anderen Seite überträgt der Kunde mit dem Kauf der Police das Risiko eines Schadens, was unkalulierbar ist, auf das Versicherungsunternehmen (Zeitpunkt und Höhe eines Schadens sind ungewiss). Zudem

kann das Versicherungsunternehmen seine Reserven investieren, woraus wieder Risiken entstehen. All diese Risiken müssen bei der Modellierung des Vermögens und der Berechnung der Ruinwahrscheinlichkeit des Versicherungsunternehmens beachtet werden und führen zu komplexen mathematischen Aufgaben, die Inhalt dieses Buches sind.

Die 2. Auflage des Buches *Ruin Probabilities* von Søren Asmussen und Hansjörg Albrecher wurde 2010 vom Verlag World Scientific veröffentlicht. Die neue Auflage ist eine überarbeitete und um einige Kapitel erweiterte Fassung der ersten Auflage von Søren Asmussen aus dem Jahr 2000. Im Gegensatz zu vielen anderen Büchern,

---

V. Fasen (✉)

RiskLab, ETH Zürich, Rämistrasse 101, Zürich, Schweiz

e-mail: [vicky.fasen@math.ethz.ch](mailto:vicky.fasen@math.ethz.ch)

die nur die klassische Ruintheorie von Cramér und Lundberg behandeln, hebt sich dieses Buch ab, da es den aktuellen Stand der Wissenschaft und Forschung der letzten Jahre mit einbezieht und zudem die Verknüpfung zur Warteschlangentheorie und anderen Bereichen der angewandten Wahrscheinlichkeitstheorie präsentiert. Das Buch geht über ein Standardlehrbuch hinaus und hat vielmehr den Charakter eines Nachschlagewerks mit einer beeindruckenden Zusammenstellung von Referenzen (insgesamt 923) und Bemerkungen am Ende jedes Abschnitts. Im Vordergrund steht der mathematische Aspekt der Ruintheorie. Auch wenn der Praxisbezug bewusst nicht im Vordergrund steht, können Modelle und Methoden für jeden Aktuar von Interesse sein.

In der neuen Auflage wurden fünf Kapitel aus aktuellen Forschungsbereichen der letzten Jahre hinzugefügt, die sich mit

- Niveau-abhängigen Risikoprozessen,
- Ruinwahrscheinlichkeiten von Lévy-Prozessen,
- Gerber–Shiu-Funktionen,
- Modellen mit Abhängigkeiten,
- Stochastischer Kontrolltheorie

beschäftigen. Daraus resultiert eine Erhöhung der Seitenzahl von 385 auf 602 Seiten und in einer deutlichen Absetzung der zweiten Auflage von der ersten. Aufgrund der inhaltlichen Dichte des Buches ist es nur möglich, einen kurzen Überblick über den Inhalt zu geben. In dem Buch werden jedoch noch weitere Themen angesprochen, die hier außer Acht gelassen werden müssen.

Das Buch beginnt in Kapitel I mit einer Einführung in das klassische Risikomodell von Cramér–Lundberg und den wesentlichen Verteilungen, die im Versicherungskontext benötigt werden. Es endet mit einer Kurzzusammenfassung über die verschiedenen Methoden der Risikotheorie, die im Laufe des Buchs näher behandelt werden.

In Kapitel II und III werden die Grundlagen über stochastische Prozesse präsentiert, die bei den weiterführenden Modellen zwingend notwendig sind. Sie werden an Beispielen aus der Risikotheorie erläutert. Im Vergleich zur ersten Auflage, in der beide Kapitel noch unter einem Kapitel zusammengefasst waren, wurden viele Beispiele hinzugefügt. Ein mathematisch erfahrener Leser kann diese Kapitel auch überspringen und im späteren Verlauf gegebenenfalls nachlesen.

Im Folgenden werden dann in den Kapiteln IV–VIII die wesentlichen Modelle der Risikotheorie dargelegt und Ruinwahrscheinlichkeiten für Schadenshöhenverteilungen mit leichten Tails untersucht. Die anschließenden Kapitel IX–XIII betrachten das Verhalten beim Ruin der vorher eingeführten Modelle detaillierter oder verallgemeinern die Modelle. Es werden dort die modernen Themen der Ruintheorie abgehandelt. Diese Kapitel sind so gestaltet, dass sie je nach Interesse des Lesers auch unabhängig voneinander gelesen werden können und beginnen meist mit einer kurzen Einführung.

Die klassische Ruintheorie in unendlicher Zeit, zurückgehend auf die Schweden Cramér und Lundberg, die den Schadensankunftsprozess mit einem Poissonprozess unabhängig von den unabhängig und identisch verteilten Schadenshöhen modellieren, wird dann in Kapitel IV behandelt und geht bei weitem über das hinaus, was man

in anderen Textbüchern findet. Nicht nur, dass die wahrscheinlichkeitstheoretischen Eigenschaften des Risikoprozesses und Schadensprozesses sowie die Pollaczek–Khinchine-Formel hergeleitet werden, es werden auch die gemeinsame Verteilung des Defizits beim Ruin und des Überschusses kurz vor dem Ruin sowie Spezialfälle der Pollaczek–Khinchine-Formel berechnet. Alle Resultate werden rigoros bewiesen. Des Weiteren werden in diesem Kapitel die Lundberg-Ungleichung und die Cramér–Lundberg-Approximation hergeleitet, eine der Kernaussagen der klassischen Risikotheorie. Andere Themen sind auch noch Inhalt dieses Kapitels: Approximation von Ruinwahrscheinlichkeiten; Vergleich von Risiken und Methoden zur Schätzung des Lundberg-Koeffizienten; Sensitivität der Ruinwahrscheinlichkeit auf Veränderungen in der Prämienrate, der Schadenshöhenverteilung und der Ankunftsrate. Die komplexere Fragestellung der Ruinwahrscheinlichkeit in endlicher Zeit ist Gegenstand von Kapitel IV.

Nachdem sich Kapitel II–V nur mit dem klassischen Risikoprozess auseinandergesetzt haben, wird das Modell in Kapitel VI erweitert, indem der Schadensankunftsprozess mit einem allgemeineren Erneuerungsprozess modelliert wird; dem *Sparre-Andersen-Prozess*. Die Modellierung des Schadensankunftsprozess durch einen Markovprozess oder periodische Fluktuationen ist dann Bestandteil von Kapitel VII.

In Kapitel VIII wird der Schadensprozess zwar wieder mit einem zusammengesetzten Poissonprozess modelliert, dafür kann der Prämienprozess von dem Überschussprozess abhängen und wächst damit nicht notwendigerweise linear. Zu Beginn werden motivierende Beispiele für dieses Modell gegeben, die dann im Folgenden näher untersucht werden. Dazu zählen Modelle, in denen das Versicherungsunternehmen sein Vermögen zu einer konstanten Zinsrate investiert, es sein Vermögen risikobehaftet anlegt, wobei die risikobehaftete Anlage mit einer geometrischen Brownschen Bewegung modelliert wird, oder es für seine Gewinne Steuern zahlen muss. Für jedes dieser Modelle wird dann das Ruinverhalten analysiert. Selbst das Ruinverhalten von stochastischen Rekurrenzgleichungen wird näher behandelt.

Phasenverteilungen sind eine wichtige Klasse von Verteilungen in der angewandten Wahrscheinlichkeitstheorie, da sie bei numerischen Berechnungen oft sehr effizient sind. Insbesondere besitzt die Ruinwahrscheinlichkeit im zusammengesetzten Poisson-Modell mit Schadenshöhen, die eine Phasenverteilung besitzen, eine analytische Darstellung. Darauf wird in Kapitel IX näher eingegangen. Zudem werden die Ruinwahrscheinlichkeiten der Modelle aus IV–VIII mit phasenverteilten Schäden betrachtet. Es wird aber auch eine kurze Einführung in Matrix-Exponentialverteilungen gegeben, die eine Verallgemeinerung von Phasenverteilungen sind.

Da man sich bisher auf Ruinwahrscheinlichkeiten für Schäden mit leichten Tails beschränkt hat (d.h. die Cramér–Lundberg-Bedingung ist erfüllt), wird in Kapitel X das Ruinverhalten für Modelle aus IV–VIII mit schweren Tails (subexponentiell) erarbeitet. Das Kapitel schließt mit einem Abschnitt über statistische Methoden aus der Extremwerttheorie zur Schätzung der Tails von Verteilungen ab, die das Ruinverhalten bestimmen.

In Kapitel XI wird das klassische Risikomodell verallgemeinert, indem der Schadenshöhenprozess nun durch einen Lévy-Prozess modelliert wird. Zunächst wird mit

einer Einführung in Lévy-Prozesse begonnen. Danach wird die Ruinwahrscheinlichkeit sowohl für den Fall mit schweren Schadensverteilungen als auch im Fall, dass die Cramér–Lundberg-Bedingung erfüllt ist, berechnet. Zudem werden Skalierungsfunktionen als Hilfsmittel zur Berechnung von zweiseitigen Ruinwahrscheinlichkeiten vorgestellt und ein Überblick über Fluktuationstheorie gegeben.

Nachdem in Kapitel IV schon die gemeinsame Verteilung des Überschusses kurz vor dem Ruin und des Defizits beim Ruin hergeleitet wurden, werden nun mit *Gerber–Shiu-Funktionen* moderne Erweiterungen dieser Größen untersucht. Es sei  $(R_t)_{t \geq 0}$  der Risikoprozess,  $\tau(u)$  der Ruinzeitpunkt bei einem Anfangskapital  $u$ ,  $|R_{\tau(u)}|$  das Defizit beim Ruin und  $R_{\tau(u)-}$  der Überschuss kurz vor dem Ruin, dann nennt man

$$m(u) = \mathbb{E}[e^{-\delta\tau(u)} w(R_{\tau(u)-}, |R_{\tau(u)}|); \tau(u) < \infty]$$

Gerber–Shiu-Funktion, wobei die Straffunktion  $w$  nicht-negativ ist. Ihre Eigenschaften und Lösungen werden im Fall des klassischen Risikomodells, Erneuerungsmodell und Lévy-Risikomodells in Kapitel XII analysiert.

Verallgemeinerte Modelle sind Bestandteil von Kapitel XIII, die zum Beispiel Abhängigkeiten in den Schadenshöhen oder Zwischenankunftszeiten erlauben (lineare Modelle, Shot-Noise-Modelle, Sparre-Andersen-Modell). Ziel ist es, Motivationen zu geben, wie die Abhängigkeitsstruktur die Ruinwahrscheinlichkeit beeinflussen kann und welche alternativen Modelle noch betrachtet werden können. Aus diesem Grund ist dieses Kapitel eher kurz gefasst.

Auch Versicherungsunternehmen müssen sich Strategien überlegen, um ihr Vermögen (Nutzen) zu maximieren, z.B. welchen Anteil ihrer Reserven in eine risikobehaftete Anlage investiert wird. Die Berechnung der besten Strategie führt zur stochastischen Kontrolltheorie in diskreter Zeit (dynamische Programmierung) und in stetiger Zeit (Lösen von Hamilton–Jacobi–Bellman-Gleichungen). Ein Überblick in das reichhaltige und sehr aktuelle Gebiet der stochastischen Kontrolltheorie mit Anwendungen im Versicherungskontext ist Gegenstand von Kapitel XIV.

Verschiedene Verfahren zur Simulierung der Ruinwahrscheinlichkeit, insbesondere Importance Sampling, in endlicher und unendlicher Zeit, mit schweren und leichten Schadensverteilungen werden in Kapitel XV präsentiert. Im letzten Kapitel XVI, unter dem Namen *diverse Themen*, wird etwas über Risikomodelle in diskreter Zeit, Approximation der Schadenshöhenverteilung, Prämienberechnung und Rückversicherung gesagt. Ein Anhang mit diversen Grundlagen schließt das Buch ab.

Das Buch stellt imposant die Entwicklung mathematischer Modelle und Methoden der klassischen und modernen Ruinthorie der letzten Jahrzehnte dar. Es zählt eindeutig zu den Monographien aus der Kategorie angewandter Wahrscheinlichkeitstheorie mit sehr vielen Ausblicken und Verweisen zu Erweiterungsmöglichkeiten und alternativer Literatur. Beeindruckend ist, wie präzise jede Aussage bewiesen wird. Teilweise werden am Ende eines Abschnitts noch alternative Beweise vorgestellt. Der Leser sollte aber eine gewisse Vertrautheit mit Wahrscheinlichkeitstheorie und stochastischen Prozessen mitbringen. Ohne dieses Grundwissen wird dieses Buch vermutlich schwer zu verstehen sein, obwohl alles rigoros definiert wird. Wer dieses Wissen mitbringt, wird dieses Buch mit großem Gewinn lesen. Die Notation richtet



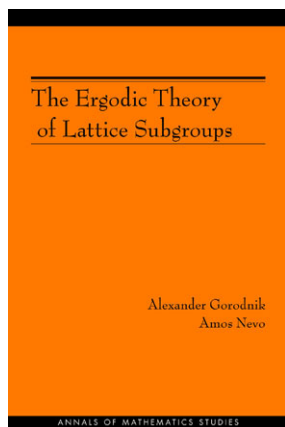
sich an die Konventionen der Warteschlangentheorie und ist deshalb für einen Versicherungsmathematiker nicht direkt zugänglich. Aber das Notationsverzeichnis im Anschluss an das Vorwort, das es in der ersten Auflage noch nicht gab, ist diesbezüglich eine Bereicherung. Zusammenfassend kann man sagen: Es ist ein ausgesprochen gelungenes Werk, das es so noch nicht auf dem Markt gibt. Es wird sicherlich in den nächsten Jahren als Standardreferenzwerk im Bereich der Ruintheorie verwendet.

**Alexander Gorodnik and Amos Nevo:**  
**“The Ergodic Theory of Lattice Subgroups”**  
Princeton University Press, 2010

**Manfred Einsiedler**

Online publiziert: 10. August 2011

© Vieweg+Teubner und Deutsche Mathematiker-Vereinigung 2011



Die Ergodentheorie beschäftigt sich mit dynamischen Systemen unter Verwendung eines statistischen Gesichtspunktes. Ein dynamisches System kann zum Beispiel aus einer Abbildung  $T : X \rightarrow X$  auf einem Raum  $X$  bestehen und ein wichtiger Aspekt der Theorie ist die Untersuchung der Bahnen

$$\{T^n(x) : n = 0, 1, 2, \dots\}$$

eines Punktes  $x \in X$ . Für einen gegebenen Ausgangspunkt kann diese Untersuchung mitunter sehr schwierig sein, aber die Untersuchung der Bahnen typischer Punkte ist oft viel einfacher. Das Wort „typisch“ ist in diesem Zusammenhang bzgl. eines invarianten Wahrscheinlichkeitsmaßes zu verstehen.

Allgemeiner besteht ein dynamisches System aus einer Gruppenwirkung

$$\cdot : G \times X \rightarrow X, \quad (g, x) \in G \times X \mapsto g \cdot x.$$

Ein Wahrscheinlichkeitsmaß  $\mu$  auf  $X$  heißt  $G$ -invariant falls  $\mu(g^{-1} \cdot B) = \mu(B)$  für alle  $g \in G$  und alle meßbaren  $B \subset X$ . Ein invariantes Wahrscheinlichkeitsmaß  $\mu$  auf  $X$  heißt ergodisch, falls für ein meßbares  $B \subset X$  mit  $g^{-1} \cdot B = B$  für alle  $g \in G$  folgt, daß  $\mu(B) \in \{0, 1\}$ . Also muß jede invariante Menge  $B$  erfüllen, daß  $B$  entweder fast leer oder fast ganz  $X$  sein muss.

Oft schränkt man sich in der Untersuchung auf ergodische invariante Maße ein, was aus zwei Gründen keine sehr drastische Einschränkung ist. Erstens sind viele natürliche invariante Maße ergodisch (außer wenn Invarianten vorhanden sind) und zweitens läßt sich (mittels Choquet's Theorem) jedes invariante Wahrscheinlichkeitsmaß in ergodische Wahrscheinlichkeitsmaße zerlegen.

Als Beispiel für obige Behauptung, daß Bahnen typischer Punkte einfacher zu untersuchen sind, wollen wir folgende Aussage beweisen. Für ein ergodisches invariantes Wahrscheinlichkeitsmaß  $\mu$  gilt für  $\mu$ -fast alle  $x \in X$ , daß der Abschluß der Bahn  $G \cdot x$  gleich dem Träger des Maßes  $\mu$  ist. Hier nehmen wir an, daß  $X$  eine abzählbare Basis der Topologie hat, und zur Einfachheit auch, daß  $G$  abzählbar ist. In diesem Fall gilt für jede offene Menge  $O$  mit  $\mu(O) > 0$ , daß die Bahn  $G \cdot O = \{g \cdot x : g \in G, x \in O\}$  invariant ist und daher  $\mu(G \cdot O) = 1$  erfüllen muß. Wendet man dies auf alle offene Mengen in der Basis, die den Träger von  $\mu$  schneiden, an, und nimmt man den abzählbaren Durchschnitt der so erhaltenen meßbaren Mengen vom Maß eins, ergibt sich die Aussage.

Eine wichtige Verallgemeinerung obiger Untersuchung des Abschlusses der Bahn ist die Untersuchung der Verteilung der Bahn. Einer der grundlegenden Sätze der Ergodentheorie ist der Birkhoffsche Ergodensatz von 1931, welcher diese Frage nach der Verteilung der Bahn für eine Abbildung beantwortet. Genauer formuliert: Für eine Abbildung  $T : X \rightarrow X$ , ein gegebenes invariantes Wahrscheinlichkeitsmaß  $\mu$  auf  $X$  und eine integrierbare Funktion  $f$  auf  $X$  gilt, daß der Grenzwert

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = F(x)$$

für  $\mu$ -fast alle  $x \in X$  existiert. Unter Verwendung der Fast-Invarianz von Intervallen in  $\mathbb{N}$  sieht man schnell, daß  $F(x) = F(T(x))$  gelten muß, wenn nur der Grenzwert für  $x$  existiert. Falls  $\mu$  zusätzlich ergodisch ist, dann ist  $F(x) = \int f d\mu$  für  $\mu$ -fast alle  $x$ . Dieser Ergodensatz wurde im letzten Jahrhundert mehrmals verallgemeinert. Ein natürlicher Gültigkeitsbereich dieses Satzes, nämlich der aller Gruppenwirkungen von Gruppen, auf denen fast invariante Teilmengen existieren – sogenannte mittelbare Gruppen, wurde 1999 in einer Arbeit von Elon Lindenstrauss erreicht.

In dem Buch ‚The Ergodic Theory of Lattice Subgroups‘ geht es um Gruppenwirkungen von halbeinfachen Liegruppen, wie zum Beispiel  $G = \mathrm{SL}(n, \mathbb{R})$ , und Gittern in diesen, zum Beispiel  $\Gamma = \mathrm{SL}(n, \mathbb{Z})$ . Allgemeiner werden auch  $p$ -adische Liegruppen und Liegruppen über einem lokalen Grundkörper mit positiver Charakteristik und Produkte von diesen Gruppen erlaubt. Solche halbeinfachen Liegruppen sind nicht mittelbar und somit ist der oben erwähnte Ergodensatz nicht anwendbar. Auf den ersten Blick ist es nicht einmal klar, wie man obige Mittelbildungen über Teile der Bahnen verallgemeinern sollte. Wie aber Gorodnik und Nevo in ihrem Buch zeigen, gibt es viele Familien von Teilmengen, wie zum Beispiel Kugeln bezüglich einer linksinvarianten Metrik auf  $G$ , für die Ergodensätze gelten. In der Tat sind diese Sätze oft viel stärker als ihre Vorgänger für mittelbare Gruppen. Um diesen Unterschied zu erklären wollen wir eine Version eines dieser Sätze formulieren.

Sei  $\Gamma = \mathrm{SL}(n, \mathbb{Z})$  für  $n \geq 2$ . Sei  $\Gamma_t = \{\gamma \in \Gamma : \log \|\gamma\| < t\}$ , wobei  $\|\cdot\|$  eine Matrixnorm bezeichnet. Sei  $(X, \mu)$  ein Wahrscheinlichkeitsraum mit einer Wirkung

von  $\Gamma$  auf  $X$  so daß  $\mu$   $\Gamma$ -invariant und ergodisch ist. Für jedes  $f \in L^p(X)$  mit  $1 < p < \infty$  gilt

$$\lim_{t \rightarrow \infty} \frac{1}{|\Gamma_t|} \sum_{\gamma \in \Gamma_t} f(\gamma^{-1} \cdot x) = \int f d\mu$$

sowohl in  $L^p$  als auch für  $\mu$ -fast alle  $x \in X$ . Falls zusätzlich ‚spectral gap‘ (siehe unten) für die Wirkung vorliegt, dann gilt sogar

$$\left| \frac{1}{|\Gamma_t|} \sum_{\gamma \in \Gamma_t} f(\gamma^{-1} \cdot x) - \int f d\mu \right| \leq C_p(f, x) e^{-\zeta_p t} \quad (1)$$

für Konstanten  $C_p(f, x)$ ,  $\zeta_p > 0$ .

Der Zusatz (1), der eine exponentielle Fehlerabschätzung liefert, ist für mittelbare Gruppen nur für sehr spezielle Wirkungen und auch dann nur für glatte Funktionen möglich. Für halbeinfache Gruppen und deren Gitter sind ‚spectral gap‘ und die Resultate für  $L^p$ -Funktionen von Gorodnik und Nevo allerdings sehr häufig erfüllt.

Wir wollen eine Definition von spectral gap (von mehreren äquivalenten) vorstellen. Eine unitäre Darstellung  $\pi$  einer Gruppe  $G$  auf einem Hilbertraum  $H$  hat *spectral gap*, falls es ein  $\epsilon > 0$  und eine kompakte Menge  $K \subset G$  gibt, so daß der Faltungsoperator

$$\chi_K(v) = \frac{1}{m_G(K)} \int_K \pi(g)v dm_G$$

( $m_G$  ist ein Haarmaß auf  $G$ ) die Abschätzung

$$\|\chi_K(v)\|_H \leq (1 - \epsilon)\|v\|_H \quad \text{für } v \in (H^G)^\perp$$

erfüllt. Hier ist  $H^G = \{v \in H : \pi(g)v = v \text{ für alle } g \in G\}$  der Teilraum, auf dem  $G$  trivial wirkt. Da  $\chi_K$  ein selbstadjungierter beschränkter Operator ist, der  $H^G$  als Eigenraum zum Eigenwert 1 hat, beschreibt obige Abschätzung in der Tat eine Lücke im Spektrum von  $\chi_K$ . Für uns interessant ist die Darstellung

$$(\pi(g)(f))(x) = f(g^{-1} \cdot x) \quad \text{für } x \in X, f \in H = L^2(X, \mu) \text{ und } g \in G,$$

die von einer Wirkung auf einem Wahrscheinlichkeitsraum mit einem invarianten Maß induziert wird, für diese Darstellung ist Ergodizität von dem Maß äquivalent zu  $\dim(H^G) = 1$ .

Für halbeinfache Gruppen ist spectral gap eine häufige Erscheinung: Jede irreduzible Darstellung einer halbeinfachen Liegruppe hat spectral gap und für manche Gruppen können  $K$  und  $\epsilon$  sogar unabhängig von der Darstellung gewählt werden – diese Eigenschaft nennt man Property (T). Die Gruppen  $G = \mathrm{SL}(n, \mathbb{R})$  und  $\Gamma = \mathrm{SL}(n, \mathbb{Z})$  haben zum Beispiel Property (T) für  $n \geq 3$ .

Es ist einfach zu sehen, warum spectral gap für die Untersuchung von Mitteln über Bahnen nützlich sein kann: Falls  $G$  auf  $(X, \mu)$  ergodisch mit spectral gap wirkt und  $f \in L^2(X)$  ist, dann folgt

$$\left\| \chi_K^n(f) - \int f d\mu \right\|_2 \leq (1 - \epsilon)^n \|f\|_2,$$

wobei  $\chi_K^n(f)$  die  $n$ -fache Anwendung des obigen Faltungsoperators auf  $f$  bezeichnet. Allerdings kann man  $\chi_K^n(f)$  auch als die Faltung einer absolut stetigen Dichteverteilung auf  $K^n \subset G$  mit  $f$  verstehen – und dies ist bis auf die Dichteverteilung eben ein Ergodensatz. Mit (viel) mehr Arbeit lässt sich diese Dichteverteilung durch eine konstante Verteilung ersetzen und so erhält man die exponentielle Abschätzung (1) zumindest im  $L^2$ -Mittel.

Eine Motivation für die Untersuchungen von Gorodnik und Nevo sind asymptotische Abzählungsergebnisse. Duke, Rudnick und Sarnak [1] haben gezeigt, daß es einen engen Zusammenhang zwischen den asymptotischen Abzählungen von Punkten in Bahnen von Gittern  $\Gamma$  auf homogenen Räumen  $G/H$  und Gleichverteilungsfragen für Bahnen der Form  $gH\Gamma \subset G/\Gamma$  gibt. Weiter haben Eskin und McMullen [3] gezeigt, wie man diese Gleichverteilungsergebnisse mit dynamischen Argumenten beweisen kann. Gorodnik und Nevo erhalten Resultate in diese Richtung, wenn auch nicht in der gleichen Allgemeinheit. In den Situationen, wo die Methode von Gorodnik und Nevo funktioniert, erhalten sie aber eine sehr gute Fehlerabschätzung (in [3] wurde keine Fehlerabschätzung bewiesen und in [1] werden nur spezielle Teilmengen für die Abzählung erlaubt). Zum Beispiel erhalten Gorodnik und Nevo für sehr allgemeine Familien von Teilmengen  $B_t \subset G$ , daß

$$\lim_{t \rightarrow \infty} \frac{|\Gamma \cap B_t|}{m_G(B_t)} = \frac{1}{m(G/\Gamma)},$$

wobei  $m$  das Maß auf  $G/\Gamma$  ist, das von  $m_G$  induziert wird. Weiter wird ein exponentiell kleiner Fehlerterm bewiesen, der wiederum von der spectral gap Eigenschaft abhängt.

Weitere Informationen zu dem Buch von Gorodnik und Nevo sind in [2] enthalten.

## Literatur

1. Duke, W., Rudnick, Z., Sarnak, P.: Density of integer points on affine homogeneous varieties. *Duke Math. J.* **71**(1), 143–179 (1993)
2. Einsiedler, M. (Reviewer): The ergodic theory of lattice subgroups, by Alexander Gorodnik and Amos Nevo. *Bull. Am. Math. Soc.* **48**, 475–480 (2011)
3. Eskin, A., McMullen, C.: Mixing, counting, and equidistribution in Lie groups. *Duke Math. J.* **71**(1), 181–209 (1993)