

# 保険と金融の統計学<sup>1</sup>

国友直人<sup>2</sup>・川崎能典<sup>3</sup> 共編

2007年3月

<sup>1</sup>この報告書は東京大学日本経済国際共同研究センター CIRJE と統計数理研究所リスク解析センターの共同で2006年度に行われた研究会「保険と金融の統計学」の内容をまとめたものである。なお、この研究会の運営の一部は東京大学公共政策大学院「損保ジャパン寄付講座」の支援を受けて実施された。

<sup>2</sup>東京大学経済学研究科教授

<sup>3</sup>統計数理研究所助教授

## 概要

この研究報告書は2006年度に開催した研究会「保険と金融の統計学」での特別講義、研究報告・議論などの一部分をまとめたものである。2006年度に開催した同研究会では、関係者の熱心な協力もあり、特にこれまであまり日本語では利用可能でなかった保険分野や金融分野における統計的手法の基礎的内容が扱われてた。例えば、志村隆影氏（統計数理研究所）による「正則変動関数の理論」に関する特別講義、西山陽一氏（統計数理研究所）による「確率過程の統計学」に関する特別講義、その他、このところ発展がめざましい統計的極値論の理論と応用、数理保険の理論と応用などが具体的内容である。これらの話題は直接に研究会に参加した関係者以外の研究者、大学院生、業界関係者にとっても有用な情報を含んだものである、と研究会幹事であった国友直人（東京大学経済学研究科）及び川崎能典（統計数理研究所）は判断している。本報告書に収録した内容は、これまで様々な応用の場所で断片的にしか利用可能でなかった確率論、統計学的議論が多く含まれているが、近年ではいずれの話題も数理保険学や数理ファイナンス・計量ファイナンスなどと呼ばれている、金融に係わる統計的リスク管理の分野で重要な役割を演じることが多くなっている。

保険分野や金融分野において実際に利用されている確率論的方法や統計的方法の理解と発展に関心がある研究者、大学院生、業界の研究開発に携わる実務家にとって、ここに収録された内容が貴重な資料として勉学の参考になれば幸いである。この報告書が特に今後の研究の一助となることを期待したい。

2007年3月

編者

# 目次

## 第 I 部：連続講義録

1. 「正則変動性と確率論」 志村隆影
2. 「確率過程の統計学：マルチンゲールと empirical process」 西山陽一

## 第 II 部：金融リスクと統計的極値論

3. 「One-parameter Families of Distortion Risk Measures」 Hideatsu Tsukahara (塚原英敦)
4. 「経験尤度法による極値パラメターの推定」 大和田孝
5. 「Extreme quantile estimation using extreme value theory」 Tatsuyoshi Okimoto (沖本竜義)
6. 「コピュラの保険への応用例」 岡賢一
7. 「コピュラの性質とその応用」 広瀬光久

## 第 III 部：保険と統計学

8. 「保険リスクと破産理論 (I),(II)」 室井芳史
9. 「変額年金保険の統計的リスク管理法：局面転換モデルの利用」 (日本アクチュアリー会編「リスクと保険」2006年より) 秋山豪太・国友直人
10. 「Swiss Solvency Test について」 川崎能典

## 第I部：連続講義録

1. 「正則変動性と確率論」 志村隆影
2. 「確率過程の統計学：マルチンゲールと empirical processs」 西山陽一

# 正則変動性と確率論

志村 隆彰\*

## はじめに

本稿は平成18年6月から7月にかけて、連続講義シリーズ「保険と金融のための応用確率論」の一環として東京大学大学院経済学研究科で5回にわたって筆者が行った正則変動関数の入門的講義の報告書として書かれたものである。

正則変動関数とは、おおざっぱに言えば、漸近挙動がべき乗関数に“近い”関数であり、正確にはべき乗関数に緩慢変動関数をかけたものである。べき乗関数（ここでは、 $x$ の実数乗）を滑らかに拡張したものであるから、数学の諸分野で高い汎用性をもち、特に確率統計では、極限定理と関連する確率分布の裾の挙動の特徴付けなどに用いられる。内容は、表題の通り、正則変動関数の理論とそれに関連する確率論の話である。全体を2章に分け、第1章では、正則変動関数の一般論を定義から述べ、第2章では正則変動性が登場する代表的な確率や統計の分野を紹介する。扱う話題は、第1章では正則変動関数の定義、例、基本的性質、重要な定理、積分の漸近挙動についてのカラマタの定理、ラプラス変換に関するアーベル型定理、タウバー型定理、正則変動性から派生した $\Gamma$ 及び $\Pi$ -変動性。第2章では、正則変動性で特徴付けされる安定分布の吸引域、極値分布の吸引域、重い裾をもつ分布の典型例である subexponential 分布について述べる。いずれの話題についても確率統計の基本的結果も書くことで理解の助けとした。ただし、個々の命題や定理の証明は省き、代わりに引用元の文献を最後にまとめて示し、主な文献については簡単な内容紹介も加えておいた。

筆者が学生のころは、確率論の研究会でも正則変動関数を使うときには、定義と例（ほとんどは対数関数）を説明していたような記憶があるが、時を経て、経済の分野でも必要とされるようになったことは感慨深く、今回、このような機会を与えてくださった東京大学大学院経済学研究科の国友直人教授と統計数理研究所の川崎能典助教授に感謝の意を表したい。

## 1 正則変動関数

ここでは、正則変動関数の定義からはじめて、例をあげた後、重要な性質、定理等について述べる。

---

\*統計数理研究所,

## 1.1 正則変動関数の定義, 例, 重要定理

$f(x) = x^\rho$  ( $\rho \in \mathbb{R} = (-\infty, \infty)$ ) の形をもつ関数をべき乗関数 (power function) と呼ぶ. 正則変動関数 (regularly varying function) を一言で言えば, “べき乗関数に近い漸近的挙動をする関数”. もう少し詳しくいえば, べき乗関数を自然な形で滑らかに拡張したものである. 正則変動関数には指数 (index 或いは exponent) が定まり, べき乗関数を正則変動関数としてみれば,  $\rho$  がそれにあたる. 指数 0 の正則変動関数を緩慢変動関数といい, その名の通りゆっくり変化する関数である.

**定義 1.1.1**  $[a, \infty)$  ( $a \in \mathbb{R}$ ) 上で定義された正值可測関数  $f(x)$  が ( $\infty$  で) 指数 (index, exponent)  $\rho \in \mathbb{R}$  の正則変動関数であるとは, 任意の  $\lambda > 0$  に対して,

$$(1.1) \quad \lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho.$$

を満たすときを言う. 特に, 指数 0 のとき, 緩慢変動関数と呼ぶ.

指数  $\rho$  の正則変動関数が, べき乗関数  $x^\rho$  と緩慢変動関数  $l(x)$  の積で表すことが出来ることは直ちにわかるであろう.

$$f(x) = x^\rho l(x).$$

指数  $\rho$  の正則変動関数全体からなる集合を  $\mathbf{R}_\rho$  で表し,  $\mathbf{R} = \cup_{\rho \in \mathbb{R}} \mathbf{R}_\rho$  とする.

**注意 1.1.1**  $l(x)$  が 0 での正則変動関数とは  $(0, a)$  ( $a > 0$ ) で定義され,  $l(1/x)$  が  $\infty$  での正則変動関数となるときをいう. 数学的には同等なので,  $\infty$  での正則変動関数のみを扱う.

上記のことから, 正則変動関数の例は緩慢変動関数の例をみれば十分である. 正定数関数は自明な緩慢変動関数であり, べき乗関数に対応する. 次に思いつくのは対数関数 (及びそのべき乗) であろうが, 緩慢変動関数にはこうしたもの以外にも, 多様なものがある. 次に示す例 5 は, 単調ではなく, 値域  $(0, \infty)$  を振動する例である. こうした例は一見意外に感じられるかもしれないが, 後述する表現定理をみれば, 感覚的にも納得できるであろう.

**例 1.1.1** 以下は緩慢変動関数である.

1.  $l(x) = c + o(1)$ . ここで,  $c$  は正の定数,  $\lim_{x \rightarrow \infty} o(1) = 0$ .
2.  $l(x) = \log x$ .
3.  $l(x) = (\log x)^2$ .
4.  $l(x) = \exp \sqrt{\log x}$ .
5.  $l(x) = \exp\{(\log x)^{\frac{1}{3}} \cos((\log x)^{\frac{1}{3}})\}$ .

**注意 1.1.2** 単調関数が緩慢変動であるためには, ある  $\lambda \neq 1$  に対して (1.1.1) 成り立てば十分である.

**例 1.3.1 (ペーター・ポール分布)** 聖ペテルスベルグのゲームとは、コインを表が出るまで投げ続け、 $k$  回目に表が出たならば、 $2^k$  円受け取るというものである。このゲームの賞金の分布  $F$  は、

$$F(x) = 1 - \sum_{2^n > x} 2^{-n}$$

となり、裾  $\sum_{2^n > x} 2^{-n}$  はジャンプする点をみれば、正則変動しないことはすぐにわかる。ところが、 $\int_0^x y dF(dy)$  や  $\int_0^x (1 - F(y)) dy$  は正則変動することが確かめられる。

## 1.4 タウバー型定理

この節では、正則変動性と関連したラプラス変換を扱う。確率分布の性質がいかにそのラプラス変換に反映されるかといったことに応用される。

**定義 1.4.1 (ラプラス変換)**  $U(x)$  を  $[0, \infty)$  上の右連続な局所有界変動関数とする。 $U(x)$  のラプラス変換 (Laplace-Stieltjes transform) は次のように定義される。

$$\hat{U}(s) = \int_0^{\infty} e^{-st} U(dt).$$

原点の近傍の  $\hat{U}(s)$  の漸近的な性質と  $U(x)$  の  $x \rightarrow \infty$  の漸近的性質が対応するといった関係をタウバー型定理 (Tauberian Theorem) という。

例をあげよう。 $U(x)$  が確率分布  $F(x)$  でそのラプラス変換が計算されているとき、その原点付近の挙動をみることで、裾  $1 - F(x)$  の  $x \rightarrow \infty$  のときの挙動がわかる。

**例 1.4.1**  $0 \leq \alpha < 1$  と緩慢変動関数  $l(x)$  に対して、次は同値である。

$$(i) 1 - F(x) \sim l(x)/(x^\alpha \Gamma(1 - \alpha)) \quad (x \rightarrow \infty),$$

$$(ii) 1 - \hat{F}(s) \sim s^\alpha l(1/s) \quad (s \rightarrow 0+).$$

ここでは、正則変動性と関連した代表的なタウバー型定理を紹介する。

**定理 1.4.1 (カラマタのタウバー型定理)**  $U(x)$  を  $[0, \infty)$  上の非減少右連続関数、 $l(x)$  を緩慢変動関数、 $c \geq 0$ ,  $\rho \geq 0$  とするとき、次は同値である。

$$U(x) \sim cx^\rho l(x)/\Gamma(1 + \rho) \quad (x \rightarrow \infty),$$

$$\hat{U}(s) \sim cs^{-\rho} l(1/s) \quad (s \rightarrow +0).$$

$c = 0$  の場合は、 $U(x) = o(x^\rho l(x))$  を意味する。後者も同様。ここで、 $\Gamma(z)$  はガンマ関数、すなわち、

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, \quad \Gamma(n+1) = n!.$$

**注意 1.4.1**  $c > 0$  のとき上の2式から次が得られ、これもまた上のふたつと同値である。

$$U(x)/\hat{U}(1/x) \rightarrow 1/\Gamma(1 + \rho) \quad (x \rightarrow \infty).$$

次の定理は前定理において、 $x$  と  $s$  を入れ替えた場合、 $x \rightarrow 0+$ 、 $s \rightarrow \infty$  としても成り立つことを主張している。

**定理 1.4.2**  $U(x)$  を  $[0, \infty)$  上の非減少関数で、十分大きな  $s$  に対して、 $\hat{U}(s) < \infty$  となるもの、 $l(x)$  を緩慢変動関数、 $c \geq 0$ 、 $\rho \geq 0$  とするとき、次は同値である。

$$\begin{aligned} U(x) &\sim cx^\rho l(1/x) / \Gamma(1 + \rho) \quad (x \rightarrow 0+), \\ \hat{U}(s) &\sim cs^{-\rho} l(s) \quad (s \rightarrow \infty). \end{aligned}$$

確率密度の存在を仮定したり、その性質を知りたいということはしばしばある。また、単峰性や裾といった条件から密度の単調性を仮定できる場合も少なくない。単調密度定理 (monotone density theorem) とは、積分の漸近的性質から、密度の漸近的性質を導くものであり、その際に密度の単調性を仮定する。無限遠と原点近傍の二通りを紹介する。

**定理 1.4.3 (単調密度定理)**  $U(x) = \int_0^x u(t) dt$  とし、ある  $c, \rho \in \mathbb{R}$ 、 $l(x) \in \mathbf{R}_0$  との間次関係が成り立っていると仮定する。

$$U(x) \sim cx^\rho l(x) \quad (x \rightarrow \infty).$$

このとき、 $u(x)$  があるところから先で単調ならば、

$$u(x) \sim c\rho x^{\rho-1} l(x) \quad (x \rightarrow \infty)$$

となる。

**注意 1.4.2** この定理では、

$$\frac{xu(x)}{U(x)} \sim \rho$$

が成り立つが、このことから  $u(x)$  の単調性を仮定せずに正値の仮定のみから正則変動性、 $u(x) \in \mathbf{R}_{\rho-1}$  と  $U(x) \in \mathbf{R}_\rho$  が出る。

次の定理は原点付近の挙動についてのものである。

**定理 1.4.4 (単調密度定理)**  $U(x)$  は  $(0, a]$  上で定義され、 $u(x) \in \mathbf{L}^1[0, a]$  により、

$$U(x) = \int_0^x u(t) dt$$

と書かれ、ある  $c \in \mathbb{R}$ 、 $\rho \geq 0$ 、 $0$  での緩慢変動関数  $l(x)$  との間次関係があることを仮定する。

$$U(x) \sim cx^\rho l(x) \quad (x \downarrow 0).$$

このとき、 $u(x)$  がある  $0$  の右近傍で単調ならば、

$$u(x) \sim c\rho x^{\rho-1} l(x) \quad (x \downarrow 0)$$

となる。



前の定理が正則変動関数とその積分との漸近的な関係についてのものであったのに対し、次の定理はその逆も真であることを主張している。すなわち、前述の漸近的な関係から、正則変動性が導かれるのである。

**定理 1.3.2**  $f(x)$  を  $[a, \infty)$  上で局所有界な正值関数とする。

(i) ある  $\sigma > -(\rho + 1)$  に対して、

$$x^{\sigma+1}f(x)/\int_a^x t^\sigma f(t)dt \rightarrow \sigma + \rho + 1$$

ならば、 $f(x)$  は指数  $\rho$  の正則変動関数である。

(ii) ある  $\sigma < -(\rho + 1)$  に対して、

$$x^{\sigma+1}f(x)/\int_x^\infty t^\sigma f(t)dt \rightarrow -(\sigma + \rho + 1)$$

ならば、 $f(x)$  は指数  $\rho$  の正則変動関数である。

次の定理は確率分布の裾と切断積率の漸近的な関係を表す。

**定理 1.3.3**  $F$  を  $[0, \infty)$  上の確率分布とし、

$$V_\sigma(x) = \int_0^x t^\sigma F(dt)$$

とする。

(i)  $V_\sigma(x)$  か  $1 - F(x)$  のどちらかが正則変動すれば、極限

$$(1.3) \quad \lim_{x \rightarrow \infty} \frac{x^\sigma(1 - F(x))}{V_\sigma(x)} = \gamma$$

が存在する。

(ii) 逆に、(1.3.1) が  $0 < \gamma < \infty$  として成り立つとき、これを

$$\gamma = \frac{\sigma - \alpha}{\alpha}$$

という形に書く。このとき、緩慢変動関数  $l(x)$  により、

$$\begin{aligned} V_\sigma(x) &\sim x^{\sigma-\alpha}l(x), \\ 1 - F(x) &\sim \frac{\sigma - \alpha}{\alpha}x^{-\alpha}l(x) \end{aligned}$$

となる。さらに、(1.3.1) が  $\gamma = 0$  として成り立てば、 $V_\sigma(x)$  が緩慢変動し、 $\gamma = \infty$  として成り立てば、 $1 - F(x)$  が緩慢変動することが導かれる。

**注意 1.3.1**  $V_\sigma(x)$  の緩慢変動性から  $1 - F(x)$  の正則変動性を導くことは出来ない。

具体的な応用では、例えば、 $p(t)$  を確率密度とし、前式で、 $f(t) = t^2 p(t)$  とすれば、切断分散を、後式で  $f(t) = p(t)$  とすれば、分布の裾を、それぞれ表す。

次の命題 1.3.1, 1.3.2, 3 から定理 1.3.1 までは  $f(x)$  を正則変動関数としたときの上記の積分の漸近挙動に関する結果である。一見複雑そうに見えるが、形としては、正則変動関数  $f(x)$  の緩慢変動部分を定数のごとく積分の外に出し、残ったべき乗関数を積分したものが、もとの積分と漸近的に等しいという式に過ぎない。

**命題 1.3.1**  $l(x)$  を  $[a, \infty)$  上で局所有界な緩慢変動関数とする。

(i)  $\alpha > -1$  ならば、

$$\int_a^x t^\alpha l(t) dt \sim x^{\alpha+1} l(x) / (\alpha + 1),$$

(ii)  $\alpha = -1$  ならば、 $\int_a^x l(t) dt / t$  は緩慢変動し、

$$\frac{1}{l(x)} \int_a^x l(t) dt / t \rightarrow \infty$$

となる。

**命題 1.3.2**  $l(x)$  を緩慢変動関数とする。

(i)  $\alpha < -1$  ならば、 $\int_x^\infty t^\alpha l(t) dt < \infty$  であり、

$$\int_x^\infty t^\alpha l(t) dt \sim -x^{\alpha+1} l(x) / (\alpha + 1)$$

となる。

(ii)  $\alpha = -1$  で、 $\int_x^\infty l(t) dt / t < \infty$  ならば、 $\int_x^\infty l(t) dt / t$  は緩慢変動し、

$$\frac{1}{l(x)} \int_x^\infty l(t) dt / t \rightarrow \infty$$

となる。

ふたつの命題を定理としてまとめておく。

**定理 1.3.1**  $f(x)$  を  $[a, \infty)$  上で局所有界な指数  $\rho$  の正則変動関数とする。

(i)  $\sigma \geq -(\rho + 1)$  ならば、 $\int_a^x t^\sigma f(t) dt$  は指数  $\sigma + \rho + 1$  の正則変動関数であり、

$$x^{\sigma+1} f(x) / \int_a^x t^\sigma f(t) dt \rightarrow \sigma + \rho + 1,$$

(ii)  $\sigma < -(\rho + 1)$ , あるいは、 $\int_x^\infty t^{-(\rho+1)} f(t) dt < \infty$  であれば、 $\sigma = -(\rho + 1)$  の場合も、 $\int_x^\infty t^\sigma f(t) dt$  は指数  $\sigma + \rho + 1$  の正則変動関数であり、

$$x^{\sigma+1} f(x) / \int_x^\infty t^\sigma f(t) dt \rightarrow -(\sigma + \rho + 1)$$

となる。

関連する事柄として conjugate pair を紹介する.

**定理 1.2.5**  $l(x)$  が緩慢変動するならば,

$$l(x)l^\#(xl(x)) \rightarrow 1, \quad l^\#(x)l(xl^\#(x)) \rightarrow 1$$

を満たす緩慢変動関数  $l^\#(x)$  が存在し, 漸近的に唯一とつに決まる. このとき,  $l^{\#\#}(x) \sim l(x)$  となる.

**定義 1.2.1** この緩慢変動関数  $l^\#(x)$  を緩慢変動関数  $l(x)$  の *de Bruijn conjugate*,  $(l(x), l^\#(x))$  を *conjugate pair* と呼ぶ.

**命題 1.2.1**  $(l(x), l^\#(x))$  を緩慢変動関数の *conjugate pair* とすれば,  $A, B, \alpha > 0$  に対して, 次の組もまた *conjugate pair* となる.

$$(l(Ax), l^\#(Bx)), \quad (Al(x), A^{-1}l^\#(x)), \quad (\{l(x^\alpha)\}^{1/\alpha}, \{l^\#(x^\alpha)\}^{1/\alpha}).$$

この節の最後として,  $f(y)/f(x)$  の大域的有界性について述べる.

**定理 1.2.6 (Potter の定理)** (i)  $l(x)$  が緩慢変動関数ならば, 任意に選ばれた  $A > 1$  と  $\delta > 0$  に対し,  $X = X(A, \delta)$  を次を満たすように取ることができる.  $x \geq X, y \geq X$  に対して,

$$l(y)/l(x) \leq A \max\{(y/x)^\delta, (y/x)^{-\delta}\}.$$

(ii)  $l(x)$  がさらに  $[0, \infty)$  の任意のコンパクト集合上で 0 と  $\infty$  にならないならば, 任意に選ばれた  $\delta > 0$  に対し,  $A' = A'(\delta)$  を次を満たすように取ることができる.  $x > 0, y > 0$  に対して,

$$l(y)/l(x) \leq A' \max\{(y/x)^\delta, (y/x)^{-\delta}\}.$$

(iii)  $f(x) \in \mathbf{R}_\rho$  ならば, 任意に選ばれた  $A > 1$  と  $\delta > 0$  に対し,  $X = X(A, \delta)$  を次を満たすように取ることができる.  $x \geq X, y \geq X$  に対して,

$$f(y)/f(x) \leq A \max\{(y/x)^{\rho+\delta}, (y/x)^{\rho-\delta}\}.$$

### 1.3 積分の漸近挙動

正則変動関数の積分の漸近挙動に関するカラマタの定理について述べる. 第2章で詳しく述べるが, 確率分布の裾や切断積率 (定義は第2章) の正則変動性は独立和の挙動のような, それが導く現象と強く結びついている. 他にも確率密度や裾を積分したものの性質がランダムな現象に反映されることもあり, 以下のような正則変動関数を被積分関数とする積分の漸近挙動を知ることは重要である.

$$\int_a^x f(t)dt, \quad \int_x^\infty f(t)dt.$$

## 1.2 大域的性質

前節で定義したように正則変動性は関数の漸近的性質であるから，漸近的な挙動は最も重要な性質である．この節では，正則変動関数の大域的性質について解説する．

例 1.1.1 で，指数 0，すなわち，緩慢変動関数には，単調なもののみならず，様々な挙動をするものがあることを示した．一方で，指数が 0 でない正則変動関数は漸近的に単調になることを意味するのが最初の定理である．

**定理 1.2.1 (Monotone equivalence)**  $f(x) \in \mathbf{R}_\rho$  を  $[a, \infty)$  ( $a \geq 0$ ) 上で局所有界とする． $\rho > 0$  ならば，

$$(i) \quad \overline{f}(x) = \sup\{f(t) : a \leq t \leq x\} \sim f(x),$$

$$(ii) \quad \underline{f}(x) = \inf\{f(t) : t \geq x\} \sim f(x).$$

$\rho < 0$  ならば，

$$(i) \quad \sup\{f(t) : t \geq x\} \sim f(x),$$

$$(ii) \quad \inf\{f(t) : a \leq t \leq x\} \sim f(x).$$

ここで， $f(x) \sim g(x)$  は  $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$  を意味する．

べき乗関数と緩慢変動関数に漸近的な発散の速さに差があることは，表現定理（定理 1.1.3）からも直ちにわかるが，このことが緩慢変動性の特徴付けにもなっている．

**定理 1.2.2** 正值可測関数  $l(x)$  が緩慢変動関数であることと任意の  $\varepsilon > 0$  に対して，非減少関数  $\phi$  と非増加関数  $\psi$  で

$$x^\varepsilon l(x) \sim \phi(x), \quad x^{-\varepsilon} l(x) \sim \psi(x)$$

を満たすものが存在することとは同値である．

**定理 1.2.3** 正值可測関数  $l(x)$  が正規化緩慢変動関数であることと任意の  $\varepsilon > 0$  に対して， $x^\varepsilon l(x)$  と  $x^{-\varepsilon} l(x)$  があるところから先でそれぞれ増加及び減少関数になることとは同値である．

指数が正の正則変動関数は漸近単調増加であるから，漸近的な意味での逆関数を考えられる．

**定理 1.2.4 (漸近的逆関数)**  $f(x) \in \mathbf{R}_\alpha$  ( $\alpha > 0$ ) ならば，

$$f(g(x)) \sim g(f(x)) \sim x$$

を満たす  $g(x) \in \mathbf{R}_{1/\alpha}$  が存在する．この  $g(x)$  を  $f(x)$  の漸近的逆関数 (*asymptotic inverse*) と呼び，漸近的に唯ひとつに決まる．具体的に，以下の一般化された逆関数をとることが出来る．

$$f^-(x) = \inf\{y \in [a, \infty) : f(y) > x\}.$$

**定理 1.1.2 (一様収束定理)** 緩慢変動関数  $l(x)$  の収束,

$$\lim_{x \rightarrow \infty} \frac{l(\lambda x)}{l(x)} = 1$$

は,  $\lambda \in (0, \infty)$  について広義一様収束する ( $(0, \infty)$  の任意のコンパクト集合上で一様収束する).

**定理 1.1.3 (緩慢変動関数の表現定理)**  $l(x)$  が緩慢変動関数であることは次の形に書けることと同値である.

$$l(x) = c(x) \exp\left\{\int_a^x \varepsilon(t) dt/t\right\}, \quad x \geq a.$$

ここで,  $a > 0$ ,  $c(x)$ ,  $\varepsilon(x)$  はそれぞれ  $\lim_{x \rightarrow \infty} c(x) = c > 0$ ,  $\lim_{x \rightarrow \infty} \varepsilon(x) = 0$  となる可測関数. さらに  $\varepsilon(x)$  は連続になるように取ることが出来る. この  $\varepsilon(x)$  を  $\varepsilon$  関数と呼ぶ.

$c(x) = c > 0$  の場合, 正規化緩慢変動関数 (normalized slowly varying function) という.

**例 1.1.3**

$$\begin{aligned} \log x &= \exp\left\{\int_e^x (\log t)^{-1} dt/t\right\}, \\ \exp(\log x)^\alpha &= \exp\left\{\int_1^x \alpha (\log t)^{1-\alpha} dt/t\right\}, \quad (0 < \alpha < 1). \end{aligned}$$

**注意 1.1.4** 1. この表現で, 仮に  $\varepsilon(x) = \rho \neq 0$  とすれば,  $\rho$  のべき乗関数となり,  $\rho + \varepsilon(x)$  とすれば, 正則変動関数の表現定理となる.

2.  $\int_a^\infty du/u = \infty$  に注意して, この表現定理をみれば,  $\varepsilon(x)$  の符号を変えながら適当に取ることによって, 緩慢変動関数の上下限が自由に取れることがわかるであろう.

3. 極端に大きい事象が起こる確率を表す分布の裾  $F(x, \infty)$  が正規化緩慢変動するならば,  $\varepsilon(x)/x$  が危険度関数 (2.2 で定義) になる.

4. 漸近的な性質に関するだけならば, 正規化緩慢変動関数だけをみれば十分だが, 一般には必ずしもそうは問屋が卸さない.

緩慢変動関数に対する定理 1.1.2 は一般の正則変動関数へ拡張される.

**定理 1.1.4 (正則変動関数に対する一様収束定理)**  $f(x) \in \mathbf{R}_\rho$  とするとき ( $\rho > 0$  のときは任意の  $a$  に対して,  $[0, a)$  上で有界を仮定), 収束

$$\lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho$$

は, それぞれ以下の区間において  $\lambda$  について一様.

(i)  $\rho = 0$  の場合,  $[a, b]$  ( $0 < a \leq b < \infty$ ),

(ii)  $\rho > 0$  の場合,  $(0, b]$  ( $0 < b < \infty$ ),

(iii)  $\rho < 0$  の場合,  $[a, \infty)$  ( $0 < a < \infty$ ).

逆に緩慢変動関数でない例も挙げる。べき乗関数より速く発散するものや、速く振動するようなものがある。

例 1.1.2 以下は正則変動関数ではない。

1.  $l(x) = \exp x$ .

2.  $l(x) = 2 + \sin x$ .

緩慢変動関数は様々な演算について閉じている。

命題 1.1.1  $l(x)$ ,  $l_1(x)$ ,  $l_2(x)$  を緩慢変動関数とする。

(i)  $\alpha \in \mathbb{R}$  に対して,  $l(x)^\alpha$  は緩慢変動関数である。

(ii) 積  $l_1(x)l_2(x)$ , 和  $l_1(x) + l_2(x)$ , さらに  $\lim_{x \rightarrow \infty} l_2(x) = \infty$  ならば, 合成  $l_1(l_2(x))$  も緩慢変動関数である。

(iii) 任意の  $\varepsilon > 0$  に対し,  $\lim_{x \rightarrow \infty} x^\varepsilon l(x) = \infty$   $\lim_{x \rightarrow \infty} x^{-\varepsilon} l(x) = 0$ .

(iv)  $\lim_{x \rightarrow \infty} \log l(x) / \log x = 0$ .

注意 1.1.3 (iii) の性質は緩慢変動関数がべき乗関数より漸近的に小さいことを主張しているが、後に示すように (定理 1.2.2), この性質こそ緩慢変動性そのものであるといえる。

正則変動関数がどのようなものかがある程度がつかめたところで, そのイメージをより明確にするべく, 基本となる定理について述べる。中にはいかにも数学的というものもあるが, いずれも重要なものである。

最初のもは, 正則変動性の定義に関連するもので, その広さについての定理である。

$f(x)$  が与えられているものとする。  $\lambda > 0$  に対して, 下記左辺の極限が存在するとき, それを  $g(\lambda)$  とおく。

$$(1.2) \quad \lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = g(\lambda) \in (0, \infty).$$

正則変動関数の定義では, (1.1.2) の極限関数  $g(\lambda)$  がべき乗関数であった。実は, 極限関数は十分に緩い条件の下でこれに限るのである。

定理 1.1.1 (特徴付定理) 正值可測関数  $f(x)$  に対し, (1.1.2) が成り立つ  $\lambda$  の集合が正の測度をもつならば, 次がいえる。

(i) (1.1.2) がすべての  $\lambda > 0$  に対して成り立つ。

(ii) 実数  $\rho$  により,  $g(\lambda) = \lambda^\rho$  とかける。

(iii)  $f(x)$  は指数  $\rho$  の正則変動関数である。

以下のふたつの定理は, 外見こそ異なるが, ほぼ同等といってもよい。特に, 後の表現定理は一般の緩慢変動関数がどのようなものなのかを感覚的に理解するのに役立つと思われる。

## 1.5 アーベル型定理

アーベル型定理 (Abelian Theorem) とは、タウバー型定理とは逆に、被積分関数に仮定を付して、積分の漸近的性質を導くという類の定理のことである。節 1.3 でカラマタの定理として正則変動関数の積分の漸近挙動を扱ったが、これを一般化した、分かりやすい例から始める。  $l(x)$  を緩慢変動関数、  $k(t)$  が  $0 < a \leq b < \infty$  に対し、  $[a, b]$  上可積分であるならば、一様収束定理 (定理 1.1.2) により、  $x \rightarrow \infty$  のとき、

$$\int_a^b k(t)l(xt)dt \sim l(x) \int_a^b k(t)dt$$

が成り立つ。  $\int_a^b k(t)dt = 0$  のときは右辺は  $o(l(x))$  の意。

このようなことを  $k(t)$  に適当な条件をつけることで、  $a = 0$  や  $b = \infty$  の場合に拡張することが目的である。

より具体的な例として、指数的な裾をもつ確率分布の合成積 (たたみこみ, convolution, 確率変数の独立和の分布) の裾の挙動を計算しよう。

**例 1.5.1 (独立和の分布の計算)**  $[0, \infty)$  上の分布  $F$  と  $G$  は指数的な裾を持つものとする。すなわち、

$$\begin{aligned}\bar{F}(x) &\sim e^{-\alpha x} l_1(e^x), \\ \bar{G}(x) &\sim e^{-\beta x} l_2(e^x).\end{aligned}$$

ここで、  $0 \leq \alpha < \beta < \infty$ ,  $l_1(x), l_2(x) \in \mathbf{R}_0$  である。

$F$  と  $G$  の合成積  $F * G$  の裾の挙動は、

$$\begin{aligned}\overline{F * G}(x) &= \int_0^x \bar{F}(x-t)G(dt) + \bar{G}(x) \\ &\sim \int_0^\infty e^{at}G(dt)\bar{F}(x).\end{aligned}$$

次の命題は、それぞれ原点及び無限遠で  $k(t)$  に条件をつけたときの結果である。

**命題 1.5.1**  $a, \delta > 0$  とする。

(i)  $\int_0^a t^{-\delta}|k(t)|dt < \infty$  ならば、  $[0, \infty)$  上で定義され、  $x^\delta l(x)$  が局所有界となる  $l(x) \in \mathbf{R}_0$  に対して、

$$\int_0^a k(t)l(xt)dt \sim l(x) \int_0^a k(t)dt \quad (x \rightarrow \infty).$$

(ii)  $\int_a^\infty t^\delta|k(t)|dt < \infty$  ならば、  $l(x) \in \mathbf{R}$  に対して、

$$\int_a^\infty k(t)l(xt)dt \sim l(x) \int_a^\infty k(t)dt \quad (x \rightarrow \infty).$$

**定理 1.5.1**  $\delta > 0$ ,  $|\sigma| \leq \delta$  に対し、  $\int_0^\infty t^\sigma|k(t)|dt < \infty$  であれば、  $l(x) \in \mathbf{R}_0^*$  に対して、

$$\int_0^\infty k(t)l(xt)dt \sim l(x) \int_0^\infty k(t)dt \quad (x \rightarrow \infty).$$

ここで、  $\mathbf{R}_0^*$  は  $(0, \infty)$  上で定義され、局所有界かつ原点近傍でも有界な緩慢変動関数の集合である。

注意 1.5.1  $k(t)$  に対する仮定は次のようにも書ける.

$$\int_0^x |k(t)| dt = O(x^\eta) \quad (x \rightarrow 0+),$$

$$\int_x^\infty |k(t)| dt = O(x^{-\eta}) \quad (x \rightarrow \infty)$$

となるような  $\eta > 0$  が存在する.

## 1.6 $\Gamma$ 変動性と $\Pi$ 変動性

正則変動関数論の発展に従い, 派生する形で多くの概念が生まれてきた. ここで紹介する  $\Gamma$ -変動と  $\Pi$ -変動は代表的なものであり, 本稿では極値分布の吸引域の特徴付けに用いられる.

**定義 1.6.1** 非減少関数  $V(x)$  が  $\Gamma$  変動 ( $\Gamma$ -varying,  $V \in \Gamma$ ) するとは,  $V(x)$  が  $(a, x_0)$  上で定義され,  $\lim_{x \uparrow x_0} V(x) = \infty$ ,  $(a, x_0)$  上で定義された正值関数  $f(x)$  があって, 任意の  $x$  に対して,

$$\frac{V(t + xf(t))}{V(t)} \rightarrow e^x \quad (t \rightarrow x_0).$$

となるときをいう.  $f(x)$  を補助関数 (*auxiliary function*) といい, 漸近的な意味で一意に決まる.

**例 1.6.1** 1.  $x_0 = \infty$ ,  $l(x) \in \mathbf{R}_0$  とするとき,  $V(x) = e^{\rho x} l(e^x)$  は補助関数  $f(t) = \rho^{-1}$  の  $\Gamma$  変動関数である.

2.  $x_0 = 1$  とするとき,  $V(x) = \exp(x/(1-x))$  は補助関数  $f(t) = (1-t)^2$  の  $\Gamma$  変動関数である.

**定義 1.6.2** 非負非減少関数  $U(x)$  が  $\Pi$  変動 ( $\Pi$ -varying,  $U \in \Pi$ ) するとは, 半直線  $(z, \infty)$  上で定義され, 関数  $a(t) \in \mathbf{R}$  と  $b(t) > 0$  が存在し,

$$\frac{U(tx) - a(t)}{b(t)} \rightarrow \log x.$$

$b(x)$  を補助関数といい, 漸近的な意味で一意に決まる.

注意 1.6.1  $a(t) = U(t)$ ,  $b(t) = U(te) - U(t)$  とすればよい.

$\Pi$  と  $\Gamma$  の間には次の関係がある.

**命題 1.6.1** (i)  $V(x) \in \Gamma$  でその補助関数が  $f(t)$  ならば,  $V^{\leftarrow}(x) \in \Pi$  であり, 補助関数は  $b(t) = f \circ V^{\leftarrow}(t)$  である.

(ii)  $U(x) \in \Pi$  でその補助関数が  $b(t)$  ならば,  $U^{\leftarrow}(x) \in \Gamma$  であり, 補助関数は  $f(t) = b \circ U^{\leftarrow}(t)$  である.



## 2 正則変動性と確率統計

この章では、正則変動関数が確率や統計においてどのように応用されているか、その代表例を紹介する。正則変動性は漸近的な性質であるから、主に極限定理を記述する際に使われる。言葉の定義をしよう。確率分布  $F$  に対して、 $F(-\infty, x]$  を  $F$  の左裾 (或いは負側の裾)、 $F(x, \infty)$  を右裾 (或いは正側の裾) とよび、(右) 裾 ((right) tail) を

$$\bar{F}(x) = 1 - F(x)$$

であらわす。確率分布  $F$  の  $\sigma$  次の切断積率 ( $\sigma$ th truncated moment) を

$$V_\sigma(x) = \int_{|y| < x} |t|^\sigma F(dt)$$

で定義する。特に、 $\sigma = 2$  のとき、切断分散 (truncated variance) とよぶ。

### 2.1 安定分布と吸引域

1次元独立同分布確率変数列に対する中心極限定理から始めて、その概念を拡張し、一般化していくことで正規分布の代わりに安定分布と呼ばれるものが極限分布として現われることを説明し、正則変動性とのかかわりについて述べる。

この章を通じて、 $\{X, X_i: i = 1, 2, \dots\}$  を実数値独立同分布確率変数列とし、 $X$  の分布は退化 (一点のみへの集中) していないものとする。

**定理 2.1.1 (中心極限定理)** 分散が有限であれば、次のように正規化された独立部分  $S_n = \sum_{i=1}^n X_i$  の分布は、正規分布に近づく。

$$\frac{S_n - nEX}{\sqrt{nVX}} \rightarrow N(0, 1).$$

ここで、 $\mathbf{E}$ 、 $\mathbf{V}$  はそれぞれ確率変数の平均と分散を、 $N(0, 1)$  は平均 0 分散 1 の正規分布を表し、収束は分布収束である。

以降、この定理の独立同分布 (i.i.d.) の仮定の下での一般化を考えていく。一般化とは分散が無限の場合の独立同分布確率変数の和の挙動を考えるという意味であり、和を線形変換により正規化して、なんらかの (非退化) 分布に収束させるということである。もちろん、分散無限の場合のすべてを扱うことが出来るわけではないが、無限への行き方によっては、中心極限定理に似た形の極限定理が成り立つことがある。尚、この場合の正規化 (normalize) は線形変換により見やすいスケールに直すという意味で正規分布とは無関係である。

まず、収束先を正規分布に限定したままでの一般化を考える。たとえ分散が発散しても、その発散の仕方が“有限”に準ずるものであれば、正規化定数を取り直すことで、中心極限定理に準ずる極限定理が成り立つ。

**定理 2.1.2** ある定数列  $B_n$  が存在して、

$$\frac{S_n - nEX}{B_n} \rightarrow N(0, 1)$$

となるための必要十分条件は、切断分散  $V(x) = \mathbf{E}X^2 \mathbf{1}(|X| < x)$  が緩慢変動することである。このとき、正規化定数  $B_n$  は漸近増加な緩慢変動関数  $l(x)$  により、 $B_n = \sqrt{nl(n)}$  と表すことが出来る。

**注意 2.1.1** 1.  $l(x)$  が漸近増加とは、増加関数  $\tilde{l}(x)$  があって、 $l(x) \sim \tilde{l}(x)$  となることをいう。分散有限の場合は、前定理で  $l(x) = \sqrt{Vx}$  (定数) ととれるが、分散無限の場合は、裾が分散有限の場合より重いから、対応する正規化定数も呼応して大きくなるであろうと考えれば、この結果は自然なものと理解できよう。

2. 次も同値な条件である。

$$\lim_{x \rightarrow \infty} x^2 \{F(-\infty, -x] + F(x, \infty)\} / V(x) = 0.$$

中心極限定理は、収束先を非正規な確率分布まで広げることで、さらに一般化することが出来る。次に考えるのは、独立同分布確率変数の和を正規化したものの極限分布であり、安定分布と呼ばれるものがそれに当たる。先に安定分布について説明する。

**定義 2.1.1** 分布  $G$  が安定分布であるとは、確率変数  $X$  と  $Y$  が独立で共に分布  $G$  をもつとき、任意の  $a_1, a_2 > 0$  に対し、 $b > 0$  と  $c \in \mathbb{R}$  が存在して、 $a_1X + a_2Y$  と  $bX + c$  の分布が等しくなるときをいう。このとき、 $\alpha (0 < \alpha \leq 2)$  が唯一つ存在し、全ての  $a_1, a_2, b, c$  に対して、

$$b^\alpha = a_1^\alpha + a_2^\alpha$$

となる。この  $\alpha$  を安定分布  $G$  の指数という。

**例 2.1.1** 次の確率分布は安定分布である。

1. 正規分布 (指数は 2) .
2. コーシー分布 (指数は 1) .
3. 次の密度関数を持つ分布は指数 1/2 の (片側) 安定分布である。

$$p(x) = (2\pi)^{-1/2} e^{-1/(2x)} x^{-3/2}.$$

安定分布は無限分解可能分布であるから、特性関数は Lévy-Khintchine の標準形をもつ。

**定理 2.1.3 (安定分布の特性関数)** 指数  $\alpha (0 < \alpha < 2)$  の安定分布の特性関数  $\hat{G}(z)$  は次のようにあらわされる。

$$(2.1) \quad \log \hat{G}(z) = i\gamma z + c_1 \int_{-\infty}^0 \left\{ \frac{dt}{|t|^{1+\alpha}} \right\} \\ + c_2 \int_0^{\infty} \left\{ e^{izt} - 1 - \frac{izt}{1+t^2} \right\} \frac{dt}{t^{1+\alpha}}$$

ここで、 $\gamma \in \mathbb{R}, c_1, c_2 \geq 0, 0 < \alpha < 2$ .

話を独立確率変数の和に戻そう。次が成り立つ。

**定理 2.1.4**  $\{X_i: i = 1, \dots\}$  を分布  $F$  に従う実数値独立同分布確率変数列とする。定数列  $B_n > 0$  と  $A_n \in \mathbb{R}$  により、正規化された部分和が (非退化) 分布  $G$  に収束するものとする。

$$(2.2) \quad \frac{S_n - A_n}{B_n} \rightarrow G.$$

このとき、極限分布  $G$  は安定分布となる。また、分布  $F$  は安定分布  $G$  に吸引されるといい、吸引される分布の全体を安定分布  $G$  の吸引域 (domain of attraction) という。  $G$  自身がその吸引域に含まれることから、全ての安定分布は空でない吸引域をもつ。

このときの正規化定数  $B_n$  は、緩慢変動関数  $l(x)$  により、

$$(2.3) \quad B_n = n^{1/\alpha} l(n)$$

と書くことが出来る。

定理 2.1.2 では、正規分布の吸引域に属するための必要条件を与えたことになる。次に非正規の安定分布  $G$  をひとつ選んだとき、確率分布  $F$  がその吸引域に入るための必要十分条件を与える。重要なのは、原点近傍ではなく、 $F$  の裾の遠方での挙動である。

**定理 2.1.5** 分布  $F$  が、特性関数 (2.1.1) をもつ安定分布  $G$  ( $0 < \alpha < 2$ ) の吸引域に含まれるための必要十分条件は、次のふたつが成り立つことである。

$$(i) \quad \lim_{x \rightarrow \infty} F(-\infty, -x] / F(x, \infty) = c_1 / c_2,$$

(ii) 任意の  $\lambda > 0$  に対して、

$$\lim_{x \rightarrow \infty} (F(-\infty, -x] + F(x, \infty)) / (F(-\infty, -\lambda x] + F(\lambda x, \infty)) = \lambda^\alpha.$$

これを正則変動関数の言葉で言い直せば、正負の裾の小さくない方が正則変動性をもち、他方が無視できるほど小さくなければ、その挙動もまた定数倍を除いた意味で同じ正則変動性をもつということになる。小さくない裾の方の正則変動性の指数は収束先の安定分布の指数のマイナスとなる。

**注意 2.1.2** 正規分布の場合と異なり、 $0 < \alpha < 2$  の場合は、(2.1.3) の正規化定数に表れる緩慢変動関数  $l(x)$  は単調なものに限定されたりはせず、本質的にあらゆるものが現われる。

**注意 2.1.3** 以上の例が示すように正則変動関数が確率統計に登場する場合、切断分散や裾などの性質であることが多く、必然的に単調なものになる。このとき、指数が 0 (緩慢変動) でなければ、漸近的な意味で正則変動関数は制限されないが、指数が 0 であれば、単調な緩慢変動関数となり、緩慢変動関数全体の中でかなり限られたものになる。

裾が緩慢変動する場合は部分和を線形変換による正規化で非退化分布に収束させることは出来ないが、次のような極限定理が成り立つことが知られている。

定理 2.1.6  $X \geq 0$  とする.  $F(x, \infty)$  が緩慢変動するならば,

$$\lim_{x \rightarrow \infty} P(n \cdot F(S_n, \infty) \geq x) = e^{-x}, \quad x > 0.$$

独立確率変数の部分和については次のようなことも知られている.

定義 2.1.2 確率分布  $F$  ないし対応する乱歩  $S_n$  が相対的安定 (*relatively stable*) であるとは,  $S_n/B_n$  が 1 に確率収束, 即ち, 任意の  $\varepsilon > 0$  に対して,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{B_n} - 1\right| > \varepsilon\right) = 0$$

となるような正規化定数  $B_n$  が取れるときをいう.

例 1.3.1 のペーター・ポール分布では,

$$S_n / \{n(\log n) / \log 2\} \rightarrow 1 \text{ in probability.}$$

が成り立つ. 一般的には, 次がいえ.

定理 2.1.7  $[0, \infty)$  上の確率分布  $F$  に対して, 次は同値である.

- (i)  $S_n$  が相対的安定である.
- (ii)  $\int_0^x t dF(dt)$  が緩慢変動する.
- (iii)  $\int_0^x \{1 - F(t)\} dt$  が緩慢変動する.
- (iv) ある緩慢変動関数  $l(x)$  に対し,  $1 - \hat{F}(s) \sim sl(1/s)$ .

## 2.1 節の補足

この章で扱った確率論に関する事柄は馴染み深いものではないかもしれないので, 基本的な知識を簡潔にまとめておくとともに, 関連事項を補足しておく.

無限分解可能分布とは, 合成積の意味で任意の数の同分布に分解できる分布である. すなわち,  $G$  が無限分解可能分布であるとは, 任意の自然数  $n$  に対して, 分布  $F$  が

$G = \overbrace{F * \dots * F}^{n \text{ 回}}$  (右辺は  $F$  の  $n$  回の合成積の意) となるように取れるときをいう. 確率変数の言葉でいえば, 任意に与えられた  $n$  に対して, (分布  $F$  に従う  $n$  個の) 独立同分布確率変数列  $\{X_i: i = 1, \dots, n\}$  の和  $\sum_{i=1}^n X_i$  の分布が  $G$  となるように取れることに相当する.

よく知られた分布のうち, 正規分布, ポアソン分布, 複合ポアソン分布, コーシー分布, 指数分布, 幾何分布, 負の二項分布に加え, Student の  $t$  分布,  $F$  分布, 対数正規分布, ロジスティック分布も無限分解可能分布である. 一方, 一様分布, 二項分布は無限分解可能分布でない.

無限分解可能分布を  $G$  としたとき, その特性関数 (Fourier 変換) はよく知られた Lévy-Khintchine の標準形で表され, 3つの特性量  $(\gamma, a, \nu)$  で特徴付けられる.

$$\int_{-\infty}^{\infty} e^{itz} G(dt) = \exp\left\{i\gamma z - \frac{1}{2}az^2 + \int_{-\infty}^{\infty} \left(e^{izt} - 1 - \frac{izt}{1+t^2}\right) \nu(dt)\right\}$$

ここで、 $\gamma \in \mathbb{R}$ ,  $a \geq 0$ ,  $\nu$  は  $\nu(\{0\}) = 0$ ,  $\int_{-\infty}^{\infty} \frac{t^2}{1+t^2} \nu(dt) < \infty$  を満たす  $\mathbb{R}$  上の測度で、Lévy 測度と呼ばれる。これらは  $\nu$  から一意的に定まり、逆に、上記条件を満たす  $\gamma, a, \nu$  に対して、対応する無限分解可能分布  $G$  がただひとつ存在する。

- 注意 2.1.4**
- 3要素のうち、正規分布を表す  $a$  と Lévy 測度  $\nu$  は内在的な意味を持つが、 $\gamma$  は必ずしもそうではない。Lévy の標準形の  $\frac{izt}{1+t^2}$  は補正項に当たり、一定の性質を満たせば、この形に限らず、この項の取り方によって、 $\gamma$  もまた変わりうるからである。その意味で、Lévy の標準形はひとつでなく、場合に応じて、補正項の違った、或いは三角関数を用いた表現が用いられる（もちろん、本質的には一つ）。ただし、Lévy 測度により、 $\gamma$  が内在的な意味を持つこともある。
  - 安定分布に関して既に触れたように、個々の無限分解可能分布を特徴付けるには、特性関数の Lévy の標準形における3要素を決めればよく、これが重要な問題となる。上述の具体例についても3要素が容易に求められるものもあるが、そうではないもの、従って、無限分解可能かどうか判定が難しいものもある。

確率論にあらわれる正則変動関数ということで中心極限定理の一般化から安定分布について触れたが、さらなる一般化についても研究されている。同分布という制限を外せば、極限分布は、L 分布（自己分解可能分布）というものになるし、無限分解可能分布さえも null array の和の極限分布として現われる。

一般に安定分布は、指数が、 $0 < \alpha < 1$ ,  $\alpha = 1$ ,  $1 < \alpha < 2$ ,  $\alpha = 2$  により性質が大きく異なる。安定分布は指数未満のモーメントはもつが（有限）、その指数以上のモーメントは発散する。(2.1.2) における正規化定数  $A_n$  は、 $1 < \alpha \leq 2$  では平均の調整に過ぎず（ $\alpha = 2$  のときは直接その形で書いた）、 $0 < \alpha < 1$  の場合も  $A_n/B_n$  が定数にいくだけである。 $A_n = 0$  とし、 $B_n$  だけで正規化すると極限分布は狭義安定分布と呼ばれる安定分布の部分族になる（定義 2.1.1 で  $c = 0$  とした場合に相当）。 $\alpha \neq 1$  の場合、狭義安定分布と一般の安定分布の違いは単に平行移動のみであり、本質的に違いはないといえる。一方、 $\alpha = 1$  の場合は、コーシー分布が唯一の狭義安定分布（非退化、一次元）であり、本質的な制限が加わる。

## 2.2 極値分布と吸引域

極値理論というのは、極端なデータ（例えば観測値の最大値など）がどのようなふるまいをするかを議論するのが中心テーマである。自然災害や破壊などの問題に加え、保険や金融など経済の分野への応用で近年幅広い関心を集めている。

**定義 2.2.1**  $X_1, X_2, \dots$  を共通の確率分布  $F$  に従う実数値独立確率変数列とし、 $X_n$  までの最大値を

$$M_n = \max\{X_1, \dots, X_n\}$$

とおく。 $M_n$  を定数列  $A_n, B_n$  により正規化し、 $n \rightarrow \infty$  としたとき、非退化分布  $G$  へ収束するならば、極限の分布  $G$  を極値分布という。

$$\frac{M_n - A_n}{B_n} \rightarrow G \quad (n \rightarrow \infty).$$

このとき、 $F$ は $G$ に吸引されるといい、そのような分布 $F$ の全体を $G$ の吸引域 ( $D(G)$ と書く) とよぶ (*Maximum domain of attraction*).

極値分布は明確に与えられている.

**定理 2.2.1 (Fisher-Tippett の定理)** 極値分布は次の3種のタイプのいずれかである. それぞれ, フレシェ (*Fréchet*) 分布, ワイブル (*Weibull*) 分布, グンベル (*Gumbel*) 分布と呼ばれる. 以下,  $\alpha > 0$  とする.

(i)

$$\Phi_\alpha(x) = \begin{cases} 0 & (x \leq 0) \\ \exp(-x^{-\alpha}) & (x > 0) \end{cases}$$

(ii)

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & (x \leq 0) \\ 1 & (x > 0) \end{cases}$$

(iii)

$$\Lambda(x) = \exp(-e^{-x}).$$

**注意 2.2.1** タイプ同値 (線形変換による同値性) の意味で3種という意味である.

3種それぞれの極値分布の吸引域に入る分布の具体例をあげる. 大雑把な特徴として,  $D(\Phi_1)$  はいくらでも大きい値をとりうる (右裾が常に正であり), 逆に,  $D(\Psi_1)$  に属する分布は取る値に限界がある (分布の台が右に有界). これらに対し,  $D(\Lambda)$  に属する分布はその両方の場合がある.

**例 2.2.1** 1. コーシー分布は  $D(\Phi_1)$  に含まれる.

2.  $(0, 1)$  上の一様分布は  $D(\Psi_1)$  に含まれる.

3. 指数分布は  $D(\Lambda)$  に含まれる.

4. 次の分布は  $D(\Lambda)$  に含まれる.

$$F(x) = 1 - \exp\left(\frac{-x}{1-x}\right) \quad (0 \leq x < 1).$$

以下, 3種の極値分布それぞれの吸引域に入るための必要十分条件について述べる. 吸引域は裾の減衰の仕方, 台の右端への近づき方によって特徴付けがなされる. フレシェ分布とワイブル分布の吸引域は, 正則変動性により比較的簡単に表現されるが, グンベル分布に対しては,  $\Gamma$ -変動性が用いられる.

**定理 2.2.2**  $F \in D(\Phi_\alpha)$  となるための必要十分条件は  $\bar{F}(x) \in \mathbf{R}_{-\alpha}$  である. 正規化定数は,  $A_n = 0, B_n = \inf\{x : \bar{F}(x) \leq 1/n\}$  ととることができる.

**定理 2.2.3**  $F \in D(\Psi_\alpha)$  となるための必要十分条件は  $x_+(F) = \sup\{x : F(x) < 1\}$  ( $F$  の上端点) が有限かつ  $\bar{F}(x_+ - 1/x) \in \mathbf{R}_{-\alpha}$  である. 正規化定数は,  $A_n = x_+$ ,  $B_n = \sup\{x : \bar{F}(x_+ - x) \leq 1/n\}$  ととることができる.

$D(\Lambda)$  を記述するための記号を準備する.

$$H(x) = -\log \bar{F}(x),$$

とし, その (右連続) 逆関数を  $H^-(x) = \inf\{s : H(s) \geq x\}$  とする. さらに,  $U(x) = (1/\bar{F}(x))^-$  とおく.  $x_+$  は  $F$  の上端点を表す.

**定理 2.2.4** 次はいずれも  $F \in D(\Lambda)$  となるための必要十分条件である.

(i)  $U(x) \in \Pi$ , すなわち, 任意の  $x, y > 0$ ,  $y \neq 1$  に対して,

$$\frac{U(tx) - U(t)}{U(ty) - U(t)} \rightarrow \frac{\log x}{\log y} \quad (t \rightarrow \infty).$$

(ii)  $1/\bar{F}(x) \in \Gamma$ , すなわち, ある正値関数  $f(x)$  に対して,

$$\frac{\bar{F}(t + xf(t))}{\bar{F}(t)} \rightarrow e^{-x} \quad (t \uparrow x_+).$$

(iii) ある緩慢変動関数  $l(x)$  があって, 任意の  $u > 0$  に対して,

$$H^-(x+u) - H^-(x) \sim ul(e^x) \quad (x \rightarrow \infty).$$

**注意 2.2.2** 1. (ii) で  $f(t) = \int_t^{x_+} \bar{F}(s) ds / \bar{F}(t)$ , ( $t < x_+$ ) ととることができる.

2. 正規化定数は次のようにとることができる.

$$\begin{aligned} A_n &= H^-(\log n), \\ B_n &= H^-(\log(ne)) - H^-(\log n). \end{aligned}$$

**注意 2.2.3** 正規分布, ガンマ分布, ワイブル分布, 対数正規分布等多くの分布がゲンベル分布の吸引域に入る. 一方, ポアソン分布や幾何分布などの離散分布はどの極値分布の吸引域にも入らない.

吸引域に入るための必要十分条件に近く, かつ簡単な, Von Mises による十分条件があるので紹介しよう.

**定義 2.2.2 (危険度関数)** 分布  $F$  が密度  $f(x)$  をもつとき,

$$h(x) = \frac{f(x)}{1 - F(x)} = H'(x)$$

を  $F$  の危険度関数 (hazard function) という.

危険度関数を用いた極値分布の吸引域に含まれるための十分条件を与える。

**定理 2.2.5 (Von Mises の条件)** (i)  $x_+ = \infty$  かつ  $xh(x) \rightarrow \alpha > 0$  ならば,  $F \in D(\Phi_\alpha)$  である.

(ii)  $x_+ < \infty$  かつ  $(x_+ - x)h(x) \rightarrow \alpha > 0$  ( $x \uparrow x_+$ ) ならば,  $F \in D(\Psi_\alpha)$  である.

(iii)  $x_+ = \infty$  かつ  $\frac{d}{dx}(1/h(x)) \rightarrow 0$  ならば,  $F \in D(\Lambda)$  である.

(iv)

$$\lim_{x \uparrow x_+} f(x) \int_x^{x_+} \bar{F}(t) dt / (\bar{F}(x))^2 = 1$$

ならば,  $F \in D(\Lambda)$  である.

### 2.3 Subexponential 分布と関係する分布族

Subexponentiality は分枝過程における粒子の寿命の分布として 1960 年代に Chistyakov により導入された概念であり ([Ch64]), 近年裾が重い確率分布 (heavy 或いは fat tail) の代表的なものとして, 破産問題, 保険を想定した確率や統計のモデルによく用いられるようになってきた ([EKM], [As00]).

**定義 2.3.1** 確率分布  $F$  が  $\bar{F}(x) \in \mathbf{R}$  を満たすとき, 正則変動する裾をもつという.

**例 2.3.1** 1. パレート分布.  $\bar{F}(x) = (\alpha/x)^\beta$  ( $\alpha, \beta > 0$ ).

2. 片側安定分布 ( $0 < \alpha < 1$ ) 及びその吸引域に属する分布. ( $1 \leq \alpha < 2$  の場合も正負の裾の和は正則変動する (定理 2.1.5)).

**命題 2.3.1** 確率分布  $F$  と  $G$  がともに正則変動する裾を持つならば,

$$\overline{F * G}(x) \sim \bar{F}(x) + \bar{G}(x)$$

となる. 従って, 任意の自然数  $n$  に対して,

$$\overline{F^{n*}}(x) \sim n\bar{F}(x)$$

である.

正則変動関数は和について閉じていることから (命題 1.1.1), 正則変動する裾をもつ分布全体は合成積について閉じていることがわかる.

$$\bar{F}(x), \bar{G}(x) \in \mathbf{R} \Rightarrow \overline{F * G}(x) \in \mathbf{R}.$$

同じことは正則変動の指数を固定した分布の全体についてもいえる.

**定義 2.3.2**  $F$  を  $[0, \infty)$  上の分布とする.



(i) 任意の  $k \in \mathbb{R}$  に対して,

$$\bar{F}(x+k) \sim \bar{F}(x)$$

となるとき,  $F$  を *long-tailed* 分布という. その全体を  $\mathcal{L}$  であらわす.

(ii)  $F$  が指数  $\gamma (> 0)$  の *exponential tail* をもつとは, 任意の  $k \in \mathbb{R}$  に対して,

$$\bar{F}(x+k) \sim e^{-k\gamma} \bar{F}(x)$$

となるときをいい, その全体を  $\mathcal{L}(\gamma)$  であらわす.  $\gamma = 0$  のときが  $\mathcal{L}$  に相当し,  $\mathcal{L}(\gamma) = \mathcal{L}$  と書く.

**注意 2.3.1** *Long-tailed* 及び *exponential tail* をもつ分布は緩慢変動関数  $l(x)$  により,

$$\bar{F}(x) = e^{-\gamma x} l(e^x)$$

と表現できる分布である. つまり, 裾が指数関数と正則変動関数の合成関数となっている. 従って, *long-tailed* 分布は任意の指数分布よりも裾が重いことがわかる (定理 1.2.2). このような簡単な変数変換によって正則変動の理論の応用範囲が格段に広がる.

**定義 2.3.3 (Subexponential class)**  $[0, \infty)$  上の分布  $F$  が

$$\frac{\overline{F * F}(x)}{\bar{F}(x)} \sim 2$$

を満たすとき,  $F$  を *subexponential* 分布という. その全体 *subexponential class* を  $S$  であらわす.

**定義 2.3.4 (Convolution-equivalent class)**  $[0, \infty)$  上の分布  $F$  が *convolution-equivalent* 分布であるとは, 次のふたつの条件を満たすをいい, その全体 *convolution-equivalent class* を  $S(\gamma)$  であらわす.

(i)  $F \in \mathcal{L}(\gamma)$ ,

(ii)  $\overline{F * F}(x) / \bar{F}(x) \sim 2 \int_0^\infty e^{\gamma t} F(dt) < \infty$ .

ここまで定義してきた分布の例をあげる.

**例 2.3.2** (i) 正則変動する裾をもつ分布は, *long-tailed* 分布であり, *subexponential* 分布でもある (命題 2.3.1).

(ii) 指数分布, ガンマ分布は *exponential tail* をもつが, *convolution-equivalent* ではない.

(iii) 逆ガウス分布は *convolution-equivalent* である.

$$p(x) = \frac{d}{\sqrt{2\pi b x^3}} \exp\left(-\frac{(ax-d)^2}{2bx}\right) \quad (x > 0).$$

基本的な結果を紹介する. (i) は *subexponential* 分布が重い分布であることを, (ii) は裾の漸近挙動で決まる性質であることをそれぞれ意味している.

定理 2.3.1 (i)  $S \subset \mathcal{L}$ .

(ii)  $G \in S$  かつ  $\bar{F}(x) \sim \bar{G}(x)$  ならば,  $F \in S$  である.

(iii)  $F \in S$  ならば, 任意の  $\varepsilon > 0$  に対して, 次を満たす  $K(\varepsilon)$  が存在する.

$$\overline{F^{n*}}(x)/\bar{F}(x) \leq K(\varepsilon)(1+\varepsilon)^n \quad (n \in \mathbb{N}, x > 0).$$

意外なことに,  $\mathcal{L}(\gamma)(\gamma \geq 1)$  と  $\mathcal{S}(\gamma)(\gamma \geq 1)$  とは合成積の閉性については異なる性質を持つ.

定理 2.3.2 (i) Long-tailed class  $\mathcal{L}$  及び指数  $\gamma$  の exponential tail をもつ分布族  $\mathcal{L}(\gamma)$  はともに合成積について閉じている.

(ii)  $\mathcal{S}$  及び  $\mathcal{S}(\gamma)$  はともに合成積について閉じてはいない!

注意 2.3.2 1. 具体的な分布が long-tailed であるかの判定は容易だが, subexponential であるかの判定はそうではなく, いくつかの十分条件が知られている. 一方,  $S \subset \mathcal{L}$  ではあるが, subexponential 分布でない long-tailed 分布は直ちに思いつくようなものではない. 意図的に構成した例がいくつかある程度なので, 実用面からいえば, 大半の long-tailed 分布は subexponential 分布とみてよい.

2.  $\mathcal{S}$  及び  $\mathcal{S}(\gamma)$  の非閉性は容易なものではなく, 90 年前後によく証明された. これもかなり特殊な例であるので, 大抵の場合は, subexponential 分布同士の合成積は再び subexponential 分布になるはずである.

最後に subexponential 分布が関連する定理をふたつ紹介する.

無限分解可能分布とその Lévy 測度の裾同士の関係について次の定理がよく知られている.

定理 2.3.3  $G$  を  $[0, \infty)$  上の無限分解可能分布でその Lévy 測度  $\nu$  が任意の  $x > 0$  に対して,  $\nu(x, \infty) > 0$  となるものとする. このとき, 次の3つは同値である.

(i)  $G \in \mathcal{S}$ ,

(ii)  $\nu(1, x)/\nu(1, \infty) \in \mathcal{S}$ ,

(iii)  $\bar{G}(x) \sim \bar{\nu}(x)$ .

例 2.3.3 以下の分布は subexponential な無限分解可能分布である.

1. 片側安定分布 ( $0 < \alpha < 1$ ), パレート分布.
2. ワイブル分布 ( $0 < \alpha < 1$ ).
3. 対数正規分布.

Subexponential 分布は破産問題でも用いられる。今、保険会社は資産  $x$  を持っているとし、保険料を一定量  $c$  の割合で受け取るものとする。支払い時期はパラメーター  $\lambda$  のポアソン過程に従い、その額は独立に  $(0, \infty)$  上の分布  $F$  に従うものとする。  $m = \int_0^\infty tF(dt)$  とする。  $c \leq \lambda m$  ならば、破産確率は 1 であるので、  $c > \lambda m$  とする。破産が起こらない確率を  $R$  とし、  $F$  の裾が重い場合の  $R$  の漸近挙動について次が成り立つ。

**定理 2.3.4**

$$F_I(x) = \int_0^x \{1 - F(t)\} dt / m$$

により *integrated tail distribution* と呼ばれる分布  $F_I$  を定める。

このとき、次の 3 つは同値である。

(i)  $F_I \in \mathcal{S}$ ,

(ii)  $R \in \mathcal{S}$ ,

(iii)  $\bar{R}(x) \sim m\lambda / (c - m\lambda) \bar{F}_I(x) = \int_x^\infty \{1 - F(t)\} dt / (c/\lambda - m)$ .

## A 定理等の引用元など

本稿では定理等の証明を省いたので、代わりに引用元を記しておく。さらに、主に講義中及びその後を受けた質問に基づいたコメントを加えた。

### 第 1 章

正則変動関数論は、カラマタ (Karamata) の論文 [Kr30] に始まるとされる。近い概念としては、以前からあったようだが、特徴付定理、一様収束定理、表現定理といった最初の方で紹介した主要な定理は、いずれも本質的にカラマタの初期の論文で証明されている。カラマタの論文は仏語なので、緩慢を意味する仏語 “lentement” の頭文字から、現在も緩慢変動関数を  $l(x)$  や  $L(x)$  と書くことが多い。

定理 1.1.1 (Characterization Theorem) [BGT] p.17, Th.1.4.1, [Se76] p.9, Th.1.3,

定理 1.1.2 (Uniform Convergence Theorem, UCT) [BGT] p.6, Th.1.2.1, [Se76] p.2, Th.1.1, [BGT] には、この定理は重要だからと証明が 5 通りも書いてある！

定理 1.1.3 (Representation Theorem) [BGT] p.12, Th.1.3.1, [Se76] p.2, Th.2, [Fe71], p.282, Cor., [Re87] p.17, Cor.  $\varepsilon$  関数が連続に取れることは必ずしも書かれていないようだが、任意の  $C^n$  級にとることができる。[Fe71] と [Re87] では表現定理は直接ではなく、積分の漸近挙動についての定理の系として証明されている。

定理 1.1.4 [BGT] p.22, Th.1.5.2, 定理 1.2.1 [BGT] p.23, Th.1.5.3,

定理 1.2.2 [BGT] p.23, Th.1.5.4, 定理 1.2.3 [BGT] p.24, Th.1.5.5,

定理 1.2.4 [BGT] p.28, Th.1.5.12, 定理 1.2.5 [BGT] p.29, Th.1.5.13,

命題 1.2.1 [BGT] p.29, Prop.1.5.14, 定理 1.2.6 [BGT] p.25, Th.1.5.6,

命題 1.3.1 [BGT] p.26, Prop.1.5.8, 1.5.9a,

積分の漸近挙動をカラマタの定理としたが、上記の基本定理もカラマタによる。積分の漸近挙動は [Fe71] で VIII に詳しく書かれている。

命題 1.3.2 [BGT] p.27, Prop.1.5.9b, 1.5.10, 定理 1.3.1 [BGT] p.28, Prop.1.5.11,  
定理 1.3.2 [BGT] p.30, Theorem.1.6.1, 定理 1.3.3 [BGT] p.331,  
Th8.1.2, [Fe71] p.283, Th.2, 例 1.3.1 [BGT] p.372, [Fe71] p.236,

タウバー型定理, アーベル型定理の用語はやや曖昧に使用した. 必ずしも厳密ではないが, 一応の説明をしておきたい. 関数の性質から, その変換の性質を導くタイプのものがアーベル型であり, その逆のタイプのものがタウバー型と呼ばれる. 一般に, タウバー型の方が難しい. 単調密度定理で単調性を仮定したが, このような仮定が必要になることが多い. 双方向が同値というタイプのものもある ([BGT] p.193, p.39 参照). 本稿にあるようなタウバー型定理は [Fe71] の XIII.5 や [Se76] の 2.2 などで扱っている.

例 1.4.1 [BGT] p.334, Cor.8.1.7, p.374, Th.8.8.2,  
定理 1.4.1 [BGT] p.37, Th.1.7.1, [Fe71] p.443, [Se76] p.59, Th.2.3,  
注意 1.4.1 [BGT] p.38, p.274, Th.5.2.4, 定理 1.4.2 [BGT] p.38, Th.1.7.1',  
定理 1.4.3 [BGT] p.39, Th.1.7.2, [Se76] p.60, Th.2.4,  
注意 1.4.2 [Re87] p.21, Prop.0.7(a), 定理 1.4.1 [BGT] p.39, Th.1.7.2b,

例 1.5.1 裾が軽い方は, 指数的な裾に限らず, より広い場合に, 同様のことがなりたつ. 指数が同じ場合はぐっと複雑になり, 様々な場合がある. さらに同分布とすれば, subexponentiality や convolution-equivalence に繋っていく.

命題 1.5.1 [BGT] p.199, Prop.4.1.2, 定理 1.5.1 [BGT] p.199, Th.4.1.3,  
注意 1.5.1 [BGT] p.198, Lem.4.11, 注意 1.6.1 [Re87] p.27,  
命題 1.6.1 [Fe71] p.278, Prop.,

## 第2章

定理 2.1.2 [BGT] p.346, Th.8.3.1(i), [GK68] p.172, Th.1, [Fe71] p.577, Th.2(b),  
定理 2.1.3 [Fe71] p.213, Th.1, 定理 2.1.4 [BGT] p.343,  
定理 2.1.5 [BGT] p.346, Th.8.3.1(ii), [GK68] p.175, Th.2, [Fe71] p.577, Th.2(c),  
定理 2.1.6 [BGT] p.29, Th.1.5.3, [Da52] 定義 2.1.2 [BGT] p.350, p.372,  
定理 2.1.7 [BGT] p.373, Th.8.8.1

極値分布をフレシェ分布, ワイブル分布, グンベル分布の3種と書いた. 統計学辞典では, フレシェ分布は逆数を取って, ワイブル分布に帰着させていて, 言葉として出てこない. もっとも数学辞典には極値理論自体が載っていないが, 海外では極値理論の書籍が応用も含めて多数出ているのと比べると大きな違いを感じる. 統計学辞典を含めグンベル分布は和書ではほとんど“ガ”ンベル分布と表記されている. しかし, グンベルの方が原音に近いらしい. 神戸大学の高橋倫也先生からお聞きしたところでは, この話題がコンファレンスで出たという. Gumbel はドイツ系ユダヤ人で, 参加していたドイツ人研究者もグンベルの方が近いといったそうだ. 本人の著作の翻訳「極値統計学 (極値の理論と工学的応用)」にもグンベルとある. ならば, いつからグンベルの方が優勢になったのだろうか? ちなみに, グンベル分布は, 他にも二重指数分布, 第一種極値分布とも呼ばれる [LLR p.4].

定理 2.2.1 [BGT] p.408, Th.8.13.1, [EKM] p.121, Th.3.2.3, [Re87] p.9, Prop.0.3,  
例 2.2.1 [EKM] p.133, 137, 139,  
定理 2.2.2 [BGT] p.409, Th.8.13.2, [Re87] p.54, Prop.1.11,  
定理 2.2.3 [BGT] p.410, Th.8.13.3, [Re87] p.59, Prop.1.13,

定理 2.2.4 [BGT] p.410, Th.8.13.4, [Re87] p.28, Prop.0.10,

注意 2.2.2 [BGT] p.411, 注意 2.2.3 [EKM] p.118, [LLR] p.26

定理 2.2.5 [BGT] p.411, Th.8.13.5, 8.13.6, p.412, Th.8.13.7, [Re87] p.62, p.64 Prop.1.17

注意 2.3.1 Subexponential を訳せば, 劣指数的となり, 任意の指数分布よりも裾が遅く 0 に近づくことを意味する. ならば, この注意から, long-tailed 分布の方がふさわしい. ちなみに, long-tailed の用語は subexponential ほど一般的ではない.

命題 2.3.1 [Fe71] p.278, Prop., 定理 2.3.1 [EKM] p.41, Lem.1.3.5, p.572, Lem.A3.15,

定理 2.3.2 [EKM] p.581,

注意 2.3.2 [BGT] p.430 (誤植有. 8 行目  $\rightarrow$  は  $\downarrow$ , 10 行目誤植  $yH'(y)$  は  $yH'(x)$ , 16, 17 行目  $/x^\alpha$  は  $x^\alpha$  が正しい), p.432, [EKM] p.581,

定理 2.3.3 [BGT] p.431, Th.A4.1, [EKM] p.581, Th.A3.22,

定理 2.3.4 [BGT] p.431, Th.A4.2,

## B 文献紹介

本稿で扱った正則変動関数及び関連する確率論の話題はこの分野の基礎をなす重要なものであるが, 反面, 正則変動関数論の入口に過ぎない. そこで, 今後の学習や研究の参考のために, 正則変動関数を扱った書籍をいくつか取り上げ, 簡潔な紹介を書く. 本稿の構成もそうだが, 正則変動関数の理論は, 確率論をはじめとする他の数学分野の基礎理論として位置付けられているようだ. そのため, 正則変動関数の理論は, 確率論の本の一部を割くという形で書かれていることが多く, [Fe71], [Ha70], [Re87], [EKM], [HF06] などがある. [Fe71] には本稿にある正則変動関数の理論の多くが書かれている. 元々確率論の教科書であり, 読みやすい. [Ha70] は, 本稿 2.2 で扱った極値分布と吸引域を主題としており, 正則変動性についてはその準備という位置づけで書かれている. [Re87] も, 同じく極値理論の予備知識という観点から簡潔に書かれていて, これも分かりやすいと思う. [EKM] では破産問題のモデルのベースとなる分布として, subexponential 分布を使っており, 付録の一部に Regular Variation and Subexponentiality として, 本の理解に必要な内容を補足している. Subexponentiality も含めた基本的な結果を知りたいというときには役立つと思われる. 最後の [HF06] は, 著者のひとりが正則変動関数の理論の大家であるのに加え, 出たばかりの本でもあるので, 最近の結果を含み, 比較的高度な内容となっている. 確率論の本ではないが, タウバー型定理と関連して [Ko04] でもひとつの章を割いて扱っている.

これらに対して, 正則変動関数そのものを主題とした書籍には, [Se76] と [BGT] がある. 前者はシュプリングーのレクチャーノートで, 基本的な内容を中心によくまとめられており, 付録では正則変動関数の一般化についても書かれている. また, どの結果を誰がいつ出したかといった歴史的なことにも触れられている点が興味深い. 本稿の内容の多くもこれに含まれている. 片や, 1987 年初版の [BGT] (現在のペーパーバックは一部加筆修整されている) は Encyclopedia of Mathematics and its Applications とあるように, それまでの正則変動関数に関する結果が網羅され, 前半で正則変動関数の一般論を扱い, 後半は数学の各分野, 解析数論, 複素解析, そして確率論への応用について書かれている. 正則変動関数について何か知りたいとき, この本を開けば, 最近の結果でない限り, ほぼ

見つかるであろう。本稿の内容の多くもこの本からの引用である。前書きには正則変動関数の研究の歴史にも簡潔に触れられている。[GH87]は、正則変動性を扱ってはいるが、教科書的というよりも、1.4で取り上げたタウバー型定理に関連する内容に特化しており、 $\Gamma$ 変動及び $\Pi$ 変動など正則変動性の拡張についても書かれている。

確率統計の話題からは、安定分布、極値理論、subexponential分布を取り上げた。[GK68]は古典的名著で独立確率変数の和の極限分布の書名通り、無限分解可能分布、安定分布、L分布などについて書かれている。安定分布は[GK68]や[Fe71]でも取り上げられているが、より専門的なものに、[Zo86]がある。安定分布は無分解可能であるから、これについて書かれた本でも扱われている。無限分解可能分布については、和書の[Sa90]が詳しい。[Sa99]はこれに加筆し、英訳したものである。Lévy過程については[Be96]もある。最近、Black-Scholes modelの影響からか、[Sc03]や[Mi03]などのような金融への応用に関する書籍もできるようになってきた。

極値理論については、和書はないものの、洋書は数多い。入門書としては、既に紹介した[EKM]や[Co01]、[FR04]がある。日本の研究者による最近の研究報告に、統計数理の極値理論特集[To04]がある。

[Ch64]に始まったSubexponential分布は、専門に扱った書籍はない。前述の[EKM]や[BGT]の他に[AN72]と[As00]で扱っているが、それ以上の知識を得るには論文を読むしかない。Embrechtsによるレビュー[Em84]やKlüppelbergによる多数の論文がある。筆者が現在興味を持っているのもsubexponential分布で、 $S$ より広い分布族に対する定理2.3.3のタイプの定理を示している([SW05])。この論文の参考文献にはsubexponential分布についてのものを多く挙げてある。

## 参考文献

### [正則変動関数主体の書籍]

- [Se76] Seneta, E. (1976). Functions of Regular Variation, Lecture note in Mathematics, vol. 508. Springer-Verlag, New York.
- [BGT] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). Regular Variation, Encyclopedia of Mathematics and its Applications, Cambridge university press, Cambridge.
- [GH87] Geluk, J.L. and de Hann, L. (1987). Regular Variation, Extensions and Tauberian Theorems., Math.Centre Tract 32, Amsterdam.

### [正則変動関数及び関連事項を扱っている主に確率統計の書籍]

- [EKM] Embrechts, P, Klüppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events, Applications of Mathematics 33, Springer.
- [Fe71] Feller, W. (1971). An Introduction to Probability Theory and its Applications, vol.II, 2nd. edn. Wiley, New York.

- [Ha70] de Haan, L. (1970). On Regular Variation and its Applications to the Weak Convergence of Sample Extremes, Math.Centre Tract 32, Amsterdam.
- [HF06] de Haan, L., Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer.
- [Ko04] Korevaar, J. (2004). Tauberian Theory: An Century of Developments. Springer
- [Re87] Resnick, S.I. (1987). Extreme Values, Regular Variation and Point Processes. Springer-Verlag, New York.

[無限分解可能分布, 安定分布を扱った書籍]

- [Be96] Bertoin, J. (1996). Lévy Processes, Cambridge university press, Cambridge.
- [GK68] Gnedenko, B.V. and Kolmogorov, A.N. (1968). Limit Distributions for Sums of Independent Random Variables. 2nd ed., Addison-Wesley, Reading, Mass., 1968(Russian original 1949).
- [Mi03] 宮原孝夫. (2003). 株価モデルとレヴィ過程, シリーズ〈金融工学の基礎〉1, 朝倉書店.
- [Sc03] Schoutens,W.(2003). Lévy Processes in Finance: Pricing Financial Derivatives, Wiley, New York .
- [Sa90] 佐藤健一. (1990). 加法過程, 紀伊国屋数学叢書 33, 紀伊国屋書店.
- [Sa99] Sato, K. (1999). Lévy Processes and Infinitely Divisible Distributions, Cambridge university press, Cambridge.
- [Zo86] Zolotarev, V.M. (1986). One-Dimensional Stable Distributions, Amer. Math. Soc., Providence, R.I. [Russian original 1983].

[極値理論関連]

- [Co01] Coles, S. G. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer Verlag, London.
- [FR04] Finkenstädt, B. and Rootzén, H. Eds. (2004). Extreme Values in Finance, Telecommunications, and the Environment. CRC/Chapman & Hall, Boca Raton.
- [LLR] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer Verlag, New York Heidelberg Berlin.
- [TS04] 統計数理 (2004). 特集 極値理論. 52, No.1

#### [Subexponential についての文献]

- [AN72] Athreya, K.B. and Ney, P.E. (1972). *Branching Processes*, Springer-Verlag, New York.
- [As00] Asmussen, S. (2000). *Ruin Probabilities*, World Scientific, Singapore.
- [Em84] Embrechts, P. (1984). Subexponential distribution on functions and their applications: a review. In M. Iosifescu (ed.), *Proceedings of the 7th Brasov Conference on Probability Theory*, pp. 125-136. Bucharest: Editura Academiei Republicii Socialiste Romania.
- [SW05] Shimura, T. and Watanabe, T. (2005) Infinite divisibility and generalized subexponentiality. *Bernoulli* **11**(3), 445-469.

#### [歴史的文献]

- [Ch64] Chistyakov, V.P. (1964). A theorem on sums of independent positive random variables and its applications to branching process. *Theory Probab. Appl.* **9**, 640-648.
- [Gu63] ゲンベル (1963). 極値統計学；極値の理論とその工学的応用 河田竜夫, 岩井重久, 加瀬滋男 監訳, Gumbel, E. J., *Statistics of extremes* の訳. 広川書店.
- [Kr30] Karamata, J. (1930). Sur un mode de croissance régulière des fonctions. *Mathematica(Cluj)* **4**, 38-53.
- [Kr33] Karamata, J. (1933). Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bull. Soc. Math. France*, **61**, 55-62.

#### [正規分布の吸引域に関連する文献]

- [Ma81] Maller, R.A. (1981). A theorem on products of random variables, with application to regression. *Austral.J.Statist.* **23** (2), 177-185.
- [Tu86] Tucker, H.G. (1968). Convolutions of distributions attracted to stable laws,. *Ann.Math.Statist.* **39**, No.5, 1381-1390.
- [Sh91] Shimura, T. (1991). Decomposition of non-decreasing slowly varying functions and the domain of attraction of Gaussian distributions, *J.Math.Soc.Japan* **43**, No.4, 775-793.

#### [その他]

- [Da52] Darling, D.A. (1952). The influence of the maximum term in the addition of independent random variables, *Trans. Amer. Math. Soc.* **73**, 95-107.



# 確率過程の統計学： マルチンゲールと empirical process

統計数理研究所 西山陽一

nisiyama@ism.ac.jp

2006 年 10 月

**要旨.** エントロピー法は、集合や関数によって添字づけられた I.I.D. データの経験分布過程に対する大数の法則や中心極限定理を確立するために、1980 年代に研究された。さらに、最近のいくつかの論文はエントロピー法が他の統計的問題にも有用であることを示してきている。この方法のある部分を I.I.D. だけでなくマルチンゲールの枠組みで体系的に考察する研究が 1990 年代の半ばごろから始められた。本稿では 2006 年の時点までの結果の一部を紹介する。

1 節では Ossiander の中心極限定理の一般化に関する直感的な説明を通じて、エントロピー法の一般化のポイントがどこにあるかを解説する。簡明のため、本稿の残りの部分は主として連続マルチンゲールを中心に話を進める。2 節では、本研究の鍵となる quadratic modulus という量を導入し、その言葉で最大不等式を与える。3 節では弱収束定理を与える。それらを用いてカーネル推定量の局所確率場の漸近挙動、いくつかの  $M$ -推定量の収束率、積分型推定量の漸近正規性などを導出する。

*Key words.* マルチンゲール, 経験分布過程, 最大不等式, 中心極限定理, カーネル推定量, 変化点, 最尤推定量.

## 1 序 —I.I.D. からマルチンゲールへ—

本研究の目的は、これまで主として I.I.D. データの経験分布過程に対して発展してきたエントロピー法のコア部分をマルチンゲールの枠組みに一般化し、確率過程の統計的推測への応用を可能ならしめる事である。ここで経験分布過程と呼んでいるのは、Dudley らによって研究されてきた集合や関数を添字にもつ確率場の事である。直感的な説明をかねて、まず Ossiander (1987) による I.I.D. の場合に対する中心極限定理の紹介をし、その後それを dependent データの最も原始的な場合に一般化する。その説明の中で、どこがコア部分であるかを解説する。

$(E, \mathcal{E})$  は可測空間であるとする.  $\{Z_i\}_{i \in \mathbb{N}}$  は  $E$ -値の I.I.D. 確率変数であるとし, その共通の分布を  $P$  と表す.  $\Psi$  は  $\mathcal{L}^2(P) = \mathcal{L}^2(E, \mathcal{E}, P)$  の部分集合であって, ある  $\varphi \in \mathcal{L}^2(P)$  に対し全ての  $\psi \in \Psi$  が  $|\psi| \leq \varphi$  を満たすものとする ( $\varphi$  は  $\Psi$  に含まれていなくてもよい). いま

$$X^n(\psi) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \psi(Z_i) - \int_E \psi(z) P(dz) \right\}$$

によって定義される確率過程  $\psi \rightsquigarrow X^n(\psi)$  を考える. Ossiander (1987) による  $L^2$ -bracketing とは次のようなものである; 各  $\varepsilon \in (0, 1]$  に対し,  $N(\varepsilon)$  個の  $\mathcal{L}^2(P)$  の要素のペア  $[l^{\varepsilon, k}, u^{\varepsilon, k}]$ ,  $k = 1, \dots, N(\varepsilon)$ , であって, 任意の  $\psi \in \Psi$  に対して関係  $l^{\varepsilon, k} \leq \psi \leq u^{\varepsilon, k}$  がいずれかの  $k$  に対して成り立ち, しかも

$$(1.1) \quad \sqrt{\int_E |u^{\varepsilon, k}(z) - l^{\varepsilon, k}(z)|^2 P(dz)} \leq \varepsilon$$

が成り立つ. Ossiander の定理の主張は「もしもこの bracketing procedure がエントロピー条件

$$(1.2) \quad \int_0^1 \sqrt{\log N(\varepsilon)} d\varepsilon < \infty,$$

を満たすように達成されるならば, 確率過程の列  $\psi \rightsquigarrow X^n(\psi)$  は  $\Psi$  によって添字付けられたブラウン橋, すなわち, 平均ゼロの正規確率過程  $\psi \rightsquigarrow G(\psi)$  であって  $EG(\psi)G(\phi) = \int_E \psi(z)\phi(z)P(dz) - \int_E \psi(z)P(dz) \int_E \phi(z)P(dz)$  を満たすものに  $\ell^\infty(\Psi)$  の中で弱収束する」という事である. ただし  $\ell^\infty(\Psi)$  は  $\Psi$  上の有界関数全体の空間とし, これに supremum norm を添付して完備距離空間とする.

さて,  $\{Z_i\}_{i \in \mathbb{N}}$  は任意の  $E$ -値確率変数の列とし,  $P_i$  は  $\mathcal{F}_{i-1} = \sigma\{Z_1, \dots, Z_{i-1}\}$  (ただし  $\mathcal{F}_0$  は null  $\sigma$ -field) を与えたもとの  $Z_i$  の条件付き分布を表すものとする. いま

$$X^n(\psi) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \psi(Z_i) - \int_E \psi(z) P_i(dz) \right\}$$

によって与えられる確率過程  $\psi \rightsquigarrow X^n(\psi)$  を考える. 上の bracketing procedure のうち, (1.1) を

$$\sqrt{\int_E |u^{\varepsilon, k}(z) - l^{\varepsilon, k}(z)|^2 P_i(dz)} \leq K_i \varepsilon \quad \text{almost surely}$$

に取り替えたものを考えよう. ただし  $K_i$  は  $\varepsilon$  や  $k$  に依存しない確率変数であって  $\mathcal{F}_{i-1}$ -可測なるものである. いまの場合, 左辺がランダムであるから右辺にもランダムな係数  $K_i$  を許したのである. このとき, もしもエントロピー条件 (1.2) が満たされ, しかも

$$\bar{K}^n = \sqrt{\frac{1}{n} \sum_{i=1}^n |K_i|^2}$$

によって定義される確率変数が確率有界であるならば、確率過程  $\psi \sim X^n(\psi)$  の緊密性は有限次元分布の収束と、ある種の Lindeberg 条件から従う。Ossiander の定理は  $P_i \equiv P$  かつ  $K_i \equiv 1$  の場合であると考えられる。本研究の鍵となるランダム量 “quadratic modulus” は、この確率変数  $\bar{K}^n$  と類似したものである。より近い表現は

$$\text{“quadratic modulus”} = \sup_{\varepsilon \in (0,1]} \max_{1 \leq k \leq N(\varepsilon)} \frac{\sqrt{n^{-1} \sum_{i=1}^n \int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P_i(dz)}}{\varepsilon}.$$

である。このランダム量は bracket の選択に見苦しい形で依存しており、また我々は重み関数  $\psi$  そのものがランダムである場合も考えたいので、実際には “quadratic modulus” は少し異なる形で定義される。それ以前に、本研究では「離散時間のマルチンゲール」、 「整数値ランダム測度」、 「連続マルチンゲール」の三つの場合に分けてアプローチをしているが、上述のものはその最初のものにすぎない。後 2 者のうちの特に「連続マルチンゲール」は異なる発想による “quadratic modulus” を定義する必要があることを注意しておこう。本稿では、紙数の関係と説明の簡明のため（これまでのイントロとは別のものとなってしまいが、敢えて）「連続マルチンゲール」の場合における結果を中心に報告する。しかし、他の 2 者についても同様のアプローチにより類似した結果が得られる事は強調しておきたい。

エントロピー法の統計学における有用性は、はじめは主としてより一般の一様大数の法則や中心極限定理（特に多次元の Donsker の定理）を求める努力を通じて認識されてきた。しかしながら、その核心部分、すなわちエントロピーによって制御された chaining と bracketing の技術は、それらの極限定理に直接関係のない統計的問題にも適用できる事が、最近のいくつかの論文により示されてきている。例えば本稿でも扱う  $M$ -推定がそのよい例である。我々にとって、エントロピー法の利点の一つは、モデルが I.I.D. でなくても  $\{\bar{K}^n \leq L\}$  という形の集合上ではその技術がそのまま適用できるという事である。それ故、確率過程の統計的推測に関する問題のいくつかは、上述のような truncation を導入し、その補集合  $\{\bar{K}^n > L\}$  が大きな定数  $L > 0$  に対して無視できる事を示す事によって解決できる。これが本研究で繰り返し用いるアプローチである。

本研究の対象としてマルチンゲールを選ぶ理由は 3 つある。理由の第 1 は、I.I.D. の場合において基本的な道具であった Bernstein の不等式が、マルチンゲールの枠組みでも、predictable quadratic variation に基づく truncation が必要であるという修正のもとで、すでに与えられている事である。第 2 に、我々の状況における有限次元分布の収束を示すために、よく発達したマルチンゲール中心極限定理を利用する事ができる。第 3 に、マルチンゲールという概念により、統計的モデルの豊富なクラスを扱う事ができる（例えば、生存解析における multiplicative intensity モデル、マルコフ連鎖、正規白色ノイズモデル、拡散過程、Lévy 過程など）。

本稿の構成は次の通り。先にもふれたが、2 節以降は敢えて「連続マルチンゲール」とその応用を中心に扱う。まず 2 節において、基本となる最大不等式を述べる。3 節では、

それを用いて連続マルチンゲールから生成される確率場に対する中心極限定理を与える。4 節において、中心極限定理を応用して、カーネル推定量の局所確率場の弱収束を導出する。5 節と 6 節では  $M$ -推定への応用を述べる。まず 5.1 節において  $M$ -推定量の収束率を判定するための一般的な定理を紹介する。それは 5.2, 5.3 および 6 節において有用となる。5.2 節では、4 節の結果も用いて、正規白色ノイズモデルの回帰関数の最大点の位置の推定問題を考察し、5.3 節では回帰関数のジャンプ点の位置の推定を行う。6 節では、正規白色ノイズモデルの回帰関数のノンパラメトリック最尤推定量の収束率を導出する。特に、回帰関数が単調である場合と滑らかである場合についての例を検討する。7 節では、Lévy 測度を Nelson-Aalen 型の推定量を用いて推定する問題を考える。本来ここで述べる結果を導出するためには、整数値ランダム測度に対する最大不等式ならびに中心極限定理を用意する必要があるが、本稿ではその解説は省略し、結果のみ与える。

若干の文献紹介をしてこの節を終える。集合や関数を添字にもつ経験分布過程の理論は I.I.D. の場合に対して Dudley (1978) によって創始され、1980 年代に列独立な三角列の場合まで研究された。van der Vaart and Wellner (1996) の本がその良い解説ならびに文献紹介を含んでいるので、ここでは敢えて述べない。1990 年代に入り、独立性をはずす努力が始められた。定常過程に対する mixing condition に関しては Doukhan et al. (1995) があり、マルチンゲールに関しては Bae and Levental (1995a,b), van de Geer (1995), Nishiyama (1997) がある。本稿は Nishiyama (1997) の発展版である Nishiyama (1999, 2000a, 2000b, 2007, 2006) の一部を紹介したものであるが、話題の中心を連続マルチンゲールとその正規白色ノイズモデルへの応用に据えたため、大部分は Nishiyama (1999) で扱ったものとなった。

Bae and Levental (1995a) は、マルコフ連鎖の経験分布過程の中心極限定理を導出するために、離散時間のマルチンゲールに対する本稿の 1 節の第三段落と類似した主張を得た。また Bae and Levental (1995b) は、パラメータ  $\psi$  に関する連続性も仮定した上で、本稿の定理 3.2 と同じ結論を得た。彼らの議論の中に、quadratic modulus という量の起源は見いだす事はできるが、その量の導入のためには van der Vaart and Wellner (1996) による新しい緊密性判定定理も大きな動機になっていることを付記しておきたい。一方 van de Geer (1995) は Bernstein の不等式のマルチンゲール版のジャンプの大きさに関する仮定をゆるめると同時に、Ossiander (1987) による bracketing の技術を用いて点過程のノンパラメトリック最尤推定量の収束率を導出した。Nishiyama (1997) は同様の技術を用いて、弱収束定理を証明し、Nelson-Aalen 推定量への応用を得た。

以上のように、本研究は、Dudley ら (特に Ossiander) による経験分布過程の研究をマルチンゲールに一般化するにあたり、特別な場合や少し異なる視点から考えた Bae and Levental や van de Geer によるアイデアと、van der Vaart and Wellner の本の出版に契機づけられて始められたものであった。Nishiyama (2000b) は 1990 年代におけるその成果の集大成である。しかしながら、そこにおいてはジャンプが無数個あるようなマルチン

ゲールは扱われていなかった。この問題を解決したのは Nishiyama (2007) である。その結果を用いて、Nishiyama (2006) は Lévy 過程のノンパラメトリック推定問題や、時間的一様性の検定問題を扱った。一方、mixing condition によるアプローチの最近の研究としては Pollard (2002) があり、拡散過程の Donsker の定理については van der Vaart and van Zanten (2005) がある。余談ではあるが、前者の論文は 1 節を割いて Ossiander の定理が証明された 1984 年頃における Seattle group の様子を紹介している。本研究の最大のメリットは関数族  $\Psi$  の取り方の自由性にあり、今後も更なる応用が産み出されることが期待される。

## 2 連続マルチンゲールの確率場に対する最大不等式

$\mathbf{B} = (\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}, P)$  は filtration が添付された確率空間であるとし、 $(\Psi, \rho)$  は距離空間であるとする。各  $\psi \in \Psi$  に対し  $\mathbf{B}$  上の連続局所マルチンゲール  $X^\psi$  であって  $X_0^\psi = 0$  なるものが与えられたとする。本節の目標は、その族  $X = (X^\psi | \psi \in \Psi)$  に対しする最大不等式、すなわち  $E \sup_{\psi, \phi} |X_t^\psi - X_t^\phi|$  の形の値を押さえる不等式を導出することである。第 1 節でも述べたが、それは以後の全ての結果の道具として有用である。

まず連続局所マルチンゲールの quadratic modulus を定義する事から始めよう。

**定義 2.1** 与えられた部分集合  $S \subset \Psi$  に対し、 $X$  の quadratic  $(S, \rho)$ -modulus を

$$\|X\|_{(S, \rho), t} = \sup_{\substack{\psi, \phi \in S \\ \psi \neq \phi}} \frac{\sqrt{\langle X^\psi - X^\phi, X^\psi - X^\phi \rangle_t}}{\rho(\psi, \phi)} \quad \forall t \in \mathbb{R}_+$$

と定義する。

$S$  が可算であれば  $[0, \infty]$ -値確率変数  $\|X\|_{(S, \rho), t}$  は  $\mathcal{F}_t$ -可測となるが、 $S$  が非可算であれば  $\mathcal{F}$ -可測性すら保障されないことに注意されたい。以下で、距離空間  $(\Psi, \rho)$  が全有界であるとは任意の  $\varepsilon > 0$  に対して  $\Psi$  が有限個の半径  $\varepsilon$  の閉球によって覆われることを意味するものとし、また  $N(\Psi, \rho; \varepsilon)$  によってそういう閉球の最小個数を表すものとする。次が本節の主定理である。

**定理 2.2**  $(\Psi, \rho)$  は全有界距離空間であるとする。  $\tau$  は  $\mathbf{B}$  上の有限停止時刻であるとする。このとき、任意の  $\delta, K > 0$  に対し

$$\sup_{\substack{\Psi^* \subset \Psi \\ \text{countable}}} E \sup_{t \in [0, \tau]} \sup_{\substack{\psi, \phi \in \Psi^* \\ \rho(\psi, \phi) \leq \delta}} |X_t^\psi - X_t^\phi| 1_{\{\|X\|_{(\Psi^*, \rho), \tau} \leq K\}} \leq CK \int_0^\delta \sqrt{\log(1 + N(\Psi, \rho; \varepsilon))} d\varepsilon,$$

が成り立つ。ただし  $C > 0$  は universal constant で、最初の supremum は  $\Psi$  の全ての可算部分集合でとられる。

ところで、我々はパス  $\psi \rightsquigarrow X_t^\psi$  や  $(t, \psi) \rightsquigarrow X_t^\psi$  が連続あるいは有界であるかどうかという疑問にしばしば出くわす。上の結果を適用して、この問題に対する 2 種類の解答を提示する事ができる。1 つ目のものは  $\Psi$  が可算集合である場合に関するものである。

**定理 2.3** 定理 2.2 の状況を考える。もしも  $\Psi$  が可算でしかも

$$P(\|X\|_{(\Psi, \rho), \tau} < \infty) = 1 \quad \text{かつ} \quad \int_0^1 \sqrt{\log N(\Psi, \rho; \varepsilon)} d\varepsilon < \infty.$$

であるならば、 $\psi \rightsquigarrow X_t^\psi$  のほとんど全てのパスは  $\Psi$  上で  $\rho$  に関して一様連続である; 特に、それらは  $\ell^\infty(\Psi)$  に属している。さらに  $\tau > 0$  が定数であるならば、 $(t, \psi) \rightsquigarrow X_t^\psi$  のほとんど全てのパスは  $\tilde{\rho}((t, \psi), (s, \phi)) = |t - s| \vee \rho(\psi, \phi)$  によって定義される  $[0, \tau] \times \Psi$  上の擬距離  $\tilde{\rho}$  に関して一様連続である; 特に、それらは  $\ell^\infty([0, \tau] \times \Psi)$  に属している。

$\Psi$  が非可算である場合、次の定理が  $\psi \rightsquigarrow X_t^\psi$  の連続ヴァージョンの存在のための十分条件を与える。

**定理 2.4** 定理 2.2 の状況を考える。もしも  $\|X\|_{(\Psi, \rho), \tau} \leq Y$  (identically) なる  $\mathcal{F}_\tau$ -可測、 $[0, \infty]$ -値確率変数  $Y$  が存在し、しかも

$$P(Y < \infty) = 1 \quad \text{かつ} \quad \int_0^1 \sqrt{\log N(\Psi, \rho; \varepsilon)} d\varepsilon < \infty$$

であるならば、ある  $\mathcal{F}_\tau$ -可測確率変数の族  $\{\tilde{X}(\psi) : \psi \in \Psi\}$  であって各  $\psi \in \Psi$  に対し確率 1 で  $\tilde{X}(\psi) = X_t^\psi$  であり、しかも  $\psi \rightsquigarrow \tilde{X}(\psi)$  のほとんど全てのパスが  $\Psi$  上で  $\rho$  に関して一様連続であるものが存在する; 特に、それらは  $\ell^\infty(\Psi)$  に属している。(そのような確率過程  $\psi \rightsquigarrow \tilde{X}(\psi)$  は  $\psi \rightsquigarrow X_t^\psi$  の  $\rho$ -連続ヴァージョンと呼ばれる。)

構成された  $\tilde{X}(\psi)$  はもはや連続局所マルチンゲールの終点値ではない事に注意されたい。

Theorem 2.2 において、 $\rho$  が  $\Psi$  上の (擬距離ではなく) 距離 であるという仮定は応用上強すぎる事もある。次の定理は、より一般に、ランダムな擬距離への修正に関するものである。ただし、エントロピーは依然としてある距離に関して計算しなければならない。

**定理 2.5** 定理 2.2 の状況を考える。もしも与えられた  $\Psi$  上のランダム擬距離  $\varrho$  が

$$\sqrt{\langle X^\psi - X^\phi, X^\psi - X^\phi \rangle_\tau} \leq \varrho(\psi, \phi) \quad \forall \psi, \phi \in \Psi \quad \text{identically}$$

を満たすならば、任意の  $\delta, K > 0$  に対し

$$\sup_{\substack{\Psi^* \subset \Psi \\ \text{countable}}} E^* \sup_{t \in [0, \tau]} \sup_{\substack{\psi, \phi \in \Psi^* \\ \varrho(\psi, \phi) \leq K\delta}} |X_t^\psi - X_t^\phi| 1_{\{\|\varrho\|_\rho \leq K\}} \leq CK \int_0^\delta \sqrt{\log(1 + N(\Psi, \rho; \varepsilon))} d\varepsilon,$$

ただし

$$\|\varrho\|_\rho = \sup_{\substack{\psi, \phi \in \Psi \\ \psi \neq \phi}} \frac{\varrho(\psi, \phi)}{\rho(\psi, \phi)},$$

が成り立つ。ここに  $C > 0$  は universal constant で、最初の supremum は  $\Psi$  の全ての可算部分集合  $\Psi^*$  でとられる。

### 3 連続マルチンゲールの確率場に対する中心極限定理

まず1つの定義を用意するところから始めよう。

**定義 3.1** 距離空間  $(\Psi, \rho)$  によって添字付けられた連続局所マルチンゲールの族  $X = (X^\psi | \psi \in \Psi)$  が  $\rho$ -可分であるとは、ある可算部分集合  $\Psi^* \subset \Psi$  と  $P$ -ゼロ集合  $N \in \mathcal{F}$  が存在して、任意の  $\varepsilon > 0$  と  $\omega \in \Omega \setminus N$  に対し

$$X_t^\psi(\omega) \in \overline{\{X_t^\phi(\omega) : \phi \in \Psi^*, \rho(\psi, \phi) < \varepsilon\}} \quad \forall t \in \mathbb{R}_+, \forall \psi \in \Psi,$$

が成り立つときにいう。ただし閉包は  $\mathbb{R} \cup \{-\infty, +\infty\}$  においてとられる。

もしも  $(\Psi, \rho)$  が可分ならば、この  $\rho$ -可分性のための十分条件は、ほとんど全てのパス  $\psi \rightsquigarrow X_t^\psi$  が  $\rho$ -連続である事である。しかし、一般にはパスの連続性をチェックするのは容易ではない。一方、添字集合  $\Psi$  が可算である場合には、どのような族も  $\Psi$  上の任意の距離  $\rho$  に関して  $\rho$ -可分である事は明らかである。

では弱収束の文脈に移ろう。 $(\Psi, \rho)$  は全有界距離空間であるとする。各  $n \in \mathbb{N}$  に対し、filtration が添付された確率空間  $B^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t \in \mathbb{R}_+}, P^n)$  が与えられたとする。 $X^n = (X^{n,\psi} | \psi \in \Psi)$  は  $B^n$  上の (必ずしも  $\rho$ -可分ではない) 連続局所マルチンゲール  $X^{n,\psi}$  であって  $X_0^{n,\psi} = 0$  であるものの族であるとする。 $\tau^n$  は同じ確率空間上で定義された有限停止時刻であるとする。ここで Metric Entropy condition を導入しよう。

[ME] 与えられた  $B^n$  上の有限停止時刻  $\tau^n$  に対し

$$\|X^n\|_{(\Psi^*, \rho), \tau^n} = O_{P^n}(1) \quad \text{かつ} \quad \int_0^1 \sqrt{\log N(\Psi, \rho; \varepsilon)} d\varepsilon < \infty.$$

**定理 3.2** 上の状況において族  $X^n = (X^{n,\psi} | \psi \in \Psi)$  は  $\rho$ -可分であるとする。 $\psi \rightsquigarrow X_{\tau^n}^{n,\psi}$  は確率1で  $\ell^\infty(\Psi)$  に値をとり、さらにその任意の有限次元マージナルはある Borel 可測分布に弱収束するとする。このとき、もしも条件 [ME] が満たされるならば、 $X_{\tau^n}^{n,\psi}$  は  $\ell^\infty(\Psi)$  の中である緊密 Borel 可測分布に弱収束する。

証明のポイントのみ述べる。一般に、全有界擬距離空間  $(T, \rho)$  によって添字付けられた、 $\ell^\infty(T)$  に値をとる確率場の列  $t \rightsquigarrow X^n(t)$ ,  $n = 1, 2, \dots$  が緊密であるための十分条件は、任意の有限次元マージナルが弱収束する事と任意の  $\varepsilon, \eta > 0$  に対しある  $\delta > 0$  が存在して

$$\limsup_{n \rightarrow \infty} P^{n*} \left( \sup_{\rho(t,s) \leq \delta} |X^n(t) - X^n(s)| > \varepsilon \right) \leq \eta$$

である事である。定理 3.2 の枠組みの場合、これは2節で与えた最大不等式 (定理 2.2) と Markov の不等式から直ちに従う。

さて、定理 2.3 によれば、もしも  $\Psi$  が可算で [ME] が満たされるならば、 $\psi \rightsquigarrow X_{\tau^n}^{n,\psi}$  が確率1で  $\ell^\infty(\Psi)$  に値をとるための十分条件は  $P^n(\|X^n\|_{(\Psi, \rho), \tau^n} < \infty) = 1$  が成り立つ事

である.  $\Psi$  が非可算である場合には,  $\psi \rightsquigarrow X_{\tau^n}^{\psi}$  そのものではなくその  $\rho$ -連続ヴァージョンを扱った次の系が有用となる.  $\rho$ -連続ヴァージョンの存在自体は定理 2.4 によって保証されている事を思い出そう.

**系 3.3** 上の状況において,  $\tau^n \equiv \tau$  は定数であるとし, ある定数の族  $\{C(\psi, \phi); \psi, \phi \in \Psi\}$  に対し

$$\langle X^{n,\psi}, X^{n,\phi} \rangle_{\tau} \xrightarrow{P^n} C(\psi, \phi) \quad \forall \psi, \phi \in \Psi$$

である事を仮定する. ある  $\mathcal{F}_{\tau}$ -可測,  $[0, \infty]$ -値確率変数  $Y^n$  であって  $\|X^n\|_{(\Psi, \rho), \tau} \leq Y^n$  identically と  $P^n(Y^n < \infty) = 1$  を満たすものが存在する事を仮定する. このとき, もしも [ME] が満たされるならば,  $\psi \rightsquigarrow X_{\tau}^{\psi}$  の  $\rho$ -連続ヴァージョン  $\psi \rightsquigarrow \tilde{X}^n(\psi)$  は, 平均ゼロの正規過程  $\psi \rightsquigarrow G(\psi)$  であって  $EG(\psi)G(\phi) = C(\psi, \phi)$  なるものに  $\ell^{\infty}(\Psi)$  の中で弱収束する.

## 4 応用 1: カーネル推定量の局所確率場の弱収束

I.I.D. データの密度関数の推定問題において, カーネル推定量が各点ごとの漸近正規性をもつことはよく知られている. 本節では, それを局所パラメータに関する汎関数の意味での漸近正規性に拡張する問題を考える. このことは, I.I.D. データをはじめとする様々な状況で可能であるが, ここでは次の式で与えられる正規白色ノイズモデルを考察する:

$$(4.3) \quad dX_t^n = f(t)dt + n^{-1/2}dB_t \quad X_0^n = x_0 \in \mathbb{R}.$$

ただし  $B = (B_t)_{t \in [0,1]}$  は標準ブラウン運動を表す. いま, ある一点  $t_0 \in (0, 1)$  を固定する. ゼロに収束する正定数の列  $b_n$  をとる. 関数  $u \rightsquigarrow f(t_0 + b_n u)$  を推定することを我々の目標とする. ただし  $u$  は  $\mathbb{R}$  の有界部分集合  $U$  の中を動くものとする. 自然な推定量は

$$\hat{f}_n(t_0 + b_n u) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-t_0}{b_n} - u\right) dX_t^n \quad \forall u \in U$$

である. これに関する 2 種類の残差確率場を考えよう:

$$\begin{aligned} Z_n(u) &= \sqrt{nb_n} \left\{ \hat{f}_n(t_0 + b_n u) - \tilde{f}_n(t_0 + b_n u) \right\} \quad \forall u \in U; \\ R_n(u) &= \sqrt{nb_n} \left\{ \hat{f}_n(t_0 + b_n u) - f(t_0 + b_n u) \right\} \quad \forall u \in U. \end{aligned}$$

ただし

$$\tilde{f}_n(t_0 + b_n u) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-t_0}{b_n} - u\right) f(t)dt \quad \forall u \in U$$

とする. カーネル関数  $K$  には次の仮定をおく.



**条件 4.1** 関数  $K : \mathbb{R} \rightarrow [0, \infty)$  は、原点对称、 $\int_{\mathbb{R}} K(x)dx = 1$ 、かつコンパクトな台をもち、しかも次の (i) あるいは (ii) のいずれかを満たす:

- (i)  $x \rightsquigarrow K(x)$  は Lipschitz 連続;
- (ii)  $x \rightsquigarrow K(x)$  は  $[0, \infty)$  上で単調減少.

**定理 4.2** カーネル関数  $K$  は条件 4.1 を満たすものとする. 定数列  $b_n$  を  $b_n \downarrow 0$  かつ  $nb_n \uparrow \infty$  となるようにとる.

(i) もし  $t \rightsquigarrow f(t)$  が点  $t_0 \in (0, 1)$  において連続であるならば、 $Z^n \xrightarrow{P} Z$  in  $\ell^\infty(U)$ . ただし  $u \rightsquigarrow Z(u)$  は平均ゼロの正規過程で

$$EZ(u_1)Z(u_2) = \int_{\mathbb{R}} K(x - u_1)K(x - u_2)dx \quad u_1, u_2 \in U$$

を満たすものである.

(ii) もし  $t \rightsquigarrow f(t)$  が点  $t_0 \in (0, 1)$  の近傍において 2 回連続的微分可能で、しかも

$$\lim_{n \rightarrow \infty} nb_n^5 = h < \infty$$

であるならば、 $R^n \xrightarrow{P} z_0 + Z$  in  $\ell^\infty(U)$ . ただし

$$z_0 = \frac{\sqrt{h}}{2} \cdot \int_{\mathbb{R}} x^2 K(x)dx \cdot \left. \frac{d^2 f(t)}{dt^2} \right|_{t=t_0}$$

上記の主張に現れる仮定は、各点ごとの漸近正規性の文脈におけるものと同一である. 従って、我々の結論は「 $f$  の局所的な滑らかさは、各点ごとでの漸近正規性のみならず、局所的汎関数の意味での分布収束をも保証する」ということである.

なお、定理 4.2 の場合には  $u \rightsquigarrow Z^n(u)$  は連続となるので、空間  $C$  に値をとる確率変数に対する古典的な弱収束理論による扱いも可能である. しかしながら、正規白色ノイズモデル以外の例では  $u \rightsquigarrow Z^n(u)$  は必ずしも連続とはならない. それ故、特に多次元の場合、空間  $\ell^\infty(U)$  に値をとる確率変数に対する新理論が自然なアプローチとなる.

## 5 応用 2: $M$ -推定量の漸近挙動

### 5.1 一般的な $M$ -推定量の収束率判定定理

まず  $M$ -推定量の収束率を導出するための一般的な判定定理を紹介する.  $(\Theta, d)$  を距離空間とする.  $\theta \rightsquigarrow \gamma(\theta)$  は  $\Theta$  上で定義された deterministic な関数であるとする. 我々は、その最大点の位置  $\theta_0 = \operatorname{argmax}_{\theta} \gamma(\theta)$  を推定したいものとする. そのために、パラメータ空間が  $\Theta$  であるような確率場  $\theta \rightsquigarrow \Gamma^n(\theta)$  が与えられたとする. 後に  $\theta \rightsquigarrow \Gamma^n(\theta)$  が  $\theta \rightsquigarrow \gamma(\theta)$  の良い近似になっているという仮定をおくが、そのような場合には、 $\theta_0$  の自

然な推定量は  $\hat{\theta}_n = \operatorname{argmax}_\theta \Gamma^n(\theta)$  であろう。以下では  $\theta \rightsquigarrow \Gamma^n(\theta)$  を criterion process,  $\theta \rightsquigarrow \gamma(\theta)$  を contrast function と呼ぶ。

この枠組みの中で、例えばパラメトリック I.I.D. モデルの最尤推定量 (MLE) は次のように捉えられる。  $\{f_\theta : \theta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}$  は密度関数の族であるとし,  $X_1, \dots, X_n$  は密度  $f_{\theta_0}$  をもつ I.I.D. のデータであるとする。いま

$$\gamma(\theta) = E_{\theta_0} \log f_\theta(X_1) \quad \forall \theta \in \Theta$$

とおく。適当な正則条件のもとで, contrast function  $\theta \rightsquigarrow \gamma(\theta)$  は,  $\theta_0$  の周りでの 2 項 Taylor 展開により,

$$\gamma(\theta) - \gamma(\theta_0) \approx -I(\theta_0)|\theta - \theta_0|^2$$

と近似的に表される。ただし  $I(\theta_0)$  は Fisher 情報量を表す。従って, 確かに contrast function  $\theta \rightsquigarrow \gamma(\theta)$  の最大点は  $\theta_0$  となる。一方, criterion process  $\theta \rightsquigarrow \Gamma^n(\theta)$  は自然に

$$\Gamma^n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f_\theta(X_i) \quad \forall \theta \in \Theta$$

定義されるが, その最大点は MLE に他ならない。正則なモデルでの MLE の収束率は  $n^{-1/2}$  であること, すなわち  $n^{1/2}|\hat{\theta}_n - \theta_0| = O_{P_{\theta_0}}(1)$  であることはよく知られている。しかし, 正則性の仮定が崩れた場合, あるいは, 未知パラメータ  $\theta_0$  を定義する contrast function  $\theta \rightsquigarrow \gamma(\theta)$  がそのものが全く別のものである場合, 収束率はどのように変化するのであろうか? この疑問に 1 つの解答を与えるのが次の一般的な判定定理である。

以下, 枠組みは I.I.D. を離れて, 本節の第一段落で与えた一般的なものに戻る。  $B_d(\theta_0, \delta)$  は  $\theta_0$  を中心とした距離  $d$  に関する半径  $\delta$  の閉球, すなわち,  $B_d(\theta_0, \delta) = \{\theta \in \Theta : d(\theta, \theta_0) \leq \delta\}$  を表すものとする。

**定理 5.1 (van der Vaart and Wellner (1996))** 次の条件 “M-CRITERION” が, ある  $p > 0$ ,  $a \in (0, p)$  および  $\delta_0 \in (0, \infty]$  (いずれも  $n$  に依存しない) と, ある関数  $\phi^n : (0, \delta_0) \rightarrow (0, \infty)$  であって  $\delta \rightsquigarrow \delta^{-a}\phi^n(\delta)$  が単調減少なるものに対して満たされたと仮定する。

M-CRITERION. 任意の  $\varepsilon > 0$  に対しある定数  $c_\varepsilon, C_\varepsilon > 0$  とある  $n_\varepsilon \in \mathbb{N}$  が存在して: 各  $n \geq n_\varepsilon$  に対し集合  $\tilde{\Omega}^n(\varepsilon) \subset \Omega^n$  が存在して

$$\gamma(\theta) - \gamma(\theta_0) \leq -c_\varepsilon d(\theta, \theta_0)^p \quad \forall \theta \in B_d(\theta_0, \delta_0)$$

と

$$E^{n*} \sup_{\theta \in B_d(\theta_0, \delta)} |(\Gamma^n - \gamma)(\theta) - (\Gamma^n - \gamma)(\theta_0)| 1_{\tilde{\Omega}^n(\varepsilon)} \leq C_\varepsilon \phi^n(\delta) \quad \forall \delta \in (0, \delta_0)$$

が成り立ち, しかも  $P^{n*}(\Omega^n \setminus \tilde{\Omega}^n(\varepsilon)) \leq \varepsilon$  である。

このとき  $\phi^n(r_n^{-1}) \leq r_n^{-p}$  を満足する定数  $r_n > 0$  をとれ. すると, もしも写像  $\hat{\theta}_n: \Omega^n \rightarrow \Theta$  が

$$(5.4) \quad \lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} P^{n*} \left( \Gamma^n(\hat{\theta}_n) < \Gamma^n(\theta_0) - Lr_n^{-p} \right) = 0$$

と

$$(5.5) \quad \lim_{n \rightarrow \infty} P^{n*} \left( d(\hat{\theta}_n, \theta_0) > \delta_0/2 \right) = 0$$

を満たすならば,

$$\lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} P^{n*} \left( r_n d(\hat{\theta}_n, \theta_0) > L \right) = 0.$$

が成り立つ. もしも  $M$ -CRITERION が  $\delta_0 = \infty$  に対して満たされている場合, 一致性の仮定 (5.5) は不必要である.

$M$ -CRITERION 中のディスプレイした 2 つ不等式をそれぞれ「第一不等式」, 「第二不等式」と呼ぶことにする. 第一不等式は, contrast function  $\theta \rightsquigarrow \gamma(\theta)$  が真値  $\theta_0$  を他の値  $\theta$  から  $p$  次のオーダーで分離する能力があることを意味し, 第二不等式は, criterion process  $\theta \rightsquigarrow \Gamma^n(\theta)$  が contrast function  $\theta \rightsquigarrow \gamma(\theta)$  をどの程度の誤差で近似する能力があるかを規定している. それらを併せて収束率  $r_n$  が関係式  $\phi^n(r_n^{-1}) \leq r_n^{-p}$  によって導出される.

上記の判定定理の  $p = 2$  場合は, 特別な場合を扱った 1990 年代前半の何人かの仕事をもとにして, van der Vaart and Wellner (1996, Theorem 3.2.5) によってほぼこの形に与えられた. さらに彼らは,  $\gamma$  や  $\theta_0$  も  $n$  に依存するランダムなものである場合への一般化も与えた (彼らの Theorem 3.4.1). 一方 Nishiyama (1997b) は, 同定理が一般の  $p$  についても成り立つことを指摘し, さらに真値  $\theta_0$  に関する一様性についても議論した. 以下の 5.2 および 5.3 節では, 敢えて一般の  $p$  に対する例のみを考察し, また簡明のため収束の真値に関する一様性は言及しない.

なお, 特に  $\Theta$  がユークリッド空間の部分集合である場合, 収束率のみならず漸近分布をも導出できる事がしばしばある. そのための一般的な手続きは以下の通りである (詳しくは van der Vaart and Wellner (1996) の Theorem 3.2.2 を見よ). いかなる場合においても, まず

$$M^n(h) = a_n \{ \Gamma^n(\theta_0 + r_n^{-1}h) - \Gamma^n(\theta_0) \}$$

というような形とした rescaled criterion process  $h \rightsquigarrow M^n(h)$  を考える事になる. 従って, 定数列  $r_n$  を見いだす事が最初の問題となるが, そのために定理 5.1 は有用である. 定数列  $a_n$  は  $r_n$  との関係を考えて決定される. 次に, argmax continuous mapping theorem (van der Vaart and Wellner (1996) の Theorem 3.2.2) を適用するために次の事を示す必要がある.

(A1) 局所列  $\hat{h}_n = r_n(\hat{\theta}_n - \theta_0)$  の一様緊密性.

(A2) 確率場  $h \rightsquigarrow M^n(h)$  が連続確率場  $h \rightsquigarrow M(h) \in \ell^\infty(K)$  の中で弱収束する事. ただし  $K$  は局所パラメータの空間の任意のコンパクト部分集合.

(A3) 極限の連続確率場  $h \rightsquigarrow M(h)$  のパスの最大点  $\hat{h}$  の一意的存在.

これらが示されたとき, “ $r_n(\hat{\theta}_n - \theta_0) \xrightarrow{P} \hat{h}$ ” という形の結果が argmax continuous mapping theorem から従う.

我々がこのアプローチに関して有限次元パラメータの場合のみに甘んずる理由は, 局所列  $\hat{h}_n$  の一様緊密性 (上のステップ (A1)) が “ $r_n|\hat{\theta}_n - \theta_0| = O_P(1)$ ” と同値であり, そしてこれがまさしく定理 5.1 の主張の形であるからである. このことは, パラメータ空間  $(\Theta, d)$  が一般の場合には必ずしも真ではない. しかし, 定理 5.1 は収束率の導出のためにやはり有用であり, その例を 6 節で提示する. なお 3 節の弱収束定理やその応用である 4 節の結果は上のステップ (A2) において用いられる.

## 5.2 例: 最高点の位置の推定

4 節で考えた正規白色ノイズモデル (4.3) を考える. ここでは関数  $t \rightsquigarrow f(t)$  の最高点の位置  $t_0$  の推定問題を考える. 採用する criterion process  $\Gamma^n(\theta)$  と contrast function  $\gamma(\theta)$  は, それぞれ  $\Gamma(\theta) = \hat{f}_n(\theta)$  と  $\gamma(\theta) = f(\theta)$  である (厳密に言えば, 後者は  $n$  に依存して  $\gamma^n(\theta) = \hat{f}_n(\theta)$  とすべきであるが, 簡明のため細かい議論は省略する).

最高点の近傍における  $f$  の滑らかさを考慮に入れた次の仮定をおく.

**条件 5.2** ある偶数  $p \geq 2$  に対し, 関数  $t \rightsquigarrow f(t)$  はある点  $t_0 \in (0, 1)$  の近傍  $N$  で  $p$ -回連続的微分可能であり, その導関数  $f^{(m)}$ ,  $m = 1, \dots, p$  は次を満たす:

(a) 各  $m = 1, \dots, p-1$  に対し  $f^{(m)}(t_0) = 0$ ;

(b)  $\sup_{t \in N} f^{(p)}(t) < 0$ .

推定したい未知母数は  $\theta_0 = t_0$  である. この条件のもとで,  $p$  に依存して, Taylor 展開により

$$f(\theta) - f(\theta_0) \approx f^{(p)}(\theta_0)|\theta - \theta_0|^p$$

という近似が成り立つ. 従って  $M$ -CRITERION の第一不等式は満たされる. 第二不等式をチェックするために, 2 節で与えた最大不等式を利用するが, その方法はカーネル関数の種類によって若干異なる. 具体的に結果のみ記すと, smooth kernel の場合は  $\phi^n(\delta) = n^{-1/2}b_n^{-2}\delta^{3/2}$  となり, monotone kernel の場合は  $\phi^n(\delta) = n^{-1/2}b_n^{-1}\delta^{1/2}$  となる. いずれにせよ, もしも bandwidth を  $b_n = n^{-1/(2p+1)}$  とおくと, 関係  $\phi^n(r_n^{-1}) \leq r_n^{-p}$  は  $r_n = n^{1/(2p+1)}$  とおくとことによって満たされる.

次に  $n^{1/(2p+1)}(\hat{\theta}_n - t_0)$  の漸近分布を導出しよう. 計算の便宜上, 一様カーネル  $K(x) = \frac{1}{2} \cdot 1_{[-1,1]}(x)$  を使用する. ステップ (A1) はすでに示されている. ステップ (A2) に関して

は、次のような分解ができる rescaled criterion function  $h \rightsquigarrow M^n(h)$  を考える.

$$\begin{aligned} M^n(h) &= b_n^{-p} \left\{ \widehat{f}_n(t_0 + b_n h) - f(t_0) \right\} \\ &= Y^n(h) + Z^n(h), \end{aligned}$$

ただし

$$\begin{aligned} Y^n(h) &= b_n^{-p} \{ f_n(t_0 + b_n h) - f(t_0) \}; \\ Z^n(h) &= b_n^{-p} \left\{ \widehat{f}_n(t_0 + b_n h) - f_n(t_0 + b_n h) \right\}. \end{aligned}$$

$\sqrt{nb_n} = b_n^{-p}$  に注意すると, 任意のコンパクト集合  $K \subset \mathbb{R}$  に対し  $Z^n \xrightarrow{P} Z$  in  $\ell^\infty(K)$ , ただし  $Z(h) = \{\mathbb{B}(h+1) - \mathbb{B}(h-1)\}/2$ , となる事が, 定理 4.2 から導かれる. ここに  $h \rightsquigarrow \mathbb{B}(h)$  は両側ブラウン運動である. 他方, 簡単な計算から, 各  $h \in \mathbb{R}$  に対し

$$\begin{aligned} \lim_{n \rightarrow \infty} Y^n(h) &= Y(h) \\ &= \frac{f^{(p)}(t_0)}{2p!} \int_{-1}^1 (h+y)^p dy \\ &= A(h)/2 \quad (\text{この式で } A(h) \text{ を定義する}) \end{aligned}$$

が成り立つ事がわかる.  $h \rightsquigarrow Y^n(h)$  と  $h \rightsquigarrow Y(h)$  はともに連続であるからこの収束は任意のコンパクト集合  $K$  上で連続である. それ故, 確率場  $h \rightsquigarrow M^n(h)$  は  $h \rightsquigarrow \{Y(h) + Z(h)\} \rightsquigarrow \ell^\infty(K)$  の中で弱収束する. これでステップ (A2) が示された. ステップ (A3) はいまの場合自明である. よって,  $\operatorname{argmax}_h M^n(h)$  は  $\operatorname{argmax}_h \{Y(h) + Z(h)\} \rightsquigarrow$  弱収束する. すなわち,  $n^{1/(2p+1)}(\widehat{\theta}_n - t_0) \xrightarrow{P} \operatorname{argmax}_h \{A(h) + \mathbb{B}(h+1) - \mathbb{B}(h-1)\}$  が得られた.

### 5.3 例: ジャンプ点の位置の推定

再び 4 節で考えた正規白色ノイズモデル (4.3) を考える. ここでは関数  $t \rightsquigarrow f(t)$  のジャンプの位置  $t_0$  の推定問題を考える.  $t_0$  の近傍における  $f$  の形状に関して次の仮定をおく.

**条件 5.3** ある  $t_0 \in (0, 1)$  に対し, ある定数  $a \in (0, 1/2)$  が存在して関数  $t \rightsquigarrow f(t)$  は区間  $[t_0 - a, t_0 + a]$  上で右連続かつ左極限をもち, しかも次の条件を満たす:

$$D = (R_* - L^*) - (L^* - L_*) \vee (R^* - R_*) > 0$$

ただし

$$\begin{aligned} L^* &= \sup_{t \in [t_0 - a, t_0)} f(t), & R^* &= \sup_{t \in [t_0, t_0 + a]} f(t), \\ L_* &= \inf_{t \in [t_0 - a, t_0)} f(t), & R_* &= \inf_{t \in [t_0, t_0 + a]} f(t). \end{aligned}$$

上で現れた定数  $a > 0$  は以下で定義する推定量を構成するために既知でなければならない。しかし、我々は関数  $t \rightsquigarrow f(t)$  の具体的な形は何も規定しない。定数  $D > 0$  の値は既知である必要はない。条件 5.3 は「関数  $t \rightsquigarrow f(t)$  は  $t_0$  において正のジャンプ、すなわち  $f(t_0) - f(t_0-) \geq R_* - L^*$  をもち、そしてそれが区間  $[t_0 - a, t_0 + a]$  の中の最大のジャンプである」という意味である。この解釈により、この仮定がいかに自然であるかがわかるであろう。

パラメータ空間を  $\Theta = [a, 1 - a]$  とおく。アイデアは「 $t \rightsquigarrow F(t) = \int_0^t f(s)ds$  の右側導関数と左側導関数の差が  $t_0$  において最大になるという事実を、それら導関数のカーネル型推定量を用いて表現する」という事である。そのために、定数  $b \in (0, a)$  を固定し、

$$(5.6) \quad w_\theta(t) = k_b(t - \theta) \quad \forall \theta \in [a, 1 - a],$$

ただし

$$k_b(x) = \begin{cases} -x - b, & x \in [-b, 0), \\ -x + b, & x \in [0, b], \\ 0, & \text{otherwise} \end{cases}$$

と定義する。採用する criterion process  $\theta \rightsquigarrow \Gamma^n(\theta)$  と contrast function  $\theta \rightsquigarrow \gamma(\theta)$  はそれぞれ

$$\begin{aligned} \Gamma^n(\theta) &= \int_0^1 w_\theta(t) dX_t^n, \\ \gamma(\theta) &= \int_0^1 w_\theta(t) f(t) dt \end{aligned}$$

である。結論からいえば、我々は  $n(\hat{\theta}^n - \theta_0) \xrightarrow{P} \operatorname{argmax}_{h \in \mathbb{R}} \{A(h) + B(h)\}$  in  $\mathbb{R}$  を得ることができる。ただし  $h \rightsquigarrow A(h)$  は

$$A(h) = \begin{cases} h \left\{ (2b)^{-1} \int_{t_0-b}^{t_0+b} f(t) dt - f(t_0) \right\}, & \forall h \geq 0, \\ h \left\{ (2b)^{-1} \int_{t_0-b}^{t_0+b} f(t) dt - f(t_0-) \right\}, & \forall h < 0, \end{cases}$$

によって与えられる deterministic process で、 $h \rightsquigarrow B(h)$  は両側ブラウン運動である。収束率の導出のポイントは、やはり次の 2 点である。まず、任意の  $\theta \in [t_0, t_0 + a - b]$  に対し

$$\gamma(\theta) - \gamma(\theta_0) \leq -|\theta - \theta_0| \{bD - |\theta - \theta_0|R_*\}$$

であり、任意の  $\theta \in [t_0 - a + b, t_0)$  に対し

$$\gamma(\theta) - \gamma(\theta_0) \leq -|\theta - \theta_0| \{bD + |\theta - \theta_0)L^*\}$$

である事から、 $M$ -CRITERION の第一不等式が  $p = 1$  について示される。一方 2 節の最大不等式を用いて、第二不等式が  $\phi^n(\delta) = n^{-1/2}\delta^{1/2}$  に対して示される。これらより収束率  $r_n = n$  が導かれる。証明の残る部分は、5.2 節と同様、criterion process  $\Gamma^n(\theta)$  の局所確率場の弱収束を示す事が鍵となる。詳細は省略する。

## 6 応用3: ノンパラメトリック MLE の収束率

この節では, 簡単のため正規白色ノイズモデルのみを考えるが, そのアプローチは, 非線形な確率微分方程式のモデルや点過程のモデルにも応用が可能である.

### 6.1 正規白色ノイズモデルにおける一般論

$\Theta$  は  $\mathcal{L}^2[0,1]$  の部分集合であるとする. 各  $n \in \mathbb{N}$  に対し  $t \rightsquigarrow X_t^n$  はある filtered measurable space  $(\Omega^n, \mathcal{F}^n, \mathbb{F}^n = (\mathcal{F}_t^n)_{t \in [0,1]})$  上で定義された連続適合過程であるとする.  $\mathbb{P}^n = \{P_\theta^n : \theta \in \Theta\}$  は  $\Theta$  によって添字づけられた  $(\Omega^n, \mathcal{F}^n)$  上の確率測度の族であるとする.  $X^n$  は  $P_\theta^n$  のもとで

$$dX_t^n = \theta(t)dt + n^{-1/2}dB_t^{n,\theta}, \quad X_0^n = x_0 \in \mathbb{R}$$

と表されるものとする. ただし  $t \rightsquigarrow B^{n,\theta}$  は  $(\Omega^n, \mathcal{F}^n, \mathbb{F}^n, P_\theta^n)$  上の標準ブラウン運動とする. ある条件の下で, 対数尤度比は

$$(6.7) \quad \log \frac{P_\theta^n | \mathcal{F}_1^n}{P_\vartheta^n | \mathcal{F}_1^n} = \int_0^1 \{\theta(t) - \vartheta(t)\} dX_t^n - \frac{1}{2} \left\{ \|\theta\|_{\mathcal{L}^2[0,1]}^2 - \|\vartheta\|_{\mathcal{L}^2[0,1]}^2 \right\} \quad \forall \theta, \vartheta \in \Theta$$

によって与えられる事はよく知られている (例えば Jacod and Shiryaev (1987) の Theorem III.5.34). よって最尤推定量 (MLE) は

$$(6.8) \quad \Gamma^n(\theta) = \int_0^1 \theta(t) dX_t^n - \frac{1}{2} \|\theta\|_{\mathcal{L}^2[0,1]}^2 \quad \forall \theta \in \mathcal{L}^2[0,1].$$

によって与えられる criterion process  $\theta \rightsquigarrow \Gamma^n(\theta)$  の最大点として定義される (“最大点” の存在自体が明らかではないので, 少し異なる正確な定義を後述の定理の中で与える). さて, データが与えられた真値  $\theta_0 \in \Theta$  から生成されたものであったとしよう. このとき, 対応する ( $P_{\theta_0}^n$  のもとでの) contrast function  $\theta \rightsquigarrow \gamma(\theta)$  は

$$(6.9) \quad \begin{aligned} \gamma(\theta) &= \int_0^1 \theta(t)\theta_0(t)dt - \frac{1}{2} \|\theta\|_{\mathcal{L}^2[0,1]}^2 \quad \forall \theta \in \mathcal{L}^2[0,1] \\ &= -\frac{1}{2} \|\theta - \theta_0\|_{\mathcal{L}^2[0,1]}^2 + \frac{1}{2} \|\theta_0\|_{\mathcal{L}^2[0,1]}^2. \end{aligned}$$

と定義するのが自然である. なぜなら, 残差  $\Gamma^n(\theta) - \gamma(\theta) = n^{-1/2} \int_0^1 \theta(t) dB_t^{n,\theta_0}$  が  $P_{\theta_0}^n$  のもとでマルチンゲールだからである. 我々は (6.7) が対数尤度比を与えるという事実を使わない. 従って,  $\Gamma^n(\theta)$  は単に (6.8) の右辺によって定義された確率変数であると考えられる. さらに, 我々は (6.8) において  $\Gamma^n(\theta)$  を  $\Theta$  の上だけではなく  $\mathcal{L}^2[0,1]$  の全ての要素について定義した. このことは, “sieved” MLE を考える事を可能にする. (6.9) と

$M$ -CRITERION の第一不等式の関係性を考慮すると,  $L^2$ -semimetric  $\|\cdot\|_{\mathcal{L}^2[0,1]}$  を  $\Theta$  上 (より広く,  $\mathcal{L}^2[0,1]$  上) の擬距離として採用するのが自然である.

いま, “sieve”  $\Theta^n$  とは,  $\mathcal{L}^2[0,1]$  の部分集合の列 ( $\Theta$  の部分集合とは限らない) であって,  $\Theta$  を “良く” 近似するようなものとする. 一般に, 大きな  $\Theta^n$  をとれば近似は “良く” できるが, 逆に確率的な制御は困難となる. そこで, 「sieve が大きすぎない」という意味のエントロピー条件と 「sieve が粗すぎない」という意味の近似条件を次のように導入する.

**条件 6.1 (エントロピー条件)** ある関数  $\varphi: (0, \infty) \rightarrow (0, \infty)$  であって  $\delta \sim \delta^{-1}\varphi(\delta)$  が非増加なるものが存在し, 十分大きな全ての  $n \in \mathbb{N}$  に対し

$$\int_0^\delta \sqrt{\log(1 + N(\Theta^n, \|\cdot\|_{\mathcal{L}^2[0,1]}; \varepsilon))} d\varepsilon \leq \varphi(\delta) \quad \forall \delta \in (0, \infty).$$

が成り立つ.

これが成り立つとき, 定数  $r_n > 0$  を  $n^{-1/2}\varphi(r_n^{-1}) \leq r_n^{-2}$  を満たすようにとる. 近似条件はこれに対して与えられる.

**条件 6.2 (近似条件)** ある定数  $M > 0$  が存在して, 十分大きな全ての  $n \in \mathbb{N}$  に対し

$$\Theta \subset \bigcup_{\theta \in \Theta^n} B(\theta, Mr_n^{-1})$$

が成り立つ. ただし  $B(\theta, \varepsilon)$  は,  $\theta$  を中心とする  $\|\cdot\|_{\mathcal{L}^2[0,1]}$  に関する半径  $\varepsilon$  の閉球を表す.

**定理 6.3** 条件 6.1 および 6.2 のもとで, 写像  $\widehat{\theta}_n: \Omega^n \rightarrow \Theta^n$  が

$$(6.10) \quad \Gamma^n(\widehat{\theta}_n) \geq \sup_{\theta \in \Theta^n} \Gamma^n(\theta) - r_n^{-2}$$

を満たすならば,

$$\lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} P_{\theta_0}^{n*} \left( r_n \|\widehat{\theta}_n - \theta_0\|_{\mathcal{L}^2[0,1]} > L \right) = 0$$

が成り立つ.

厳密には, この定理を証明するためには, 定理 5.1 のパラメータ集合  $\Theta$  を  $\Theta^n$  に取り替えたものを用意する必要がある. しかし, 大まかな流れは同じであるから, 詳細は省略する. 要約すると, 第一不等式は contrast function の定義 (6.9) から従い, また第二不等式は 2 節の最大不等式がそのまま適用できる形になっている, という事である.



## 6.2 例 1: 単調関数の族

$\Theta$  は  $[0, 1]$  から  $[0, 1]$  への単調関数の全体を表すものとする. すると, van der Vaart and Wellner (1996) の Theorem 2.7.5 より

$$\int_0^\delta \sqrt{\log N(\Theta, \|\cdot\|_{\mathcal{L}^2[0,1]}; \varepsilon)} d\varepsilon \leq \text{const.} \delta^{1/2} \quad \forall \delta > 0$$

が成り立つ. このことは, sieve  $\Theta^n$  を適切にとることにより, 条件 6.1 は  $\varphi(\delta) = \text{const.}(\delta^{1/2} \vee \delta)$  について成り立つようにするべきである, ということを示唆している. この場合, 収束率は  $r_n = n^{1/3}$  となり, この率に対して条件 6.2 が満たされるようにするべきである. 実際, 簡単な計算により, 次の命題が成り立つ.

**命題 6.4** 分点  $0 = t_0^n < t_1^n < \dots < t_n^n = 1$  を  $t_i^n - t_{i-1}^n \leq n^{-2/3}$  なるように選び, また  $V^n = \{j \cdot n^{-2/3} : j \in \mathbb{Z}\} \cap [0, 1]$  とおく.  $\Theta^n$  を  $[0, 1]$  から  $V^n$  への単調関数であって各区間  $[t_{i-1}^n, t_i^n)$  上で一定値をとるものであるとする. このとき  $\Theta \subset \bigcup_{\theta \in \Theta^n} B(\theta; \sqrt{2}n^{-1/3})$  が成り立つ.

この  $\Theta^n$  の構成の場合,  $\Theta^n \subset \Theta$  であるから, 条件 6.1 も実際満たされている. 結局, (6.10) をこのように構成した  $\Theta^n$  について満たす  $\hat{\theta}_n$  であれば,

$$\lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} P_{\theta_0}^{n*} \left( n^{1/3} \|\hat{\theta}_n - \theta_0\|_{\mathcal{L}^2[0,1]} > L \right) = 0$$

が成り立つ. この収束率  $n^{1/3}$  は, I.I.D. のノイズをもつ回帰関数の推定の文脈における “optimal rate” と一致している. 注意すべき点は,  $n^{-2/3}$  というオーダーの分点がこの収束率を得るためには十分であり, 従って確率過程  $t \rightsquigarrow X_t^n$  をこれらの分点上で離散的に観測できれば, その推定量を計算する事ができる, という事である.

## 6.3 例 2: 滑らかな関数の族

定数  $\alpha > 1/2$  と  $H > 0$  が与えられたとする.  $f : [0, 1] \rightarrow \mathbb{R}$  に対し  $\|f\|_\alpha$  を

$$\|f\|_\alpha = \max_{k \leq \alpha} \sup_{t \in (0,1)} |D^k f(t)| + \sup_{\substack{t,s \in (0,1) \\ t \neq s}} \frac{|D^k f(t) - D^k f(s)|}{|t - s|^{\alpha - k}}$$

と定義する. ただし  $\alpha$  は  $\alpha$  よりも真に小さい最大の整数であり, また

$$D^k f(t) = \frac{d^k f(t)}{dt^k}$$

である. 関数の族  $C_H^\alpha([0, 1])$  を  $\{f : [0, 1] \rightarrow \mathbb{R} : \|f\|_\alpha \leq H\}$  によって定義し, 真値の族  $\Theta = C_H^\alpha([0, 1])$  を考える. よく知られているように,

$$\int_0^\delta \sqrt{\log N(C_H^\alpha([0, 1]), \|\cdot\|_\infty; \varepsilon)} d\varepsilon \leq \text{const.} \delta^{1-(1/2\alpha)}$$

が成り立つので, 条件 6.1 は  $\varphi(\delta) = \text{const.}(\delta^{1-(1/2\alpha)} \vee \delta)$  に対して成り立つべきである. この場合, 収束率は  $r_n = n^{\alpha/(2\alpha+1)}$  となる.

分点  $0 = t_0^n < t_1^n < \dots < t_{k_n}^n = 1$  を,  $t_i^n - t_{i-1}^n \leq n^{-\alpha/(2\alpha+1)}$  を満たすように選び, 写像  $\pi^n : \Theta \rightarrow \ell^\infty[0, 1]$  を

$$\pi^n \theta(t) = \sum_{i=1}^{k_n} \frac{1_{[t_{i-1}^n, t_i^n)}(t)}{t_i^n - t_{i-1}^n} \int_{t_{i-1}^n}^{t_i^n} \theta(s) ds \quad \forall t \in [0, 1]$$

によって定義する. このとき次の 2 つの事実が成り立つ:

$$\|\pi^n \theta - \pi^n \vartheta\|_{\mathcal{L}^2[0,1]} \leq \|\theta - \vartheta\|_{\mathcal{L}^2[0,1]} \leq \|\theta - \vartheta\|_\infty \quad \forall \theta, \vartheta \in \Theta;$$

$$\|\theta - \pi^n \theta\|_{\mathcal{L}^2[0,1]} \leq \|\theta - \pi^n \theta\|_\infty \leq H n^{-\alpha/(2\alpha+1)} \quad \forall \theta \in \Theta.$$

従って, もしも  $\pi^n \Theta$  の可算部分集合の列  $\Theta^n$  であって  $\pi^n \Theta \subset \bigcup_{\theta \in \Theta^n} B(\theta; M n^{-\alpha/(2\alpha+1)})$  が適当な定数  $M > 0$  ( $n$  に依存しない) に対して成り立つものをとれば, 条件 6.1 と 6.2 は満たされ収束率  $r_n = n^{\alpha/(2\alpha+1)}$  を得る. 前の例と同様, この率は回帰分析の文脈における “optimal rate” に一致する. また, オーダー  $n^{-\alpha/(2\alpha+1)}$  の分点における離散的観測が得られれば, この収束率を達成するためには十分であることに注意すべきである.

## 7 応用 4: 積分型推定量の漸近正規性

各  $k = 1, \dots, n$  に対し,  $t \rightsquigarrow Z_t^k$  は  $Z_0^k$  を出発し, Lévy 測度

$$L(dt, dz) = dt \alpha(t, dz)$$

をもつような 1 次元 Lévy 過程であるとする. 過程  $t \rightsquigarrow Z_t^k$  はランダム時間区間  $[\sigma^k, \tau^k] \subset [0, 1]$  でのみ観測可能であるとする. ただし  $\sigma^k$  と  $\tau^k$  は停止時刻である. 典型的な例は  $Z_0^k > 0$ ,  $\sigma^k = 0$  かつ  $\tau^k = \inf\{t \geq 0 : Z_t^k \leq 0\} \wedge 1$  である. 右センサーが入った場合, すなわち,  $\sigma^k = 0$  かつ  $\tau^k = \inf\{t \geq 0 : Z_t^k \leq 0\} \wedge C^k \wedge 1$  を考えることもできる. しかし, 我々の状況設定はもっと一般的である. 実際, 我々が仮定するのは以下で出てくる条件 (7.12) だけである.

整数値ランダム測度

$$N^k(\omega; dt, dz) = \sum_{s \in (\sigma^k(\omega), \tau^k(\omega))} 1_{\{\Delta Z_s^k(\omega) \neq 0\}} \varepsilon_{(s, \Delta Z_s^k(\omega))}(dt, dz)$$

を導入する. ただし  $\varepsilon_a$  は点  $a$  における Dirac 測度を表す. このとき,  $N^k$  の予測可能カンペンセイターは

$$1_{] \sigma^k, \tau^k ]}(\omega, t) dt \alpha(t, dz)$$

によって与えられる.  $k$  に関して和をとることにより, 整数値ランダム測度

$$\mu^n(\omega; dt, dz) = \sum_{k=1}^n N^k(\omega; dt, dz)$$

を得る. ただし  $Z^k$  たちは独立であると仮定する.  $\mu^n$  の予測可能カンペンセイターは

$$\nu^n(\omega; dt, dz) = Y_t^n(\omega) dt \alpha(t, dz)$$

によって与えられる. ただし

$$Y_t^n(\omega) = \sum_{k=1}^n 1_{\llbracket \sigma^k, \tau^k \rrbracket}(\omega, t)$$

である. 我々は Nelson-Aalen 推定量を使って  $\alpha$  を推定することを考える. そのために, Nishiyama (2000a, 2000b, 2007) によるマルチンゲール中心極限理論を使う.

$[0, 1] \times \mathbb{R}$  上で定義された可測関数の可算族  $\Psi$  が与えられたとする. ある可測関数  $\varphi$  が存在して  $|\psi| \leq \varphi$  が全ての  $\psi \in \Psi$  に対して成り立っていることを仮定する.

$$\int_{[0,1] \times \mathbb{R}} \varphi(t, z) \vee \varphi(t, z)^2 dt \alpha(t, dz) < \infty$$

を仮定しておく. 我々は実際には

$$(t, \psi) \rightsquigarrow A(t, \psi) = \int_{[0,t] \times \mathbb{R}} \psi(s, z) ds \alpha(s, dz), \quad (t, \psi) \in [0, 1] \times \Psi$$

という汎関数を推定することを考える.

セミノルム

$$\|\psi\|_{L^2(\alpha)} = \sqrt{\int_{[0,1] \times \mathbb{R}} |\psi(t, z)|^2 dt \alpha(t, dz)}$$

に関する bracketing 数を  $N_{[\cdot]}(\Psi, \|\cdot\|_{L^2(\alpha)}; \varepsilon)$  によって表す. 我々は bracketing エントロピー条件

$$(7.11) \quad \int_0^1 \sqrt{\log N_{[\cdot]}(\Psi, \|\cdot\|_{L^2(\alpha)}; \varepsilon)} d\varepsilon < \infty$$

を仮定する.

我々は Nelson-Aalen 推定量

$$(t, \psi) \rightsquigarrow \hat{A}^n(t, \psi) = \int_{[0,t] \times \mathbb{R}} \psi(s, z) 1_{\mathbb{R} \setminus [-c_n, c_n]}(z) Y_s^{n-} \mu^n(ds, dz),$$

を提案する. ただし

$$Y_s^{n-} = \begin{cases} 1/Y_s^n, & \text{if } Y_s^n \geq 1, \\ 0, & \text{otherwise,} \end{cases}$$

であり,  $\{c_n\}$  は適当な非負定数列であるとする.

**定理 7.1** Bracketing エントロピー条件 (7.11) が成り立つと仮定する. ある真に正である可測関数  $y(t)$  であって,

$$(7.12) \quad \sup_{t \in [0,1]} |nY_t^{n-} - 1/y(t)| \rightarrow 0 \quad \text{in probability, as } n \rightarrow \infty$$

を満たすものが存在するとする. さらに

$$(7.13) \quad \sqrt{n} \int_{[0,t] \times \mathbb{R}} \varphi(s, z) 1_{[-c_n, c_n]}(z) ds \alpha(s, dz) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

を仮定する.

このとき  $\sqrt{n}(\hat{A}^n - A)$  は  $n \rightarrow \infty$  とするとき  $\ell^\infty([0, 1] \times \Psi)$  の中で, 平均ゼロで共分散が

$$E(G(t, \psi)G(s, \phi)) = \int_{[0, t \wedge s] \times \mathbb{R}} \frac{\psi(s, z)\phi(s, z)}{y(s)} ds \alpha(s, dz)$$

によって与えられる正規確率過程  $G$  に分布収束する. さらに  $(t, \psi) \rightsquigarrow G(t, \psi)$  の殆ど全てのパスは  $\rho((t, \psi), (s, \phi)) = |t - s| \vee \|\psi - \phi\|_{L^2(\alpha)}$  によって与えられる擬距離  $\rho$  に関し一様連続である.

さらにこれが convolution theorem (例えば van der Vaart and Wellner (1996) の Theorem 3.11.2) や asymptotic minimax theorem (同じく Theorem 3.11.5) の意味において漸近有効であることも示される. 詳しい議論は Nishiyama (2000b) の 4 章や Nishiyama (2006) を参照されたい.

## 参考文献

- [1] Bae, J. and Levental, S. (1995a). Uniform CLT for Markov chains and its invariance principle: a martingale approach. *J. Theoret. Probab.* **8** 549-570.
- [2] Bae, J. and Levental, S. (1995b). A uniform CLT for continuous martingales. *J. Korean Statist. Soc.* **24** 225-231.
- [3] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré Probab. Statist.* **31** 393-427.
- [4] Dudley, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899-929. Correction, **7** 909-911.
- [5] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin Heidelberg.

- [6] Nishiyama, Y. (1997). Some central limit theorems for  $\ell^\infty$ -valued semimartingales and their applications. *Probab. Theory Relat. Fields* **108** 459-494.
- [7] Nishiyama, Y. (1999). A maximal inequality for continuous martingales and  $M$ -estimation in a Gaussian white noise model. *Ann. Statist.* **27** 675-696.
- [8] Nishiyama, Y. (2000a). Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **28** 685-712.
- [9] Nishiyama, Y. (2000b). *Entropy Methods for Martingales*. CWI Tract **128**, Centrum voor Wiskunde en Informatica, Amsterdam.
- [10] Nishiyama, Y. (2007). On the paper "Weak convergence of some classes of martingales with jumps". *To appear in Ann. Probab.* **35** No.2 or 3.
- [11] Nishiyama, Y. (August 2006). Nonparametric estimation and testing time homogeneity for Lévy processes. *Research Memorandum 997*, The Institute of Statistical Mathematics, Tokyo.
- [12] Ossiander, M. (1987). A central limit theorem under metric entropy with  $L_2$  bracketing. *Ann. Probab.* **15** 897-919.
- [13] Pollard, D. (2002). Maximal inequalities via bracketing with adaptive truncation. *Ann. I. H. Poincaré – PR* **38** 1039-1052.
- [14] van de Geer, S. (1995). Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.* **23** 1779-1801.
- [15] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York.
- [16] van der Vaart, A.W. and van Zanten, H. (2005). Donsker theorems for diffusions: necessary and sufficient conditions. *Ann. Probab.* **33** 1422-1451.

## 第II部：金融リスクと統計的極値論

3. 「One-parameter Families of Distortion Risk Measures」 Hideatsu Tsukahara (塚原英敦)
4. 「経験尤度法による極値パラメターの推定」 大和田孝
5. 「Extreme quantile estimation using extreme value theory」 Tatsuyoshi Okimoto (沖本竜義)
6. 「コピュラの保険への応用例」 岡賢一
7. 「コピュラの性質とその応用」 広瀬光久

# One-parameter Families of Distortion Risk Measures

Hideatsu Tsukahara\*

Department of Economics, Seijo University, Japan

and

Department of Mathematics, ETH Zürich, Switzerland

February 18, 2006

## Abstract

This paper introduces parametric families of distortion risk measures, investigates their properties, and discusses its use in risk management. Their derivation is based on Kusuoka's representation theorem of law invariant and comonotonically additive coherent risk measures. Much simplified proofs of some known results under slightly weaker condition are also given. Our approach is then to narrow down a tractable class of risk measures by requiring their comparability with the traditional expected shortfall. We make numerical comparison among them, and propose a method of estimating the value of the distortion risk measures based on data. Their use and interpretation in risk management will also be discussed.

**KEY WORDS:** coherent risk measure, convex risk measure, distortion, insurance premium principle, parametric family.

---

\*The author would like to thank Freddy Delbaen for making the preprint by Jouini, Schachermayer and Touzi (2005) available, suggesting a proof of Lemma 2.3, and his helpful comments. The author's e-mail address is [tsukahar@seiyo.ac.jp](mailto:tsukahar@seiyo.ac.jp).

# 1 Introduction

The problem of quantifying risk led us to the notion of risk measure. The value at risk (VaR) is the first example of such a measure and is now the standard measure used in practice. An axiomatic approach by Artzner et al. (1997, 1999) suggests, however, that the VaR fails to satisfy one of the properties that any *coherent* risk measure should possess: namely the subadditivity, which reflects the reduction of risk by diversification. The expected shortfall has emerged as a coherent risk measure with many nice features (Acerbi and Tasche (2002) and Tasche (2002)).

In practice, depending on one's purpose and the type of risk to be handled, it would be desirable to have more coherent risk measures with some nice characteristics. In financial economics, there are many examples of utility functions, many of which are parametrized by a few number of parameters, to express different investors' attitude towards risk. The same thing applies to risk measures; risk managers' assessment of risk may be subjective, and we would need various risk measures to reflect their subjective degree and form of conservatism.

In this paper, we narrow down a good class of risk measures by imposing several conditions successively. First, we use the usual axioms of coherence; then additional conditions called *law invariance* and *comonotonic additivity* are imposed. By Kusuoka representation (Kusuoka (2001)), we arrive at the class of distortion risk measures of the following form: for a random variable  $X$  representing loss,

$$(1.1) \quad \rho(X) = \int_{[0,1]} F_X^{-1}(u) dD(u) = \int_{\mathbb{R}} x dD \circ F_X(x).$$

where  $D$  is a *convex distortion function*.  $F_X$  denotes the distribution function (df) of  $X$  and  $F_X^{-1}$  is its *quantile* defined by

$$F_X^{-1}(u) := \inf\{x \in \mathbb{R} : F_X(x) \geq u\}, \quad 0 < u < 1.$$

We note that the essentially same class of risk measures has been introduced in Acerbi (2002, 2004) under the name *spectral measures of risk*.

Here is the precise definition of distortion function:

**Definition 1.1** Any df  $D$  on  $[0, 1]$  is called a *distortion function*: Namely a distortion function  $D$  is a right-continuous increasing function on  $[0, 1]$  with values in  $[0, 1]$  such that  $D(0) = 0$  and  $D(1) = 1$ .

Given a distortion  $D$  and a df  $F$ ,  $G = D \circ F$  is again a df, which we will call the *distorted distribution* under  $D$ . Notice that a distortion can be viewed as a transformation on the space of df's. And the expectation under the distorted df  $G$ , i.e.,  $\int x dG(x)$ , will be called the *distorted expectation* under  $D$ . A distorted expectation is also called the *Choquet integral* (Choquet (1955)), the *non-additive integral* (Denneberg (1994)), or the *fuzzy*



*integral* (Grabisch et al. (2000)). We would not use these terms because our distorted expectation is a proper integral; we distort a df, not a probability measure.

The concept of distortion has been recently popularized in financial and actuarial research: Denneberg (1990) introduced the idea of distortion in constructing premium principles, and Wang (1996, 2000) developed it further. However, it has appeared much earlier in other fields. In economics of measuring the inequality in income distributions, a special class of distortion functions appears as *Lorenz curve* (see e.g., Kleiber and Kotz (2003)). Furthermore the connection with the dual theory of choice under risk in economics has now been well known. As an alternative theory to the expected utility theory, Yaari (1987) proposed the dual theory of choice under risk. Instead of the usual independence axiom, he puts the dual independence axiom, and proved that any preference relation satisfying his axioms can be represented as a distorted expectation. See also Schmeidler (1989) in this connection. In statistical literature, the distorted df's are known as *Lehmann alternatives*, which were introduced in Lehmann (1953) and generalized in estimation context by Miura and Tsukahara (1993).

The class of risk measures in (1.1) is still broad, and in order to get more explicit form, we parametrize distortion  $D$  to get new risk measures. Our approach is to take the expected shortfall as the benchmark, and put the same conditions as the expected shortfall satisfies. In other words, we would like the parametrized family  $\{D_\theta\}$  to be comparable to the one for the expected shortfall. With some regularity conditions, it is shown that  $D_\theta$  of the form

$$(1.2) \quad D_\theta(u) = \Psi(\Psi^{-1}(u) + \log \theta),$$

where  $\Psi$  is a strictly increasing, continuous df on  $\mathbb{R}$ , has many nice properties, including comparability with the expected shortfall.

In the next section, we first recall some definitions in the theory of risk measures. Assuming throughout that the underlying probability space is atomless, we give a much shorter and simpler proof of two facts; the first one, due to Jouini, Schachermayer and Touzi (2005), states that the Fatou property automatically follows from the law invariance for convex risk measures. The second one is the above-mentioned Kusuoka representation with the form suitable for us.

After showing that the convex distortion transformation preserves some stochastic orders, we introduce, in Section 3, parametric families of distortion functions in the form (1.2). We use an elementary result on the translation equation to identify the class (1.2) as the one having required properties. We argue that the class (1.2) of distortion functions can be interpreted in financial terms, and give examples such as proportional hazards and proportional odds distortions.

Section 4 discusses some computational issues. In applications, we need to examine the condition(s) for  $\rho(X)$  to be finite for unbounded  $X$ . We take some well-known distributions and distortions to illustrate this point. Some classes are shown to have undesirable properties that the range of permissible parameter values is restricted severely

by the tail of the underlying distribution. Also we display some numerical figures of several distortion risk measures including VaR and the expected shortfall so that one can see the difference among those risk measures. Finally we discuss briefly some methods of estimating the value of the distorted expectation based on real or simulated i.i.d. data. A natural estimator is a simple form of  $L$ -statistics; hence one can use the existing theory to investigate stability and efficiency of the estimators. Section 5 concludes the paper with some remarks.

## 2 Coherent risk measures

### 2.1 Notation and Terminology

We first recall the definition of coherence of risk measures according to Artzner et al. (1997, 1999). Let  $(\Omega, \mathcal{F}, \mathbf{P})$  be a probability space, and  $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbf{P})$  be the set of all  $\mathbf{P}$ -essentially bounded random variables. In what follows, a random variable  $X$  represents the loss, not the value of a financial position.

**Definition 2.1** A *coherent risk measure* is a functional  $\rho: L^\infty \rightarrow \mathbb{R}$  satisfying the following four properties:

[PO] (**positivity**): If  $X \leq 0$ , then  $\rho(X) \leq 0$ .

[PH] (**positive homogeneity**): For any  $\lambda > 0$ ,  $\rho(\lambda X) = \lambda\rho(X)$ .

[SA] (**subadditivity**):  $\rho(X + Y) \leq \rho(X) + \rho(Y)$ .

[TE] (**translation equivariance**): For any  $c \in \mathbb{R}$ ,  $\rho(X + c) = \rho(X) + c$ .

For a nice representation theorem, the following continuity property is needed.

[FA] (**Fatou property**): If the  $X_n$  are uniformly bounded in absolute value by 1 and  $X_n \xrightarrow{\mathbf{P}} X$ , then  $\rho(X) \leq \liminf \rho(X_n)$ .

Delbaen (2002) proved that any coherent risk measure with the Fatou property can be represented as

$$\rho(X) = \sup \{ \mathbf{E}^{\mathbf{Q}}(X) : \mathbf{Q} \in \mathcal{Q} \},$$

where  $\mathcal{Q}$  is a set of probability measures and each member of  $\mathcal{Q}$  is absolutely continuous with respect to  $\mathbf{P}$ .

Next we need the following definition:

**Definition 2.2** Random variables  $X_1, \dots, X_d$  are said to be *comonotone* if one of the following equivalent statements holds:

- (i) For every  $i \neq j$ ,  $(X_i(\omega) - X_i(\omega'))(X_j(\omega) - X_j(\omega')) \geq 0$  for  $\mathbf{P} \otimes \mathbf{P}$ -almost all  $(\omega, \omega')$ .

- (ii) The joint df of  $X_1, \dots, X_d$  is given by the upper Fréchet bound  $F_{x_1}(x_d) \wedge \dots \wedge F_{x_d}(x_d)$ .
- (iii) For  $U \sim U(0, 1)$ ,  $(X_1, \dots, X_d)$  and  $(F_{x_1}^{-1}(U), \dots, F_{x_d}^{-1}(U))$  have the same distribution.
- (iv) There exist a random variable  $Z$  and increasing functions  $f_1, \dots, f_d$  such that  $(X_1, \dots, X_d)$  and  $(f_1(Z), \dots, f_d(Z))$  have the same distribution.

Note that we define the comonotonicity as a distributional property rather than a property of random variables. See Dhaene et al. (2002) for an overview of the concept of comonotonicity in actuarial science and finance.

Now we introduce two more axioms:

- [LI] (**law invariance**): For any two random variables  $X$  and  $Y$  with the same law, we have  $\rho(X) = \rho(Y)$ .
- [CA] (**comonotonic additivity**):  $\rho(X + Y) = \rho(X) + \rho(Y)$  if  $X$  and  $Y$  are comonotone.

It is strongly desirable that we are able to estimate the value of risk measures statistically (as also claimed in Acerbi (2002)), and to this end the law invariance [LI] is necessary. When  $X$  and  $Y$  are comonotone, they move together in one direction with probability 1, so, as rightly described in Yaari (1986), neither of them is a hedge against the other. The comonotonic additivity [CA] then means that for such two risks, we should have additivity instead of subadditivity, which is a very natural requirement.

Recently the concept of convex risk measure has been introduced by Föllmer and Schied (2002a, 2002b). A *convex risk measure* is a functional  $\rho: L^\infty \rightarrow \mathbb{R}$  satisfying the translation equivariance [TE] and

- [MO] (**monotonicity**): If  $X \leq Y$ , then  $\rho(X) \leq \rho(Y)$ .
- [CX] (**convexity**):  $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$  for  $0 \leq \lambda \leq 1$ .

Note that every coherent risk measure is a convex risk measure and that [TE] and [MO] imply (see Föllmer and Schied (2002a))

$$(2.1) \quad |\rho(X) - \rho(Y)| \leq \|X - Y\|_\infty.$$

## 2.2 Fatou property and law invariance

Jouini, Schachermayer and Touzi (2005) proved that for convex risk measures, the Fatou property [FA] automatically follows from the law invariance [LI]. Here we give a direct and shorter proof of this result (with a slightly weaker condition on the underlying probability space) although our method is somewhat similar to theirs. Let us recall that a set  $A \in \mathcal{F}$  is called an *atom* of  $\mathbf{P}$  if  $\mathbf{P}(A) > 0$  and for every  $B \subset A$  with  $B \in \mathcal{F}$ , one has either  $\mathbf{P}(B) = 0$  or  $\mathbf{P}(B) = \mathbf{P}(A)$ . Then  $(\Omega, \mathcal{F}, \mathbf{P})$  is called *atomless* if  $\mathbf{P}$  does not have any atoms.

We first prove the following lemma, which may be of independent interest.

**Lemma 2.3** Suppose that  $(\Omega, \mathcal{F}, \mathbf{P})$  is atomless. Let  $(X_n)$  be a sequence of random variables uniformly bounded in absolute value by 1, and  $X_n \xrightarrow{\mathbf{P}} X$ . Then one can find a triangular array of random variables  $Y_{nk}$ ,  $k = 1, \dots, N_n$  and  $n \in \mathbb{N}$  on  $(\Omega, \mathcal{F}, \mathbf{P})$  with the following two properties:

(i) For each  $n$ ,  $Y_{nk}$ ,  $k = 1, \dots, N_n$  has the same distribution as  $X_n$ .

(ii)  $\frac{1}{N_n} \sum_{k=1}^{N_n} Y_{nk} \rightarrow X$  in  $L^\infty$ .

*Proof (due to Delbaen).* Fix any  $m \in \mathbb{N}$  and set

$$A_j = \{\omega \in \Omega: j/m \leq X(\omega) < (j+1)/m\}$$

for  $j = -m, -m+1, \dots, -1, 0, 1, \dots, m-2$ . For  $j = m-1$ , we define  $A_{m-1}$  by the above equation with  $<$  replaced by  $\leq$ . Without loss of generality, we may assume  $\mathbf{P}(A_j) > 0$  for all  $j$  (otherwise define everything to be 0 on such sets of measure 0). Now, for some  $n_m$  large enough, we have

$$\mathbf{P}(\{|X - X_n| > 1/m\} \cap A_j) \leq \mathbf{P}(A_j)/m$$

for all  $j = -m, -m+1, \dots, -1, 0, 1, \dots, m-1$  and  $n \geq n_m$ . We can take the sequence  $(n_m)_{m \in \mathbb{N}}$  to be increasing. Let  $C_{jn} = \{|X_n - j/m| > 2/m\} \cap A_j$ . Then  $\mathbf{P}(C_{jn}) \leq \mathbf{P}(A_j)/m$ . Using the assumption that  $\mathbf{P}$  has no atom, we can divide  $A_j$  into  $m$  sets  $B_j^1, \dots, B_j^m$  of equal  $\mathbf{P}$ -probability so that  $C_{jn} \subset B_j^1$  and

$$\mathbf{P}(B_j^i | A_j) = 1/m, \quad i = 1, \dots, m.$$

Applying Proposition 6.9 in Delbaen (2002) to the restrictions of  $\mathbf{P}$  to the  $B_j^i$ , for each  $i$  and  $j$ , we can construct  $V_{nk}$  so that, for  $k = 1, \dots, m$ ,

$$\mathbf{P}(\{V_{nk} \leq y\} \cap B_j^i) = \mathbf{P}(\{X_n \leq y\} \cap B_j^{i+k-1}).$$

Here and in the rest of the proof, we use the convention ‘mod  $m$ ’ in the superscript of  $B$ . It then follows that

$$\begin{aligned} \mathbf{P}(V_{nk} \leq y) &= \sum_{j=-K}^{K-1} \sum_{i=1}^m \mathbf{P}(\{V_{nk} \leq y\} \cap B_j^i) \\ &= \sum_{j=-K}^{K-1} \sum_{i=1}^m \mathbf{P}(\{X_n \leq y\} \cap B_j^{i+k-1}) = \mathbf{P}(X_n \leq y). \end{aligned}$$

This shows that the  $V_{nk}$  have the same distribution as  $X_n$  under  $\mathbf{P}$ . Moreover, by construction,  $\mathbf{P}$ -a.s. on  $A_j$

$$\left| V_{nk}(\omega) - \frac{j}{m} \right| \leq \begin{cases} 2, & \text{if } \omega \in B_j^{m+2-k} \\ 2/m, & \text{otherwise.} \end{cases}$$

Therefore,  $\mathbf{P}$ -a.s. on  $A_j$ ,

$$\left| X - \frac{1}{m} \sum_{k=1}^m V_{nk} \right| \leq \left| X - \frac{j}{m} \right| + \left| \frac{j}{m} - \frac{1}{m} \sum_{k=1}^m V_{nk} \right| \leq \frac{1}{m} + \frac{m-1}{m} \frac{2}{m} + \frac{2}{m} < \frac{5}{m}$$

Note now that the  $V_{nk}$ 's in fact depend on  $m$ ; thus write  $V_{nk}^m$ . What we have shown is that for each  $m$ , there exists an  $n_m$  such that for all  $n \geq n_m$ , we can construct random variables  $V_{n1}^m, \dots, V_{nm}^m$  satisfying that, for each  $n$ ,  $V_{nk}^m$ ,  $k = 1, \dots, m$  has the same distribution as  $X_n$ , and  $|X - (1/m) \sum_{k=1}^m V_{nk}^m| \leq 5/m$ . For each  $n$ , there exists a unique  $m$  satisfying  $n_m \leq n < n_{m+1}$ . Then letting  $N_n$  equal to this  $m$  and define  $Y_{nk} = V_{nk}^m$ . It is clear that the  $Y_{nk}$  thus defined satisfy the required properties. ■

Now we can easily prove the following theorem.

**Theorem 2.4** *Assume that  $(\Omega, \mathcal{F}, \mathbf{P})$  is atomless. Then every law invariant convex risk measure  $\rho(X)$  satisfies the Fatou property.*

*Proof.* Let  $|X_n| \leq 1$  for every  $n$  and  $\omega$ , and  $X_n \xrightarrow{\mathbf{P}} X$ . Let  $Y_{nk}$  be a triangular array of random variables satisfying the properties stated in Lemma 2.3. Write

$$Y_n = \frac{1}{N_n} \sum_{k=1}^{N_n} Y_{nk}.$$

The convexity of  $\rho$  and the law invariance yield

$$\rho(Y_n) = \rho\left(\frac{1}{N_n} \sum_{k=1}^{N_n} Y_{nk}\right) \leq \frac{1}{N_n} \sum_{k=1}^{N_n} \rho(Y_{nk}) = \rho(X_n).$$

We use (2.1) and take limit inferior on both sides to get

$$\rho(X) = \lim_{n \rightarrow \infty} \rho(Y_n) \leq \liminf_{n \rightarrow \infty} \rho(X_n),$$

as required. ■

### 2.3 Representation of law invariant, comonotonically additive coherent risk measures

Under the assumption that  $(\Omega, \mathcal{F}, \mathbf{P})$  is a *standard* probability space, Kusuoka (2001) showed that every law invariant, comonotonically additive coherent risk measure can be represented as the expectation of the risk under a convexly distorted distribution. Here we present his result in the form suitable for us and give an easy proof under a slightly weakened assumption of atomlessness, emphasizing the distorted expectation aspect of the risk measures. As a preparation, we recall the following results:

**Proposition 2.5** Let  $(X, Y)$  be random variables with joint df  $H(x, y)$ , and let  $F$  and  $G$  denote the marginal df's of  $X$  and  $Y$  respectively.

(i) **Hoeffding's lemma:** Suppose  $\mathbf{E}|X|$ ,  $\mathbf{E}|Y|$  and  $\mathbf{E}|XY|$  are all finite. Then

$$\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [H(x, y) - F(x)G(y)] dx dy$$

(ii) **Fréchet-Hoeffding bounds:** We have

$$(F(x) + G(y) - 1) \vee 0 \leq H(x, y) \leq F(x) \wedge G(y),$$

where  $a \vee b$  and  $a \wedge b$  denote the maximum and minimum of  $a$  and  $b$  respectively.

For proofs of (i) and (ii), see Lehmann (1966), or Lemma 5.24 and Theorem 5.7 of McNeil et al. (2005).

Now we can give a simple proof of the following theorem due to Kusuoka (2001).

**Theorem 2.6** Suppose that  $(\Omega, \mathcal{F}, \mathbf{P})$  is atomless and that a coherent risk measure  $\rho(X)$  satisfies [LI] and [CA]. For  $X \in L^\infty$ , let  $F_X$  be the df of  $X$  under  $\mathbf{P}$  and  $F_X^{-1}$  be its quantile. Then there exists a convex distortion  $D$  such that

$$(2.2) \quad \rho(X) = \int_{[0,1]} F_X^{-1}(u) dD(u) = \int_{\mathbb{R}} x dD \circ F_X(x).$$

Conversely,  $\rho$  given by (2.2) with a convex distortion  $D$  is a coherent risk measure satisfying [LI] and [CA].

*Proof.* Let  $\mathbf{Q} \ll \mathbf{P}$  and put  $Y_{\mathbf{Q}} = d\mathbf{Q}/d\mathbf{P}$ . We denote by  $F_{Y_{\mathbf{Q}}}$  be the df of  $Y_{\mathbf{Q}}$  under  $\mathbf{P}$ . Also let  $F$  be the joint df of  $X$  and  $Y_{\mathbf{Q}}$  under  $\mathbf{P}$ . Then, for  $X \in L^\infty$ , by Hoeffding's lemma (apply twice) and Fréchet-Hoeffding upper bound, we get

$$\begin{aligned} \mathbf{E}^{\mathbf{Q}}(X) &= \mathbf{E}^{\mathbf{P}}(XY_{\mathbf{Q}}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x, y) - F_X(x)F_{Y_{\mathbf{Q}}}(y)] dx dy - \mathbf{E}^{\mathbf{P}}(X)\mathbf{E}^{\mathbf{P}}(Y_{\mathbf{Q}}) \\ &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F_X(x) \wedge F_{Y_{\mathbf{Q}}}(y) - F_X(x)F_{Y_{\mathbf{Q}}}(y)] dx dy - \mathbf{E}^{\mathbf{P}}(X)\mathbf{E}^{\mathbf{P}}(Y_{\mathbf{Q}}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy [F_X(dx) \wedge F_{Y_{\mathbf{Q}}}(dy)] = \int_0^1 F_X^{-1}(u)F_{Y_{\mathbf{Q}}}^{-1}(u) du, \end{aligned}$$

where  $F_{Y_{\mathbf{Q}}}^{-1}$  is the quantile function of  $Y_{\mathbf{Q}}$ . We have thus proved that for every  $X \in L^\infty$  and  $\mathbf{Q} \ll \mathbf{P}$ ,

$$(2.3) \quad \mathbf{E}^{\mathbf{Q}}(X) \leq \int_0^1 F_X^{-1}(u)F_{Y_{\mathbf{Q}}}^{-1}(u) du$$

(with the equality when  $X$  and  $Y_{\mathbf{Q}}$  are comonotone). In passing we remark that the same argument with the lower bound would yield a proof of *Hardy-Littlewood inequality*; see Föllmer and Schied (2002c), Theorem 2.76.

Now let  $\mathcal{P}_0 = \{\mathbf{Q}: \mathbf{Q} \ll \mathbf{P}, \mathbf{E}^{\mathbf{Q}}(X) \leq \rho(X) \text{ for all } X \in L^\infty\}$ . It follows from Theorem 3.2 of Delbaen (2002) and (2.3) that

$$(2.4) \quad \rho(X) = \sup_{\mathbf{Q} \in \mathcal{P}_0} \mathbf{E}^{\mathbf{Q}}(X) \leq \sup_{\mathbf{Q} \in \mathcal{P}_0} \int_0^1 F_X^{-1}(u) F_{Y_{\mathbf{Q}}}^{-1}(u) du.$$

On the other hand, since  $(\Omega, \mathcal{F}, \mathbf{P})$  is atomless, for any  $\mathbf{Q} \ll \mathbf{P}$  and  $X \in L^\infty$ , we can construct  $\tilde{X}$  which has the same distribution as  $X$  and is comonotone with  $Y_{\mathbf{Q}}$ . It then follows that for  $\mathbf{Q} \in \mathcal{P}_0$ ,

$$\rho(X) = \rho(\tilde{X}) \geq \mathbf{E}^{\mathbf{Q}}(\tilde{X}) = \mathbf{E}^{\mathbf{P}}(\tilde{X} Y_{\mathbf{Q}}) = \int_0^1 F_X^{-1}(u) F_{Y_{\mathbf{Q}}}^{-1}(u) du,$$

and hence

$$\rho(X) \geq \sup_{\mathbf{Q} \in \mathcal{P}_0} \int_0^1 F_X^{-1}(u) F_{Y_{\mathbf{Q}}}^{-1}(u) du.$$

But we have shown in (2.4) the reverse inequality, so the equality must hold in the above:

$$(2.5) \quad \rho(X) = \sup_{\mathbf{Q} \in \mathcal{P}_0} \int_0^1 F_X^{-1}(u) F_{Y_{\mathbf{Q}}}^{-1}(u) du.$$

Define

$$(2.6) \quad D(u) := \int_0^u F_{Y_{\mathbf{Q}}}^{-1}(v) dv.$$

Since  $Y_{\mathbf{Q}}$  is positive, so is  $F_{Y_{\mathbf{Q}}}^{-1}$ , whence  $D$  is increasing. And  $D$  has an increasing density  $F_{Y_{\mathbf{Q}}}^{-1}(v)$ , so it is convex. Finally  $D(1) = \mathbf{E}^{\mathbf{P}}(Y_{\mathbf{Q}}) = 1$ . Thus  $D$  is a convex distortion.

Let  $\mathcal{D}_{\text{cx}}$  denote the set of convex distortions. We shall prove the following two claims:

(i)  $\mathcal{D}_{\text{cx}}$  is compact with respect to the topology of weak convergence.

(ii)  $D \mapsto \int_{\mathbb{R}} x dD \circ F(x)$  is continuous on  $\mathcal{D}_{\text{cx}}$  for each df  $F$  of  $X \in L^\infty$ .

Let  $\mathcal{D}[0, 1]$  be the set of df's on  $[0, 1]$ , which is compact. Since  $\mathcal{D}_{\text{cx}} \subset \mathcal{D}[0, 1]$ , we need only to show that  $\mathcal{D}_{\text{cx}}$  is closed. Suppose that  $D_n \in \mathcal{D}_{\text{cx}}$  and  $D_n \xrightarrow{d} D$ . Denoting by  $C$  the set of all continuity points of  $D$ , this means that  $D_n(u) \rightarrow D(u)$  for every  $u \in C$ . If  $u$  and  $v$  belongs to  $C$  and if  $0 < \lambda < 1$  is such that  $\lambda u + (1 - \lambda)v \in C$ , then we clearly have

$$D(\lambda u + (1 - \lambda)v) \leq \lambda D(u) + (1 - \lambda)D(v).$$

But  $C$  is dense in  $[0, 1]$  (in fact,  $C^c$  is countable), so the right-continuity of  $D$  establishes the above inequality for all  $u, v$  in  $[0, 1]$  and  $0 < \lambda < 1$ , i.e.,  $D \in \mathcal{D}_{\text{cx}}$ . Note that  $v = 1$

cannot be approached from the right, but we always have  $D(1) = 1$ . This shows the compactness of  $\mathcal{D}_{\text{cx}}$ .

For (ii), note that  $D \in \mathcal{D}_{\text{cx}}$  is continuous on  $[0, 1)$  because it is convex and  $D(0) = 0$ . Thus the only possible discontinuity point of any  $D \in \mathcal{D}_{\text{cx}}$  is  $u = 1$ . Now suppose  $D_n \xrightarrow{d} D$  in  $\mathcal{D}_{\text{cx}}$ . Then for any  $x$  with  $F(x) < 1$ , one has  $D_n \circ F(x) \rightarrow D \circ F(x)$ . Also  $D_n(1) = 1$  and  $D(1) = 1$ , so  $D_n \circ F(x) \rightarrow D \circ F(x)$  holds for  $x$  with  $F(x) = 1$  as well. Thus obviously  $D_n \circ F \xrightarrow{d} D \circ F$ . Since  $X$  is bounded a.s., this entails that

$$\int_{\mathbb{R}} x dD_n \circ F(x) \rightarrow \int_{\mathbb{R}} x dD \circ F(x),$$

which proves (ii).

What we have shown in (2.5) can be rephrased as follows: there exist  $\mathcal{D}_0 \subset \mathcal{D}_{\text{cx}}$  such that

$$(2.7) \quad \rho(X) = \sup_{D \in \mathcal{D}_0} \int_{[0,1]} F_X^{-1}(u) dD(u).$$

For the rest of the proof, we may argue as in Kusuoka (2001) (see his Theorem 7 and Proposition 21). We include the proof for completeness. If we put

$$\mathcal{D}_\rho = \left\{ D \in \mathcal{D}_{\text{cx}} : \int_{[0,1]} F_X^{-1}(u) dD(u) \leq \rho(X) \text{ for all } X \in L^\infty \right\},$$

$\mathcal{D}_0$  in (2.7) may clearly be replaced by  $\mathcal{D}_\rho$ . By (ii) shown above,  $\mathcal{D}_\rho$  is closed, and hence compact. Therefore for each fixed  $X \in L^\infty$ , one can find a  $D_X \in \mathcal{D}_\rho$  for which one has  $\rho(X) = \int_{[0,1]} F_X^{-1}(u) dD_X(u)$ . Let

$$\mathcal{D}_\rho(F_X) = \left\{ D \in \mathcal{D}_\rho : \int_{[0,1]} F_X^{-1}(u) dD(u) = \rho(X) \right\}.$$

This is clearly closed, and what we have shown amounts to  $\mathcal{D}_\rho(F_X) \neq \emptyset$ .

Now we make the following assertion: For any finite number of df's  $F_1, \dots, F_k$  with bounded supports,  $\mathcal{D}_\rho(F_1), \dots, \mathcal{D}_\rho(F_k)$  has a nonempty intersection. To prove this, let  $X_1, \dots, X_k$  be comonotone random variables with df's  $F_1, \dots, F_k$  respectively (just let  $X_i = F_i^{-1}(U)$  for a uniform random variable  $U$ ). Put  $X = \sum_{i=1}^k X_i$ , and denote its df by  $F$ . Then by [CA], we have  $\rho(X) = \sum_{i=1}^k \rho(X_i)$ .

Note that since  $X_1, \dots, X_k$  are comonotone, we have  $F^{-1} = \sum_{i=1}^k F_i^{-1}$  (see Denneberg (1994) or Dhaene et al. (2002) for a proof). If we choose  $D \in \mathcal{D}_\rho(F)$ , then

$$\sum_{i=1}^k \int_{[0,1]} F_i^{-1}(u) dD(u) = \int_{[0,1]} F^{-1}(u) dD(u) = \rho(X) = \sum_{i=1}^k \rho(X_i)$$

On the other hand, we must have,

$$\int_{[0,1]} F_i^{-1}(u) dD(u) \leq \rho(X_i), \quad i = 1, \dots, k.$$



Hence the equality must hold in each of the above inequalities, which implies  $D \in \bigcap_{i=1}^k \mathcal{D}_\rho(F_i)$ . We have proved that  $\{\mathcal{D}_\rho(F) : F \text{ df of } X \in L^\infty\}$  has the finite intersection property in  $\mathcal{D}_{\text{cx}}$ , which is compact, Thus there exists an element  $\widehat{D} \in \mathcal{D}_{\text{cx}}$  which belongs to  $\mathcal{D}_\rho(F)$  for every df  $F$  of  $X \in L^\infty$ . Namely, for any  $X \in L^\infty$ ,

$$\rho(X) = \int_{[0,1]} F_X^{-1}(u) d\widehat{D}(u).$$

This proves the first half of Theorem 2.6.

Finally, suppose that  $\rho$  is of the form (2.2). Then  $\rho$  is clearly a law invariant coherent risk measure. If  $X$  and  $Y$  are comonotone, then  $F_{X+Y}^{-1} = F_X^{-1} + F_Y^{-1}$  as we noted above. Thus  $\rho$  is comonotonically additive. ■

Theorem 2.6 suggests that we should look at risk measures of the form (2.2):

$$\rho(X) = \int_0^1 F_X^{-1}(u) dD(u) = \int_{-\infty}^{\infty} x dD \circ F_X(x),$$

the distorted expectation under  $D$ , if we are willing to accept the very plausible requirements of coherence, law invariance and comonotonic additivity. In fact, we know more from (2.6); the distortion function  $D$  has a density which is the quantile of the df of the Radon-Nikodym density  $Y_{\mathbf{Q}}$ .

The reason that we stick to the expression above and not the original form of Kusuoka representation is as follows: We can regard  $F_X^{-1}$  as the *rearrangement* of  $X$  in increasing order in the sense of Hardy et al. (1952), Section 10.12. Since  $X$  represents the loss, the larger values in the unit interval  $(0, 1)$  correspond to unfavorable events. Thus if  $D$  in  $\rho(X)$  above put much weight to the the values near 1, it reveals decision maker's conservative attitude in assessing the risk  $X$ . In this way, intuitive interpretation of the distortion function  $D$  is possible.

Our next task is to go one step further to introduce parametric families of convex distortions for the practical use of risk measures such as capital allocation rules, portfolio selection, and portfolio performance evaluation. But before we go on, let us note that the renowned value at risk and the conditional tail expectation may be represented as a distorted expectation.

**Example 2.7 (Value at risk)** Taking  $D(u) = \mathbf{1}_{[1-\alpha, 1]}(u)$ ,  $0 < \alpha < 1$  produces the well-known value at risk (VaR), which is just the  $(1 - \alpha)$ -th quantile  $F_X^{-1}(1 - \alpha)$ . It is well known (Artzner et al. (1997)) that the VaR is not a coherent risk measure. From the distortion point of view, it is because the corresponding distortion  $D$  is *not* convex.

**Example 2.8 (Expected Shortfall)** For  $0 < \alpha < 1$ , let

$$D_\alpha^{\text{ES}}(u) = \frac{1}{\alpha} [u - (1 - \alpha)]_+,$$

where  $a_+ = a \vee 0$ . Then the resulting risk measure is the renowned expected shortfall:

$$\text{ES}_\alpha(X) = \frac{1}{\alpha} \int_{1-\alpha}^1 F_X^{-1}(u) du.$$

One can easily verify that for  $0 < \alpha_1, \alpha_2 < 1$ ,  $D_\alpha^{\text{ES}}$  satisfies

$$(2.8) \quad D_{\alpha_1}^{\text{ES}} \circ D_{\alpha_2}^{\text{ES}} = D_{\alpha_1 \alpha_2}^{\text{ES}}.$$

Also we have

$$(2.9) \quad D_1^{\text{ES}}(u) = u, \quad \lim_{\alpha \downarrow 0} D_\alpha^{\text{ES}}(u) = \mathbf{1}_{\{1\}}(u)$$

so that  $\text{ES}_1(X) = \mathbf{E}(X)$  and  $\lim_{\alpha \downarrow 0} \text{ES}_\alpha(X) = \text{ess sup } X$ .

Here we note that many premium principles in actuarial mathematics can be written as a distorted expectation. Wang (1996) is considered to be the first to propose the distortion approach to premium principles although Denneberg (1990) had already indicated possible generalization using distorted probabilities. See also Wang (1995). He also showed that for a convex  $D$ , the distortion premium principle possesses many desirable properties as a premium principle. Wang et al. (1997) take the axiomatic approach similar to that given above and prove the analogous representation theorem for premium principles.

## 2.4 Preservation of Some Stochastic Orders

If  $D$  is a convex distortion, then, as a transformation on the space of df's, it preserves some well-known stochastic orders. Recall the following definitions:

**Definition 2.9** Let  $F$  and  $G$  be df's on  $\mathbb{R}$ .

- (i)  $G$  is called *stochastically larger than*  $F$  ( $F \preceq_{\text{st}} G$ ) if  $G(x) \leq F(x)$  for all  $x \in \mathbb{R}$ .

For the next two stochastic orders, we assume that the means of  $F$  and  $G$  exists.

- (ii)  $F$  is called *less dangerous than*  $G$  ( $F \preceq_D G$ ) if there exists  $x_0$  such that  $F(x) \leq G(x)$  for  $x < x_0$  and  $G(x) \leq F(x)$  for  $x \geq x_0$ , and if  $\int x dF(x) \leq \int x dG(x)$ .
- (iii)  $G$  is called *larger than*  $F$  in increasing convex order ( $F \preceq_{\text{icx}} G$ ) if  $\int h(x) dF(x) \leq \int h(x) dG(x)$  for every increasing convex function  $h$  on  $\mathbb{R}$  for which the integrals exist.

Note that the increasing convex order is also known as *stop loss order* in actuarial mathematics. See Müller and Stoyan (2002) for properties of these stochastic orders.

**Theorem 2.10** Let  $D \in \mathcal{D}_{\text{cx}}$ . Then we have

- (i) If  $F \preceq_{st} G$ , then  $D \circ F \preceq_{st} D \circ G$ .
- (ii) If  $F \preceq_D G$ , then  $D \circ F \preceq_D D \circ G$ .
- (iii) If  $F \preceq_{icx} G$ , then  $D \circ F \preceq_{icx} D \circ G$ .

For (ii) and (iii), we assume that the means of  $F$ ,  $G$ ,  $D \circ F$  and  $D \circ G$  all exist.

*Proof.* (i) is trivial (we do not need convexity for (i)).

(ii). Let  $F \preceq_D G$ , and  $x_0$  be a point such that  $F(x) \leq G(x)$  for  $x < x_0$  and  $G(x) \leq F(x)$  for  $x \geq x_0$ . Then clearly  $D \circ F(x) \leq D \circ G(x)$  for  $x < x_0$ , and  $D \circ G(x) \leq D \circ F(x)$  for  $x \geq x_0$ . So we need only to prove  $\int x dD \circ F(x) \leq \int x dD \circ G(x)$ . First suppose that  $0 < G(x_0) < 1$ . Note that

$$\int_{-\infty}^{\infty} x dD \circ G(x) - \int x dD \circ F(x) = \int_{-\infty}^{\infty} [D \circ F(x) - D \circ G(x)] dx =: I_1 + I_2,$$

where

$$I_1 := \int_{-\infty}^{x_0} [D \circ F(x) - D \circ G(x)] dx, \quad I_2 := \int_{x_0}^{\infty} [D \circ F(x) - D \circ G(x)] dx.$$

Since  $F(x) \leq G(x)$  for  $x \in (-\infty, x_0)$ , we have  $D \circ G(x) - D \circ F(x) \leq d(G(x))(G(x) - F(x))$ , where  $d$  is the left derivative of  $D$ . Then

$$I_1 \geq -d(G(x_0)) \int_{-\infty}^{x_0} [G(x) - F(x)] dx.$$

One can show in a similarl manner that

$$I_2 \geq d(G(x_0)) \int_{x_0}^{\infty} [F(x) - G(x)] dx.$$

Hence,

$$I_1 + I_2 \geq d(G(x_0)) \int_{-\infty}^{\infty} [F(x) - G(x)] dx \geq 0$$

because  $d$  is positive and  $\int_{-\infty}^{\infty} [F(x) - G(x)] dx = \int x dG(x) - \int x dF(x) \geq 0$ .

If  $G(x_0) = 1$ , then we must have  $F(x) \leq G(x)$  for all  $x \in \mathbb{R}$ . On the other hand,  $0 \leq \int x dG(x) - \int x dF(x) = \int_{-\infty}^{\infty} [F(x) - G(x)] dx$  implies that  $F \equiv G$ .

Finally, if  $G(x_0) = 0$ , then  $F = G \equiv 0$  on  $(-\infty, x_0)$ . This means  $F(x) \geq G(x)$  for all  $x \in \mathbb{R}$ , i.e.,  $F \preceq_{st} G$ . By (i),  $D \circ F \preceq_{st} D \circ G$  then holds, and so we get  $\int x dF(x) \leq \int x dG(x)$ .

(iii) Define the dual  $D^*$  of  $D$  by  $D^*(u) = 1 - D(1 - u)$ . Then the survivor function of the distorted df  $D \circ F$  is given by  $D^* \circ \bar{F}$  By Theorem 1.5.7 of Müller and Stoyan (2002), to prove  $D \circ F \preceq_{icx} D \circ G$ , it is enough to show that

$$\int_t^{\infty} D^* \circ \bar{F}(x) dx \leq \int_t^{\infty} D^* \circ \bar{G}(x) dx$$

holds for each  $t \in \mathbb{R}$ .

Convexity of  $D$  implies that  $D^*$  is concave, so there exists a sequence of piecewise linear concave distortions  $D_n^*$  such that  $D_n^* \uparrow D^*$  pointwise. Such a  $D_n^*$  can be written as a convex combination of functions having the form  $(u/\alpha) \wedge 1$ ,  $0 < \alpha \leq 1$  (In fact, this is precisely the dual of the distortion associated with the expected shortfall in Example 2.8). Therefore, in view of monotone convergence theorem, we have only to show, for a fixed  $t \in \mathbb{R}$  and  $0 < \alpha \leq 1$ ,

$$(2.10) \quad \int_t^\infty \left( \frac{\bar{F}(x)}{\alpha} \wedge 1 \right) dx \leq \int_t^\infty \left( \frac{\bar{G}(x)}{\alpha} \wedge 1 \right) dx.$$

Note that  $\bar{G}(x) \leq \alpha$  iff  $x \geq G^{-1}(1 - \alpha)$ . So we separate two cases: When  $t \geq G^{-1}(1 - \alpha)$ , (2.10) follows easily. Using the assumption  $F \preceq_{\text{icx}} G$ , we get

$$\int_t^\infty \left( \frac{\bar{G}(x)}{\alpha} \wedge 1 \right) dx = \int_t^\infty \frac{\bar{G}(x)}{\alpha} dx \geq \int_t^\infty \frac{\bar{F}(x)}{\alpha} dx \geq \int_t^\infty \left( \frac{\bar{F}(x)}{\alpha} \wedge 1 \right) dx.$$

When  $t < G^{-1}(1 - \alpha)$ , splitting the range of the integral at  $G^{-1}(1 - \alpha)$  yields

$$\begin{aligned} \int_t^\infty \left( \frac{\bar{G}(x)}{\alpha} \wedge 1 \right) dx &= \int_t^{G^{-1}(1-\alpha)} 1 dx + \int_{G^{-1}(1-\alpha)}^\infty \frac{\bar{G}(x)}{\alpha} dx \\ &\geq \int_t^{G^{-1}(1-\alpha)} \left( \frac{\bar{F}(x)}{\alpha} \wedge 1 \right) dx + \int_{G^{-1}(1-\alpha)}^\infty \left( \frac{\bar{F}(x)}{\alpha} \wedge 1 \right) dx \\ &= \int_t^\infty \left( \frac{\bar{F}(x)}{\alpha} \wedge 1 \right) dx, \end{aligned}$$

as required. ■

**Remark 2.11** Another way of proving (iii) is to apply Theorem 1.5.19 in Müller and Stoyan (2002), together with (ii) although a careful treatment of the step involving limits is necessary.

When  $X$  and  $Y$  have df's  $F$  and  $G$  respectively, we write  $X \preceq_{\text{st}} Y$  when  $F \preceq_{\text{st}} G$ ; and similarly for  $X \preceq_{\text{D}} Y$  and  $X \preceq_{\text{icx}} Y$ . Since  $X \preceq_{\text{st}} Y$  implies  $X \preceq_{\text{D}} Y$ , which in turn implies  $X \preceq_{\text{icx}} Y$ , we have the following corollary.

**Corollary 2.12** *Any of  $X \preceq_{\text{st}} Y$ ,  $X \preceq_{\text{D}} Y$  or  $X \preceq_{\text{icx}} Y$  implies  $\rho_D(X) \leq \rho_D(Y)$ .*

**Remark 2.13** A proof of the above corollary in Wang (1996) [Theorem 1] contains an error. One can find a similar statement in Yaari (1987). A simpler proof is also given in Leitner (2005). In the literature, this is a common statement concerning the preservation of stochastic order by risk measures. Note, however, that Theorem 2.10 is a stronger result in that if  $D \in \mathcal{D}_{\text{cx}}$ , then  $F \preceq_{\text{icx}} G$  implies  $\int h(x) dD \circ F(x) \leq \int h(x) dD \circ G(x)$  for all convex increasing  $h$ , not just for  $h(x) = x$ .

### 3 One-parameter Families of Distortions

In this section, we construct parametric families of distortions by putting some more restrictions. This produces several families of law invariant, comonotonically additive coherent risk measures.

We use the expected shortfall as a guide; remark in particular that the distortion function  $D_\alpha^{\text{ES}}$  satisfies (2.8) and (2.9). Then the following conditions seem to be a natural requirement for a family of distortion functions  $\{D_\theta\}$ :

[A.1] For each  $\theta > 0$ ,  $u \mapsto D_\theta(u)$  is a strictly increasing, continuous distortion;

[A.2] For each  $u \in (0, 1)$ ,  $\theta \mapsto D_\theta(u)$  is strictly increasing and continuous;

[G]  $D_{\theta_1} \circ D_{\theta_2} = D_{\theta_1 \theta_2}$  for  $\theta_1, \theta_2 > 0$ .

[A.1] and [A.2] are rather regularity conditions in nature, but the latter means that  $\{D_\theta\}$  is ordered by  $\theta$  in a nice way; it implies that for any df  $F$ ,  $D_{\theta_1} \circ F \preceq_{\text{st}} D_{\theta_2} \circ F$  (in the strict sense) if  $\theta_1 < \theta_2$ . Also note that  $D$  is necessarily (absolutely) continuous on  $(0, 1)$  when  $D \in \mathcal{D}_{\text{cx}}$  (which will be considered shortly). They may seem mathematically convenient because the class of all strictly increasing, continuous distortions forms a group under composition. Finally, [G] means that  $\{D_\theta\}$  is closed under composition as in the case of the expected shortfall (“G” stands for *group*).

If we impose the above conditions, then a certain class of distortion functions emerges, as shown in the following theorem.

**Theorem 3.1** *Assume that  $\{D_\theta\}$  satisfies [A.1], [A.2] and [G]. Then there exists a strictly increasing, continuous df  $\Psi$  on  $\mathbb{R}$  such that  $D_\theta$  can be written in the following form*

$$(3.1) \quad D_\theta(u) = \Psi(\Psi^{-1}(u) + \log \theta).$$

*Conversely, the family  $\{D_\theta\}$  given by (3.1) with a strictly increasing, continuous df  $\Psi$  on  $\mathbb{R}$  satisfies [A.1], [A.2] and [G].*

The converse part is obvious. To prove the direct part, we make the transformation  $\mu = \log \theta$ , and define  $\tilde{D}_\mu(u) = D_{e^\mu}(u)$ . Then  $\tilde{D}_\mu$  satisfies  $\tilde{D}_{\mu_1} \circ \tilde{D}_{\mu_2} = \tilde{D}_{\mu_1 + \mu_2}$ . This is called the *translation equation* in the functional equation theory. See Section 6.1 of Aczél (1966) for the details.

Solving this is in fact easy: Fix  $u_0 \in (0, 1)$  arbitrarily and put  $\Psi(x) := D(u_0, x)$ ,  $x \in \mathbb{R}$ . Then, by [G], we have  $\Psi(x + \mu) = D(\Psi(x), \mu)$ . [A.2] ensures that  $\Psi$  has a well-defined inverse, so we get  $D(u, \mu) = \Psi(\Psi^{-1}(u) + \mu)$ . It remains to check that  $\lim_{x \rightarrow -\infty} \Psi(x) = 0$  and  $\lim_{x \rightarrow \infty} \Psi(x) = 1$ , which may be shown by a simple *reductio ad absurdum* using [A.1].

**Definition 3.2** *If a one-parameter family of distortions  $\{D_\theta : \theta > 0\}$  satisfies [A.1], [A.2] and [G], we call it a *one-parameter distortion semigroup*.*

With this definition, Theorem 3.1 states that a one-parameter distortion semigroup  $\{D_\theta\}$  must be of the form  $D_\theta(u) = \Psi(\Psi^{-1}(u) + \log \theta)$  for some continuous and strictly increasing df  $\Psi$  on  $\mathbb{R}$ .

Note that we have  $D_1(u) = u$  and that  $\lim_{\theta \downarrow 0} D_\theta(u) = 1_{\{1\}}(u)$  as the distortion function for the expected shortfall (see (2.8) and (2.9)). This form of distortion has been considered by Dabrowska et al. (1989) in the context of transformation models in statistics, and by Arnold et al. (1987) to construct families of Lorenz curves in economics.

Members of a one-parameter distortion semigroup  $\{D_\theta\}$  are not convex in general. To get a simple condition for it, let us suppose that  $\Psi$  in (3.5) is twice differentiable and denote the first and second derivatives by  $\psi$  and  $\psi'$  respectively. Then the distortion  $D_\theta(u)$  is also twice differentiable with respect to  $u$ , and we have

$$d_\theta(u) := \frac{\partial}{\partial u} D_\theta(u) = \frac{\psi(x + \log \theta)}{\psi(x)},$$

and

$$d'_\theta(u) := \frac{\partial^2}{\partial u^2} D_\theta(u) = \frac{\psi'(x + \log \theta) - \psi(x + \log \theta)\psi'(x)/\psi(x)}{[\psi(x)]^2}.$$

For notational simplicity, we have set  $x = \Psi^{-1}(u)$ . Since we are assuming that  $\Psi$  is twice differentiable,  $D_\theta(u)$  is convex in  $u$  if and only if its second derivative  $d'_\theta(u)$  is positive. From the above expression of  $d'_\theta(u)$ , one finds that

$$d'_\theta(u) \geq 0 \quad \text{for all } u \in (0, 1) \quad \iff \quad \frac{\psi'(x + \log \theta)}{\psi(x + \log \theta)} \geq \frac{\psi'(x)}{\psi(x)} \quad \text{for all } x \in \mathbb{R}$$

If the density  $\psi$  is strongly unimodal (i.e.,  $\log \psi$  is concave), then  $(\psi'/\psi)$  is decreasing, so  $0 < \theta \leq 1$  is a necessary and sufficient condition for the convexity of  $D_\theta(u)$  in  $u$ .

Summarizing, *if we choose  $\Psi$  with a strongly unimodal density in the one-parameter distortion semigroup*

$$D_\theta(u) = \Psi(\Psi^{-1}(u) + \log \theta),$$

*then  $D_\theta$  is convex for  $0 < \theta \leq 1$ , and*

$$(3.2) \quad \rho_\theta(X) := \int_{[0,1]} F_X^{-1}(u) dD_\theta(u) = \int_{\mathbb{R}} x dD_\theta \circ F_X(x)$$

*is a coherent risk measure satisfying [LI] and [CA]. Also we have*

$$D_1(u) = u, \quad \lim_{\theta \downarrow 0} D_\theta(u) = 1_{\{1\}}(u), \quad \text{and} \quad D_{\theta_1} \circ D_{\theta_2} = D_{\theta_1 \theta_2}.$$

Thus we have established the ground on which we can reasonably compare our distorted risk measure given by (3.2) with the expected shortfall.

Before we turn to numerical and statistical comparison of those risk measures, we make an attempt to interpret the distorted risk measures (3.2), and then give some examples. Note that if  $\theta_1 \geq \theta_2$ , then  $\rho_{\theta_1}(X) \leq \rho_{\theta_2}(X)$ , and  $\mathbf{E}(X) \leq \rho_{\theta}(X) \leq \text{ess sup } X$ ; smaller values of  $\theta$  thus represents a more pessimistic view of the decision maker, and the most optimistic [resp. pessimistic] evaluation of risk corresponds to the case  $\theta = 1$  [resp.  $\theta \downarrow 0$ ] in which one has  $\rho_{\theta}(X) = \mathbf{E}(X)$  [resp.  $\rho_{\theta}(X) = \text{ess sup } X$ ].

It can be easily seen that the structure of one-parameter distortion groups imply the following proportional relation of a certain quantity derived from the original df  $F$  and the distorted df  $G = D_{\theta} \circ F$ . Let  $\kappa(u) = \exp[\Psi^{-1}(u)]$ . Then we have

$$(3.3) \quad \kappa(G(x)) = \theta \kappa(F(x)),$$

The above proportional relation gives us some idea about the interpretation of the parameter  $\theta$ .

Now we give some examples of one-parameter distortion semigroup. It is a straightforward calculation to verify that  $\psi$  is strongly unimodal in all of the following examples.

**Example 3.3 (proportional odds)** Taking  $D_{\theta}^{\text{PO}}(u) = \theta u[1 - u + \theta u]^{-1}$  yields the proportional odds  $\kappa(u) = u/(1 - u)$  (odds ratio). It is in the form (3.1) with  $\Psi(x) = 1/(1 + e^{-x})$ , i.e., the standard logistic df. McCullagh (1980) first introduced this model in statistics.

When  $X$  represents the loss of a financial institution, it is natural to define the event  $\{X > x_0\}$  as insolvency for some fixed  $x_0$ . Then denoting the df of  $X$  by  $F$ , the odds

$$\frac{F(x_0)}{1 - F(x_0)} = \frac{\mathbf{P}(X \leq x_0)}{\mathbf{P}(X > x_0)}$$

represents the degree to which the event of solvency is likely to happen. On the other hand, the above proportional distortion yields

$$\frac{G(x_0)}{1 - G(x_0)} = \theta \frac{F(x_0)}{1 - F(x_0)}$$

Hence we can interpret this distortion as follows: No matter what value is used to define solvency, the odds for solvency under the distorted distribution is proportional to that under the actual distribution. Using very small values for the proportionality constant  $\theta$ , thus lowering the odds for solvency, the decision maker distorts the distribution to express his/her conservatism.

**Example 3.4 (proportional hazards)** The family  $D_{\theta}^{\text{PH}}(u) = 1 - (1 - u)^{\theta}$  corresponds to the proportional hazards relation:  $\kappa(u) = -\log(1 - u)$  (cumulative hazard function). The associated  $\Psi$  is given by  $\Psi(x) = 1 - \exp\{-e^x\}$  (Gumbel df). This is a well-known model in statistics (Cox (1972)). See Wang (1995) for applications of this family to actuarial science.

**Example 3.5 (proportional  $\gamma$ -odds)** If we take the distortion

$$D_{\theta}^{\text{PGO}}(u) = \begin{cases} 1 - \left[ \frac{\theta(1-u)^{\gamma}}{1 - (1-u)^{\gamma} + \theta(1-u)^{\gamma}} \right]^{1/\gamma} & \gamma > 0, \\ 1 - (1-u)^{1/\theta} & \gamma = 0, \end{cases}$$

we have the proportional  $\gamma$ -odds;  $\kappa$  is given by the  $\gamma$ -odds ratio:

$$\kappa(u) = \begin{cases} [(1-u)^{-\gamma} - 1]/\gamma & \text{for } \gamma > 0, \\ -\log(1-u) & \text{for } \gamma = 0. \end{cases}$$

Note that the case  $\gamma = 1$  corresponds to Example 3.3 and that the case  $\gamma = 0$  is defined by the limit operation  $\gamma \downarrow 0$ . It is straightforward to check that the choice

$$\Psi(x) = 1 - (1 + \gamma e^x)^{-1/\gamma}$$

in (3.1) gives this distortion (it may be called exp-GPD (generalized Pareto distribution)). Clayton and Cuzick (1986) is the first to investigate this model for the regression problem in statistics.

**Example 3.6 (Gaussian distortion)** To price both financial and insurance risks, Wang (2000) chose the standard normal distribution function  $\Phi$  as  $\Psi$  in a one-parameter distortion semigroup:

$$D_{\theta}^{\text{GA}}(u) = \Phi(\Phi^{-1}(u) + \log \theta).$$

According to the discussion above, with this distortion, the df  $F$  of a loss variable  $X$  and its distorted df  $G$  are related as

$$\exp[\Phi^{-1}(G(x))] = \theta \exp[\Phi^{-1}(F(x))],$$

but it is not clear how to interpret the quantity  $\exp[\Phi^{-1}(F(x))]$ .

This model has been mentioned in Dabrowska et al. (1989) in studying semiparametric parameter estimation for a class of transformation models.

Wang (2000) asserts that using the Gaussian distortion together with the (log-)normality assumption, one can interpret the parameter  $\mu = \log \theta$  as the Sharpe ratio, which is based on the mean-variance CAPM theory, and Black-Scholes pricing formula for put and call options can be written as the expectation under the distorted distribution. But this is a coincidence due to the shift nature of the normal distributions, and it does not mean that the arbitrage pricing theory and the CAPM are special cases of the distortion “pricing”. The CAPM equation is a consequence of equilibrium pricing with quadratic utility or under multivariate normality, and the Black-Scholes formula is derived from no-arbitrage argument with the geometric Brownian motion assumption.

Figure 1 below shows the graphs of distortion densities in the above examples. From this, we can see visually how these distortions put weights in a different way.



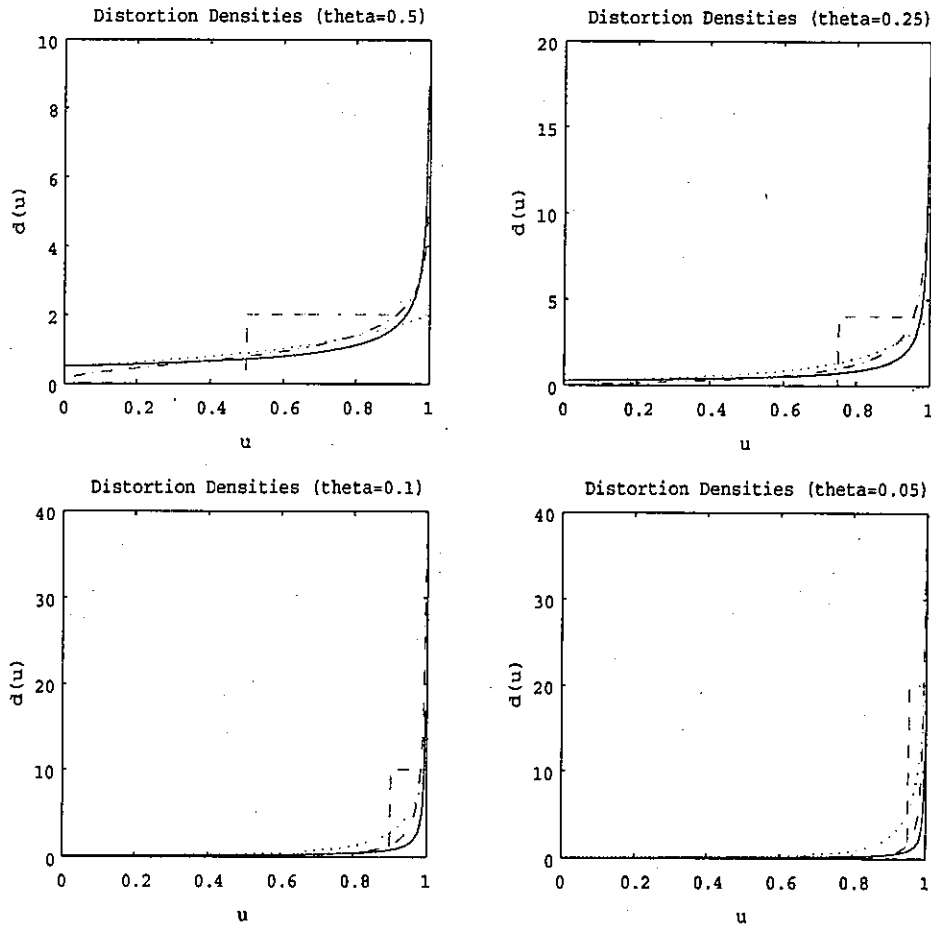


Figure 1: Distortion densities with  $\theta = 0.5, 0.25, 0.1, 0.05$ ; dashed line (--) for ES, solid line (—) for PH, dotted line (···) for PO, dash-dot line (-·-) for GA.

### Mixtures of extremes

Here we quickly illustrate the second approach to constructing parametric families of distortion functions.

Let  $c_k(\theta) \geq 0$ ,  $k = 1, 2, \dots$  satisfy  $\sum c_k(\theta) = 1$ , and set

$$(3.4) \quad D_\theta(u) = \sum_{k=1}^{\infty} c_k(\theta) u^k.$$

Since  $[F(x)]^k$  is the df of the maximum of  $k$  i.i.d. random variables with df  $F$ , the distorted df  $G = D_\theta \circ F$  may be interpreted as a mixture of extremes. Since the powers  $u^k$ ,  $k = 1, 2, \dots$ , are convex, so is the distortion (3.4), and hence the resulting risk measure is coherent.

**Example 3.7**

- (i) If we take  $c_i(\theta)$  from Bin(1,  $\theta$ ), then  $D_\theta(u) = (1 - \theta)u + \theta u^2$ , where  $0 \leq \theta \leq 1$ . (Lehmann (1952))
- (ii) If we take  $c_i(\theta)$  from positive Poisson distribution, i.e.,

$$c_i(\theta) = \frac{\theta^i}{(e^\theta - 1)i!}, \quad \theta > 0,$$

then we obtain

$$D_\theta(u) = \frac{e^{\theta u} - 1}{e^\theta - 1}.$$

We can put  $D_1(u) = u$  by taking a limit (see Ferguson (1967), Section 5.7). A special case of this example also appears in Delbaen (2002).

Also if we take  $c_i(\theta) = (1 - \theta)^{i-1}\theta$ , i.e. geometric distribution, then an easy calculation shows that we get the proportional odds distortion of Example 3.4. Thus these two approaches are not mutually exclusive.

## 4 Computation and Statistical Estimation

### 4.1 Computation

We assumed in Section 2 that, for theoretical reasons, the random variable  $X$  representing the loss belongs to  $L^\infty$ . In applications, however, we need to use  $\rho_\theta$  even when  $X$  is presumably unbounded. Therefore, before any computation, we need to make sure that

$$\rho_\theta(X) = \int_{[0,1]} F_X^{-1}(u) dD_\theta(u) = \int_{\mathbb{R}} x dD_\theta \circ F_X(x)$$

is finite for a given  $D_\theta$ ,  $F_X$  and  $\theta$ .

First we note that if  $X$  is not bounded,  $D_\theta$  cannot put any mass on 1, so  $D_\theta$  must be continuous on  $[0, 1]$ . In this case, we can write

$$(4.1) \quad \rho_\theta(X) = \int_0^1 F_X^{-1}(u) d_\theta(u) du = \int_{\mathbb{R}} x d_\theta(F_X(x)) dF_X(u)$$

where  $d_\theta$  is either the left or right derivative. The second expression may be preferred for computational stability. From this, we see that the tail of  $F_X$ , or equivalently, the growth of  $F_X^{-1}$  and the behavior of  $d_\theta$  near 1 are decisive factors for the finiteness of  $\rho_\theta(X)$ . A simple sufficient condition is given by

$$|F_X^{-1}(u) d_\theta(u)| \leq (1 - u)^{-1+\delta}, \quad \delta > 0.$$

For example, the proportional hazards distortion has a density  $d_\theta(u) = \theta(1-u)^{\theta-1}$ . Since  $0 < \theta \leq 1$ , this put a stringent restriction to the growth of  $F_X^{-1}$  near 1.

We illustrate the situation by taking the generalized Pareto distribution (GPD) as an example for  $F$ .

**Example 4.1** The value of the scale parameter  $\beta$  is irrelevant for our purpose, so we put it equal to 1. The df of this GPD is given by

$$F_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, & \xi \neq 0 \\ 1 - e^{-x}, & \xi = 0 \end{cases}$$

For large  $\xi > 0$ , it is heavy-tailed (the mean is infinite if  $\xi \geq 1$ ). We have

$$F_\xi^{-1}(x) = \frac{1}{\xi} [(1-u)^{-\xi} - 1].$$

Hence direct calculation can be done easily and one concludes that:

- (i) For the PH-distortion, we have  $d_\theta^{\text{PH}}(u) = \theta(1-u)^{\theta-1}$ . This is unbounded and tends to  $+\infty$  relatively quickly as  $u \uparrow 1$ . As a consequence, we need  $\xi < \theta$  for  $\rho_\theta^{\text{PH}}(X)$  to be finite. This means that the permissible range for  $\theta$  is affected by the tail of the underlying distribution; for relatively heavy-tailed distributions, one cannot choose small values such as 0.05 for  $\theta$ . This limits the usefulness of this distortion, which we regard as undesirable.
- (ii) The PO-distortion has the density given by  $d_\theta^{\text{PO}}(u) = \theta[1 - (1-\theta)u]^{-2}$ . This is bounded by  $1/\theta$ , and hence we need only  $\xi < 1$ .
- (iii) The density for the GA-distortion is  $d_\theta^{\text{GA}}(u) = \exp[-\log \theta \Phi^{-1}(u) - (\log \theta)^2/2]$ . We have not found a clear region of  $\theta$  for which  $\rho_\theta^{\text{PH}}(X)$  is finite, but since  $1 - \Phi(x) \sim x^{-1}\phi(x)$ ,  $x > 0$ ,  $\{\exp[\Phi^{-1}(u)]\}^{-\log \theta}$  cannot be the order of some power of  $\log(1-u)$  ( $u \uparrow 1$ ), so there must be a moderate restriction on the range for  $\theta$ .

We complement the numerical comparison of the VaR and expected shortfall in McNeil et al. (2005) (p. 47) by adding the values of the PO-distortion risk measures. The result is summarized in Table 1. Following the above-mentioned book, we use two different models for the loss distribution  $F_X$ : one is a normal distribution with mean 0 and standard deviation 126.5, and the other is a  $t$  distribution with four degrees of freedom, scaled to have the same standard deviation.

We used the last expression in (4.1) for the computation of the PO-distortion risk measure; i.e., we generate 100,000 observations  $X_i$  from  $F_X$  and calculate the mean of  $X_i d_\theta^{\text{PO}}(F_X(X_i))$ . This was repeated 1,000 times and their mean and standard error are calculated.

We did try computing the PH- and GA-distortion risk measures, but the results were too unstable, probably due to the huge variance for small values of  $\theta$ , to be included here. For the  $t$  model, their values are not even finite.

Table 1: Values of risk measures in normal and  $t$  model

$\alpha = \theta$	0.10	0.05	0.025	0.01	0.005
VaR $_{\alpha}$ (normal)	162.1	208.1	247.9	294.3	325.8
VaR $_{\alpha}$ ( $t$ )	137.1	190.7	248.3	335.1	411.8
ES $_{\alpha}$ (normal)	222.0	260.9	295.7	337.2	365.8
ES $_{\alpha}$ ( $t$ )	223.4	286.3	356.7	465.8	563.5
PO $_{\theta}$ (normal)	157.5	199.7	238.5	285.5	318.3
standard error	(1.5)	(2.3)	(3.7)	(6.5)	(9.9)
PO $_{\theta}$ ( $t$ )	159.7	215.7	278.3	373.5	458.3
standard error	(1.7)	(3.2)	(5.5)	(11.4)	(18.5)

A summary of computation goes as follows. The PH-distortion risk measure cannot even be stably estimated with Monte Carlo methods with one million observations. The GA-distortion risk measure may be more stable. For the PO-distortion risk measure, its numerical computation is stable. It seems that its value is smaller than the VaR when the underlying distribution has short tail, and it is between the VaR and the expected shortfall in the case of heavy-tailed distribution. Thus it seems suitable for some use.

## 4.2 Statistical Estimation

Here we assume, as a first step towards statistical estimation, that i.i.d. sample  $X_1, \dots, X_n$  from the df  $F$  are available. A natural estimate of  $\rho_D(X)$  is given by replacing  $F$  with the empirical df  $\mathbb{F}_n$  based on the sample  $X_1, \dots, X_n$ :

$$\hat{\rho}(X) = \int_0^1 \mathbb{F}_n^{-1}(u) dD_{\theta}(u) = \sum_{i=1}^n X_{(i)} \left[ D_{\theta} \left( \frac{i}{n} \right) - D_{\theta} \left( \frac{i-1}{n} \right) \right]$$

This is a very simple form of  $L$ -statistic, and we can apply standard asymptotics as given in Shorack and Wellner (1986) to get their limiting distributions:

**Proposition 4.2** *Assume that*

$$\begin{aligned} |d_{\theta}(u)| &\leq Mu^{-b_1}(1-u)^{-b_2}, & u \in (0, 1) \\ |F^{-1}(u)| &\leq Mu^{-d_1}(1-u)^{-d_2}, & u \in (0, 1) \end{aligned}$$

Put  $a := (b_1 + d_1) \vee (b_2 + d_2)$ . With certain smoothness assumptions, we have the following:

- (i) (SLLN): If  $a < 1$ , then  $\hat{\rho}(X) \rightarrow \rho(X)$ ,  $\mathbf{P}$ -a.s.

(ii) (**CLT**): If  $a < 1/2$ , then  $\sqrt{n}(\widehat{\rho}(X) - \rho(X)) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$ , where

$$\sigma^2 := \int_0^1 \int_0^1 (u \wedge v - uv) d_\theta(u) d_\theta(v) dF^{-1}(u) dF^{-1}(v)$$

**Example 4.3** If  $F$  has bounded support, then we can take  $d_1 = d_2 = 0$ :

- (i) For the PH distortion,  $0 < \theta$  for SLLN;  $1/2 < \theta$  for CLT
- (ii) For the PO distortion,  $0 < \theta \leq 1$  for SLLN and CLT

**Example 4.4** If  $F$  is a normal distribution, then  $d_1$  and  $d_2$  may be taken arbitrarily close to zero. Thus we get the same range for  $\theta$  as in the preceding example.

- (i) For the PH distortion,  $0 < \theta$  for SLLN;  $1/2 < \theta$  for CLT
- (ii) For the PO distortion,  $0 < \theta \leq 1$  for SLLN and CLT

From the above two examples, we see that the PH distortion risk measure is hard to estimate; we can barely consistently estimate it, but for small value of  $\theta$ , which we would be much interested in, the accuracy of the estimate is impossible to assess.

Naturally, the situation is worse when the underlying distribution has much heavier tail:

**Example 4.5** If  $F$  is GPD ( $\beta = 1$ ), then as in Example 4.1, we have

- (i) For the PH distortion, need  $\theta > \xi$  for SLLN;  $\theta > \xi + 1/2$  for CLT
- (ii) For the PO distortion, need  $\xi < 1$  for SLLN;  $\xi < 1/2$  for CLT.

**Remark 4.6** Insurance loss data may very well be incomplete because of deductibles and limits, so it is important to be able to adapt the methods to incomplete data. In this case, we can replace  $\mathbb{F}_n$  by the Kaplan-Meier estimate with left-truncated right-censored data.

## 5 Concluding Remarks

In this paper, under slightly weaker assumptions, we gave simple proofs of two facts; one is that the law invariance implies the Fatou property, and the other is the representation theorem for law invariant and comonotonically additive coherent risk measures. Using them, with some additional conditions regarding comparability to the expected shortfall, we proposed several one-parameter families of distortion risk measures, among which the proportional odds distortion is most promising because of its financial interpretability and numerical stability.

Considering the recent development of integrating financial and insurance risks, the distortion approach should be useful and indispensable tools, and will be applied to many

practical problems in finance and insurance. Further problems includes portfolio optimization, portfolio performance evaluation and capital allocation. Also since we may not have i.i.d. data in practice, sophisticated estimation methods with time series data would be necessary to implement a historical method.

## REFERENCE

- Acerbi, C. (2002): Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance*, **26**, 1505–1518.
- Acerbi, C. (2004): Coherent representation of subjective risk-aversion, in *Risk Measures for the 21st Century*, ed. G. Szegö. Chichester: John Wiley & Sons, 147–207.
- Acerbi, C. and D. Tasche (2002): On the coherence of expected shortfall, *Journal of Banking & Finance*, **26**, 1487–1503.
- Aczél, J. (1966): *Lectures on Functional Equations and Their Applications*. New York-London: Academic Press.
- Arnold, B. C., C. A. Robertson, P. L. Broukett and B.-Y. Shu (1987): Generating ordered families of Lorenz curves by strongly unimodal distributions. *Journal of Business & Economic Statistics*, **5**, 305–308.
- Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1997): Thinking coherently. *RISK Magazine*, **10**, November, 68–71.
- Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999): Coherent measures of risk, *Mathematical Finance*, **9**, 203–228.
- Choquet, G. (1955): Theory of Capacities. *Annales de l'Institut Fourier*, **5**, 131–295.
- Clayton, D. and J. Cuzick (1986): The semi-parametric Pareto model for regression analysis of survival times, in *Papers on Semiparametric Models at the ISI Centenary Session*, eds. R. D. Gill and M. N. Voors. Amsterdam: Centre for mathematics and Computer Science, 19–30.
- Cox, D. R. (1972): Regression models and life table. *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- Dabrowska, D. M., K. A. Doksum and R. Miura (1989): Rank estimates in a class of semiparametric two-sample models. *Annals of the Institute of Statistical Mathematics*, **41**, 63–79.

- Delbaen, F. (2002): Coherent risk measures on general probability spaces, in *Advances in Finance and Stochastics*, eds. K. Sandmann and P. J. Schönbucher. Berlin: Springer-Verlag, 1–37.
- Denneberg, D. (1990): Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, **20**, 181–190.
- Denneberg, D. (1994): *Non-additive measures and integral*. Dordrecht-Boston-London: Kluwer Academic Publishers.
- Dhaene, J., M. Denuit, M. J. Goovaerts, R. Kaas and D. Vyncke (2002): The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**, 3–33.
- Ferguson, T. S. (1967): *Mathematical Statistics*. New York: Academic Press.
- Föllmer, H. and A. Schied (2002a): Robust preferences and convex measures of risk, in *Advances in Finance and Stochastics*, eds. K. Sandmann and P. J. Schönbucher. Berlin: Springer-Verlag, 39–56.
- Föllmer, H. and A. Schied (2002b): Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**, 429–447.
- Föllmer, H. and A. Schied (2002c): *Stochastic Finance: An Introduction in Discrete Time*. Berlin: Walter de Gruyter.
- Grabisch, M., T. Murofushi and M. Sugeno (Eds.) (2000): *Fuzzy Measures and Integrals: Theory and Applications*. Heidelberg: Physica-Verlag.
- Hardy, G. H., J. E. Littlewood and G. Pólya (1952): *Inequalities*. 2nd ed. Cambridge: Cambridge University Press.
- Jouini, E., W. Schachermayer and N. Touzi (2005): *Law Invariant Risk Measures have the Fatou Property*, Preprint.
- Kleiber, C. and S. Kotz (2003): *Statistical Size Distributions in Economics and Actuarial Sciences*. New Jersey: John Wiley & Sons.
- Kusuoka, S. (2001): On law invariant coherent risk measures. *Advances in Mathematical Economics*, **3**, 83–95.
- Lehmann, E. L. (1953): The power of rank tests. *Annals of Mathematical Statistics*, **24**, 23–43.
- Lehmann, E. L. (1966): Some concepts of dependence. *Annals of Mathematical Statistics*, **37**, 1137–1153.

- Leitner, J. (2005): A short note on second-order stochastic dominance preserving coherent risk measures, *Mathematical Finance*, **15**, 649–651.
- McCullagh, P. (1980): Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, **42**, 109–142.
- McNeil, A. J., R. Frey and P. Embrechts (2005): *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton, New Jersey: Princeton University Press.
- Miura, R. and H. Tsukahara (1993): One-sample estimation for generalized Lehmann's alternative models. *Statistica Sinica*, **3**, 83–101.
- Müller, A. and D. Stoyan (2002): *Comparison Methods for Stochastic Models and Risks*. Chichester: John Wiley & Sons.
- Schmeidler, D. (1989): Subjective probability and expected utility without additivity. *Econometrica*, **57**, 571–587.
- Shorack, G. R. and J. A. Wellner (1986): *Empirical Processes with Applications to Statistics*. New York: John Wiley & Sons.
- Tasche, D. (2002): Expected shortfall and beyond. *Journal of Banking & Finance*, **26**, 1519–1533.
- Wang, S. S. (1995): Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, **17**, 43–54.
- Wang, S. S. (1996): Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, **26**, 71–92.
- Wang, S. S. (2000): A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, **67**, 15–36.
- Wang, S. S., V. R. Young and H. H. Panjer (1997): Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**, 173–183.
- Yaari, M. E. (1987): The dual theory of choice under risk. *Econometrica*, **55**, 95–115.



# Extreme quantile estimation using extreme value theory

Tatsuyoshi Okimoto\*

January 2007

## Abstract

This paper considers extreme quantile estimation based on extreme value theory (EVT). In particular, we consider the case where the distribution function  $F$  belongs to the maximum domain of attraction of either the Gumbel distribution with infinite right endpoint or the Fréchet distribution. In this case, EVT suggests two extreme quantile estimators depending on the domain of attraction of the distribution. We derive the asymptotic distribution of these estimators for  $p_n \rightarrow 1$  and  $n(1 - p_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . It turns out that those estimators use only the first  $k$  largest observations from the samples and the optimal choice of  $k$  is also treated in both asymptotic and finite sample cases.

## 1 Introduction

The study of extreme events, such as currency crises, market crashes, and natural disasters, is essential for risk control and policy management. For example, the probability and magnitude of natural disasters are integral to calculating the premium on property insurance.

Since we cannot collect many observations of extreme events, developing theories to describe extreme events has been one of the most attractive topics in mathematics and statistics, therefore it is not in itself new. Extreme value theory (EVT) can give us a powerful tool to examine distribution theories for extreme events. As a consequence, EVT has recently attracted increasing attention among theorists, empirical researchers and practitioners.

Following this theme, this paper considers the extreme quantile estimation based on EVT.<sup>1</sup> In particular, we consider the case where the distribution function  $F$  belongs to the maximum domain of attraction of the Gumbel distribution  $\Lambda$  with infinite right endpoint or Fréchet distribution  $\Phi_\alpha$ . In these case, extreme value theory suggests the following two large quantile estimators depending on the domain of attraction of the distribution as discussed in detail in section 4:

---

\*International Graduate School of Social Sciences, Yokohama National University, 79-4 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501 Japan. E-mail : okimoto@ynu.ac.jp

<sup>1</sup>Since we can treat small quantiles as large quantiles by considering  $Y = -X$ , we will concentrate on large quantiles without loss of generality.

**Case 1:**  $F \in D(\Lambda)$

$$\hat{x}_{p_n} = X_{k+1,n} - \left( \frac{1}{k} \sum_{j=1}^k X_{j,n} - X_{k+1,n} \right) \ln \frac{n(1-p_n)}{k}.$$

**Case 2:**  $F \in D(\Lambda)$  with  $x_F = \infty$  or  $D(\Phi_\alpha)$

$$\hat{x}_{p_n}^* = \left\{ \frac{n(1-p_n)}{k} \right\}^{-\left\{ \frac{1}{k} \sum_{j=1}^k \ln(1 \vee X_{j,n}) - \ln(1 \vee X_{k+1,n}) \right\}} (1 \vee X_{k+1,n}),$$

where  $X_{k,n}$  is the  $k$ th largest order statistic of  $X_1, \dots, X_n$ ,  $x_F$  is the right endpoint of  $F$  defined by  $x_F = \sup\{x | F(x) < 1\}$ , and  $k$  is a certain number less than  $n$ .

There are a number of studies that consider statistical properties of related estimators. Weissman (1978) first considers the joint limiting distribution of the  $k$  largest observations from random variables in the domain of attraction of  $\Lambda$ . He derives large quantile estimators as asymptotic MLE and asymptotic UMVUE, both of which are asymptotically equivalent to  $\hat{x}_{p_n}$ . The approach is based on the fact that the joint distribution of the  $k$  largest order statistics converges to the joint distribution of the  $k$  largest order statistics from a location and scale exponential distribution. Therefore, when the tails of the distribution are not exactly exponential, the estimators are biased. When large  $k$  is used, this bias becomes severer, but the variance of the estimator gets smaller. To control this bias and variance some attention must be given to the number of order statistics  $k$  that is used. Boos (1984) investigates the finite sample properties of Weissman's estimators via simulation. His findings include

1. If a distribution has a lighter tail compared to exponential tails such as the normal and Weibull distributions, then Weissman's estimators using  $k = 4n(1-p)$  can outperform the empirical quantiles when  $50 \leq n \leq 500, p \geq 0.95$  and  $500 < n \leq 5000, p \geq 0.99$ .
2. If a distribution has an approximately exponential tail such as  $t_8, \chi_1^2, \chi_4^2, \chi_8^2$ , and lognormal  $(0, 1/2)$ , then Weissman's estimators using  $k = 0.2n(1-p)$  can outperform the empirical quantiles when  $50 \leq n \leq 500, p \geq 0.95$  and Weissman's estimators using  $k = 0.1n(1-p)$  can outperform the empirical quantiles when  $500 < n \leq 5000, p \geq 0.95$ .
3. If a distribution has a heavier tail compared to exponential tails like  $t_3$  and standard lognormal, then it is better to use empirical quantiles.

Davis and Resnick (1984) adapt a similar idea as that of Weissman's (1978) to estimate the tail of the distribution consistently for  $p_n \rightarrow 1$  and  $n(1-p_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Dekkers, Einmahl and de Haan (1989) consider more general situation  $F \in D(H_\alpha)$ , where  $H_\alpha$  is the generalized extreme

value distribution and show that the empirical quantile, which is a special case of  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$ , is asymptotically normal for  $p_n \rightarrow 1$  and  $n(1 - p_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . They also consider the situation  $p_n \rightarrow 1$  and  $n(1 - p_n) \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$ . They propose the following estimator to estimate large quantile:

$$\hat{x}_{p_n}^{DEH}(\hat{\alpha}) = \frac{\left(\frac{k}{n(1-p)}\right)^{\hat{\alpha}_n} - 1}{\hat{\alpha}_n} \cdot \hat{a}_{\frac{n}{k}}^{DEH} + \hat{b}_{\frac{n}{k}}$$

with

$$\begin{aligned} \hat{a}_{\frac{n}{k}}^{DEH} &= \frac{X_{k+1,n} M_n^{(1)}}{\rho_1(\hat{\alpha}_n)}, \\ \hat{b}_{\frac{n}{k}} &= X_{k+1,n}, \\ M_n^{(r)} &= \frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k+1,n})^r, \\ \rho_1(\alpha) &= \begin{cases} 1 & \alpha \geq 0, \\ \frac{1}{1-\alpha} & \alpha < 0, \end{cases} \end{aligned}$$

and  $\hat{\alpha}_n$  is some consistent estimator of  $\alpha$ . They derive the asymptotic distribution of  $\hat{x}_{p_n}^{DEH}(\hat{\alpha})$  for  $p_n \rightarrow 1$  and  $n(1 - p_n) \rightarrow c \in (0, \infty)$  as  $n \rightarrow \infty$  (see also Dekkers and de Haan (1989)). Note that for the special case  $\alpha = 0$  it turns out that  $\hat{x}_{p_n}^{DEH}(\hat{\alpha})$  with different estimator of norming constant for  $a_{\frac{n}{k}}$  coincides with  $\hat{x}_{p_n}$ . Following Dekkers et al. (1989), de Haan and Rootzén (1993) and de Haan (1994) show asymptotic normality of  $\hat{x}_{p_n}^{DEH}(\hat{\alpha})$  for  $p_n \rightarrow 1$  and  $n(1 - p_n) \rightarrow 0$  as  $n \rightarrow \infty$ .

The first contribution of this paper is that we derive asymptotic distribution of  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$  for  $p_n \rightarrow 1$  and  $n(1 - p_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . This result can be considered as the generalization of the result of asymptotic distribution of empirical quantile by Dekkers et al. (1989). This result also provides the base of the asymptotic optimal choice of  $k$ , the number of the largest order statistics used to compute  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$ .

The finite sample optimal choice of  $k$  is also considered in this paper. Although Boos (1984) gives some rules of thumb to choose  $k$ , there are few suggestions in the previous literature which concentrates on the extreme quantile estimation. Pickands (1975) considers the estimation problem of the tail index and proposes using the smallest value of  $k$  that minimizes the Kolmogorov-Smirnov distance between the empirical distribution of the normalized spacings  $d_1, \dots, d_k$  and the exponential distribution with mean some estimates of  $a_{\frac{n}{k}}$ . Zelterman (1993) remarks the Weissman's (1978) result and the test proposed by Gail and Gastwirth (1978) to choose  $k$  to use bootstrap technique for simulating extreme order statistics. Zelterman proposes a sequential test to the normalized spacings  $d_1, \dots, d_k$  until some value of  $k$  is found to reject the hypothesis that the  $d_1, \dots, d_k$  are iid

exponential random variables. Gail and Gastwirth (1978) show that under the null hypothesis that the  $d_1, \dots, d_k$  are iid exponential random variables, the Gini statistic

$$G_k = \frac{\sum_{i=1}^{k-1} i(k-i)(d_{i,k} - d_{i+1,k})}{(k-1) \sum_{i=1}^k d_i},$$

where  $d_{i,k}$  is the  $i$ th largest order statistic of  $d_1, \dots, d_k$ , is approximately normal with mean  $\frac{1}{2}$  and variance  $\frac{1}{12(k-1)}$  when  $k$  is greater than 5. Zelterman (1993) suggests using  $G_k$  in a sequential test procedure examining successive values of  $k$  until the null hypothesis is rejected. Wang (1995) proposes two criteria based on generalized least squares and the Shapiro-Wilk goodness-of-fit statistics to select the  $k$  largest order statistics for the domain of attraction of the Gumbel distribution. Other related studies that consider the optimal choice of  $k$  for the Hill estimator include Beirlant, Vynckier and Teugels (1996a, 1996b), and Danielsson, de Haan, Peng, and de Vries (2001). The second contribution of this paper is that we propose a way to apply Zelterman's (1993) bootstrap method to choose  $k$  and confirm that the method works fairly well via simulation.

The remainder of the paper is organized as the following. Section 2 provides several examples where we need a large quantile estimation and reviews two typical methods to estimate large quantiles. Section 3 presents the basic EVT. Section 4 focuses on deriving two large quantile estimators using extreme value theory. Section 5 derives the asymptotic distributions of those estimators and asymptotic optimal choice of  $k$ , which is the number of the largest order statistics to compute those estimators, is provided. Section 6 considers optimal choice of  $k$  in finite sample using bootstrap. Lastly, Section 7 provides concluding remarks.

## 2 Motivating Examples and Previous Methods

In this section, we present several examples where a large quantile estimation is needed to reach the solution. We then introduce some quantile estimation methods used in literature.

Here are several examples where a large quantile estimation is needed:

### Example 1 Value-at-Risk (VaR)

The  $100(1 - \alpha)\%$  VaR for period  $T$  of a certain portfolio is defined as the maximum loss of the portfolio with probability greater than  $\alpha$  for period  $T$ . Therefore, calculating VaR is equivalent to estimating the  $1 - \alpha$  quantile of the portfolio's loss distribution. Generally, since small values, such as 0.05 or 0.01, are chosen as  $\alpha$ , calculating VaR is the same as the large quantile estimation.  $\square$

### Example 2 Constructing a bank

Suppose a government wants to construct a river bank so that the probability of a flood happening in 100 years is less than 0.01. If the annual maximum flow of the river is independently and identically distributed, then since  $0.99^{1/100} \approx 0.9999$ , this problem is equivalent to estimating 0.9999 quantile of the annual maximum flow distribution, so that the government can construct a bank to endure that quantity of flow.  $\square$

### Example 3 Defining Crises

There are a variety of ways to define a crisis. One of the clearest ways to define a crises is to use a phenomenon which occurs with probability less than some small probability. For example, we can define a currency crisis if a currency depreciates more than 0.99 quantile of the currency depreciation distribution. As we can see this example, large quantile estimation can be useful also to define crises.  $\square$

### Example 4 Optimal Forecasts

Consider the situation where we want to forecast  $y_{t+h}$  from some information set  $\Omega_t$  with a loss function

$$L(e) = \{p + (1 - 2p)I(e < 0)\} |e|^q, \quad 0 < p < 1, \quad q > 0$$

where  $e = y_{t+h} - \hat{y}_{t+h}$  and  $I(\cdot)$  is the indicator function. Then optimal forecast  $y_{t+h}^*$  is

$$y_{t+h}^* = \begin{cases} E(y_{t+h}|\Omega_t) & \text{if } q = 2, p = 1/2, \text{ (MSE case)} \\ F_{y_{t+h}|\Omega_t}^{-1}(p) & \text{if } q = 1, \text{ (Lin Lin case)}. \end{cases}$$

Hence if  $q = 1$  and  $p$  is close to one, in other words, if people care about underestimation very much, but don't care about overestimation very much, the optimal forecast is a large quantile. Such a situation indeed occurs, for example, if policy makers want to estimate sustainable fiscal deficit, they tend to care about only underestimation.  $\square$

As we can see from the above examples, large quantile estimation is very useful for risk control and they are one of the hot topics in finance and insurance field.

Since quantile estimation is one of the most important themes in statistics, there are many studies about it and two of the major basic methods are the following.

### Parametric method

Parametric method estimates quantiles assuming the distribution  $F(x; \theta)$  which data follows. Thus parametric quantile estimator is defined as  $\hat{x}_p = F_{\hat{\theta}}^{-1}(p)$ . Since we assume the distribution, estimating quantile is identical to estimating the parameters  $\theta$  of distribution. If the assumption about the

distribution is correct, the parametric method can bring relatively accurate quantile estimates, even if it is large quantile. In practice, however, it is very rare that we would know what distribution the data follows.  $\square$

### Non-parametric method: Empirical quantile

Empirical quantile is defined as

$$\hat{x}_p = X_{n-k+1,n}, \quad p \in \left( \frac{k-1}{n}, \frac{k}{n} \right], \quad \text{where } X_{k,n} \text{ is the } k\text{th largest order statistics of } X_1, \dots, X_n.$$

Since empirical quantile does not need any assumptions about the distribution, it is very useful in practice. In addition, it is known that with fixed  $p$  empirical quantile is consistent and asymptotically normally distributed. However it is very hard to estimate extreme quantiles using empirical quantile, since order statistics often have little information regarding extreme quantiles.  $\square$

Although the above two methods are very useful in certain situation, neither of them give fully satisfactory estimates of large quantiles. Since they are two extreme methods, thus parametric and nonparametric methods, it is expected that there exist some median, or semiparametric methods, which requires weak assumptions (only about tails), but still have some good properties for estimating large quantiles. One possibility of that is the method using the extreme value theory.

## 3 Basic Theory

EVT studies behavior of extreme values such as maximum and values exceeding a certain threshold and was developed in astronomy and meteorology. Recently it is drawing the attention in finance and insurance in terms of risk control and analysis of the fat-tail distribution.

In the following we introduce the basic concept of EVT and then state the detail of extreme value distributions, their maximum domain of attraction, and norming constants based on Embrechts, Klüppelberg, and Mikosh (1997).

Throughout the paper we consider the following model:

$$X_1, X_2, \dots, X_n, \dots \text{ iid } X_i \sim F.$$

Thus  $\{X_i\}_{i=1}^{\infty}$  is a sequence of iid non-degenerate random variables with a common df  $F$ . Also define  $M_n$  as

$$M_n = \max(X_1, X_2, \dots, X_n).$$

EVT considers the situations where there exist some sequences  $a_n > 0$ ,  $b_n \in \mathbb{R}$  ( $n = 1, 2, \dots$ ) and a non-degenerate distribution  $H$  such that:

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H, \quad n \rightarrow \infty.$$

If there exist those sequences and a distribution, we call  $(a_n, b_n)$  as norming constants and  $H$  as a extreme value distribution. Also we say that  $F$  belongs to the maximum domain of attraction of  $H$  and denote  $F \in D(H)$ .

It is, actually, known that there exist only three extreme value distributions. The following theorem was discovered first by Fisher and Tippett (1928) and later proved in complete generality by Gnedenko (1943).

**Theorem 1** Fisher-Tippett Theorem, Limit Laws for Maxima

Extreme value distributions belong to the type of one of the following three distributions:

$$\begin{aligned} \Phi_\alpha(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} & \alpha > 0, & \text{Fréchet distribution} \\ \Psi_\alpha(x) &= \begin{cases} \exp(-(-x)^\alpha) & x \leq 0 \\ 1 & x > 0 \end{cases} & \alpha > 0, & \text{Weibull distribution} \\ \Lambda(x) &= \exp(-e^{-x}) & x \in \mathbb{R} & \text{Gumbel distribution} \end{aligned}$$

□

It is known that the necessary and sufficient condition for that  $F$  belongs to the maximum domain of attraction of each extreme value distribution and its norming constants are given by the next theorem.

**Theorem 2**

The maximum domain of attractions for each extreme value distribution are characterized as the following:

$$\begin{aligned} F \in D(\Phi_\alpha) &\iff F(x) < 1, \forall x \quad \text{and} \quad \lim_{x \rightarrow \infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = t^{-\alpha}, \quad t > 0 \\ F \in D(\Psi_\alpha) &\iff x_F < \infty \quad \text{and} \quad \lim_{x \uparrow x_F} \frac{\overline{F}(x_F - (x_F - x)t)}{\overline{F}(x)} = t^\alpha, \quad t > 0 \\ F \in D(\Lambda) &\iff \exists s(\cdot) > 0 \quad \text{s.t.} \quad \lim_{x \uparrow x_F} \frac{\overline{F}(x + ts(x))}{\overline{F}(x)} = e^{-t} \end{aligned}$$

Also, norming constants for each extreme value distribution are characterized as the following:<sup>2</sup>

$$F \in D(\Phi_\alpha), F(a_n) = 1 - \frac{1}{n}, b_n = 0$$

$$F \in D(\Psi_\alpha), F(x_F - a_n) = 1 - \frac{1}{n}, b_n = x_F$$

$$F \in D(\Lambda), F(b_n) = 1 - \frac{1}{n}, a_n = s(b_n)$$

□

### Remark

It is known that the function  $s(\cdot)$ , which is used for defining the maximum domain of attraction of the Gumbel distribution, can be chosen as

$$s(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)} dt, \quad x < x_F.$$

□

As you can see from above, according to the extreme value theory, we can classify the whole set of distributions into three families given they have an extreme value distribution. Consider now the characteristics of each family. The first family, which belongs to the maximum domain of attraction of the Weibull distribution,  $D(\Psi_\alpha)$ , consists of distributions with support bounded to the right, and hence they have a finite tail. The second family, which belongs to the maximum domain of attraction of the Gumbel distribution,  $D(\Lambda)$ , consists of a large variety of distributions with very different tail characteristics. In particular, tails may range from moderately heavy (lognormal, and heavy tailed Weibull) to very light (exponential, and distributions with support bounded to the right). Also, note that  $\forall F \in D(\Lambda), E(X^+)^\delta < \infty, \forall \delta$ , where  $X^+ = \max(X, 0)$ . The third family, which belongs to the maximum domain of the attraction of the Fréchet distribution,  $D(\Phi_\alpha)$ , has an infinite right endpoint and are heavy-tailed, in the sense that  $E(X^+)^\delta = \infty, \forall \delta > \alpha$ .

## 4 Large Quantile Estimation

In this section assuming that  $F$  belongs to the maximum domain of attraction of the Gumbel distribution, we derive a large quantile estimator of  $F$  using the relationship given by the extreme value theory. In addition, we show that the method can be extended to the case where  $F$  belongs to  $D(\Phi_\alpha)$  or  $D(\Lambda)$  with  $x_F = \infty$  instead. In the previous section, we stated that we can divide whole set of distribution into three families, although, in practice those two families are most important, since those maximum domain of attraction contain almost all distributions used in practice.

<sup>2</sup>It is, actually, enough for norming constants to satisfy the equation asymptotically. Thus, norming constants are not uniquely decided and those sequences which satisfy the equation asymptotically could be all norming constants.



#### 4.1 Case 1: $F \in D(\Lambda)$

It is known that if  $F \in D(\Lambda)$ ,

$$\lim_{n \rightarrow \infty} n\bar{F}(a_n x + b_n) = -\ln H(x) = \exp(-x), \quad \forall x$$

must be hold. This implies

$$\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \lim_{n \rightarrow \infty} \exp\left(\frac{b_n - u_n}{a_n}\right), \quad \forall x,$$

where  $u_n = a_n x + b_n$ . By taking logarithm of both sides we have

$$\lim_{n \rightarrow \infty} \left\{ \ln n\bar{F}(u_n) - \frac{b_n - u_n}{a_n} \right\} = 0, \quad \forall x$$

Hence<sup>3</sup>

$$u_n \sim b_n - a_n \ln n\bar{F}(u_n), \quad \forall x. \tag{1}$$

A handwaving consequence of this is that for large  $u_n = a_n x + b_n = x_p$ ,<sup>4</sup>

$$x_p \sim b_n - a_n \ln n(1-p).$$

Therefore a large quantile estimator could take on the form

$$\hat{x}_p = \hat{b}_n - \hat{a}_n \ln n(1-p).$$

Consider now the estimation of  $a_n$  and  $b_n$ . By Theorem 2, the natural estimators of  $a_n$  and  $b_n$  are

$$\begin{aligned} \hat{b}_n &= \hat{F}^{-1}\left(1 - \frac{1}{n}\right) = X_{2,n} \\ \hat{a}_n &= s(\hat{b}_n) = \int_{\hat{b}_n}^{X_{1,n}} \frac{\bar{\hat{F}}(y)}{\bar{\hat{F}}(\hat{b}_n)} dy = n \int_{X_{2,n}}^{X_{1,n}} \bar{\hat{F}}(y) dy = X_{1,n} - X_{2,n} \end{aligned}$$

There are, however, some doubts about finite-sample properties of those estimators. Since  $\hat{b}_n$  is the  $\left(1 - \frac{1}{n}\right)$  quantile estimator, in practice the accuracy of estimation is, in general, unsatisfactory.<sup>5</sup>

Furthermore since  $\hat{a}_n$  is a function of  $\hat{b}_n$ , its accuracy might be even worse than  $\hat{b}_n$ . One reason for this is that  $\hat{a}_n$  and  $\hat{b}_n$  depends only on the first two largest value. Taking into consideration that the extreme value distributions are the limit distributions of a normalized maxima, it is very natural.

We can consider, however, that the  $k$  largest order statistics also contain some information about

<sup>3</sup>Note that accuracy of this approximation depends of the value of  $x$ .

<sup>4</sup>Since (1) holds for all  $x$ ,  $u_n$  could take any value. However, since for all  $x$ ,  $u_n$  diverges to  $\infty$ , it is natural to consider  $x_p$  as a large quantile.

<sup>5</sup>Actually this is our motivation to consider large quantile estimators using extreme value theory.

the tail of  $F$ . Taking this into account, we can introduce the variable  $k = k_n$  ( $1 \leq k \leq n-1$ ) which depends on  $n$  satisfying

$$\begin{cases} k \rightarrow \infty, & n \rightarrow \infty, \\ \frac{k}{n} \rightarrow 0, & n \rightarrow \infty. \end{cases}$$

With the following relationship

$$x_p \sim b_{\frac{n}{k}} - a_{\frac{n}{k}} \ln \frac{n(1-p)}{k}, \quad \forall x, \quad (2)$$

each estimator then changes into<sup>6</sup>

$$\begin{aligned} \hat{b}_{\frac{n}{k}} &= \hat{F}^{-1} \left( 1 - \frac{k}{n} \right) = X_{k+1,n} \\ \hat{a}_{\frac{n}{k}} &= s(\hat{b}_{\frac{n}{k}}) = \int_{\hat{b}_{\frac{n}{k}}}^{X_{1,n}} \frac{\overline{\hat{F}}(y)}{\overline{\hat{F}}(\hat{b}_{\frac{n}{k}})} dy = \frac{n}{k} \int_{X_{k+1,n}}^{X_{1,n}} \overline{\hat{F}}(y) dy \\ &= \frac{n}{k} \left\{ \frac{k}{n} (X_{k,n} - X_{k+1,n}) + \dots + \frac{1}{n} (X_{1,n} - X_{2,n}) \right\} \\ &= \frac{1}{k} \sum_{j=1}^k X_{j,n} - X_{k+1,n} \\ \hat{x}_p &= \hat{b}_{\frac{n}{k}} - \hat{a}_{\frac{n}{k}} \ln \frac{n(1-p)}{k} \\ &= X_{k+1,n} - \left( \frac{1}{k} \sum_{j=1}^k X_{j,n} - X_{k+1,n} \right) \ln \frac{n(1-p)}{k} \end{aligned}$$

#### 4.2 Case 2: $F \in D(\Phi_\alpha)$ or $F \in D(\Lambda)$ with $x_F = \infty$

In order to extend the method stated in the previous subsection to the case where  $F \in D(\Phi_\alpha)$  or  $F \in D(\Lambda)$  with  $x_F = \infty$ , we will use the next result.

##### **Theorem 3**

Let  $X_1, \dots, X_n$  be iid with df  $F \in D(\Phi_\alpha)$ , or  $F \in D(\Lambda)$  with  $x_F = \infty$ . Then  $X_1^*, \dots, X_n^*$  where  $X_i^* = \ln(1 \vee X_i)$  are iid with df  $F^* \in D(\Lambda)$  with  $x_{F^*} = \infty$ .  $\square$

Suppose  $F \in D(\Phi_\alpha)$  or  $F \in D(\Lambda)$  with  $x_F = \infty$ , then  $F^* \in D(\Lambda)$  from Theorem 4.1, for sufficiently large  $n$  and  $x_p > 1$ ,

$$\ln x_p = x_p^* \sim b_{\frac{n}{k}}^* - a_{\frac{n}{k}}^* \ln \frac{n(1-p)}{k}, \quad (3)$$

<sup>6</sup>  $\frac{n}{k}$  is not necessarily an integer. So rigorously speaking we have to use  $[\frac{n}{k}]$  instead of  $\frac{n}{k}$ , but for the simplicity of notation we denote  $[\frac{n}{k}]$  as  $\frac{n}{k}$ .

so that a large quantile estimator could take on the form

$$\begin{aligned}\hat{x}_p &= \left\{ \frac{n(1-p)}{k} \right\}^{-\hat{a}_{\frac{n}{k}}} \exp \left( \hat{b}_{\frac{n}{k}} \right) \\ &= \left\{ \frac{n(1-p)}{k} \right\}^{-\left\{ \frac{1}{k} \sum_{j=1}^k \ln(1 \vee X_{j,n}) - \ln(1 \vee X_{k+1,n}) \right\}} (1 \vee X_{k+1,n}).\end{aligned}$$

## 5 Asymptotic distribution and optimal choice of $k$

In this section, we derive the asymptotic distributions of the estimators  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$ , and consider asymptotic optimal choice of  $k$ .

### 5.1 Case 1 : $F \in D(\Lambda)$ with $x_F = \infty$

In this subsection, we consider the asymptotic distribution of the estimators  $\hat{x}_{p_n}$ . To derive the asymptotic distribution for the estimator  $\hat{x}_{p_n}$  we refer the following theorem from Davis and Resnick (1984).

#### **Theorem 4** Davis and Resnick (1984)

Suppose  $F \in D(\Lambda)$  with  $x_F = \infty$ . Suppose further  $\{k\}$  satisfies

$$\begin{cases} \lim_{n \rightarrow \infty} k = \infty \\ \lim_{n \rightarrow \infty} \frac{k}{n} = 0 \end{cases} \quad (4)$$

Then

$$\left( \sqrt{k} \left( \frac{\hat{a}_{\frac{n}{k}}}{a_{\frac{n}{k}}} - 1 \right), \sqrt{k} \left( \frac{\hat{b}_{\frac{n}{k}} - b_{\frac{n}{k}}}{a_{\frac{n}{k}}} \right) \right)' \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_2)$$

where  $a_{\frac{n}{k}} = \frac{n}{k} \int_{b_{\frac{n}{k}}}^{\infty} 1 - F(y) dy = \frac{n}{k} \int_{b_{\frac{n}{k}}}^{\infty} \bar{F}(y) dy$  and  $b_{\frac{n}{k}} = F^{-1} \left( 1 - \frac{k}{n} \right)$ .  $\square$

From theorem 4, we can easily show the following.

#### **Theorem 5**

Suppose  $F \in D(\Lambda)$  with  $x_F = \infty$ . Suppose further  $\{p_n\}$  satisfies

$$\begin{cases} \lim_{n \rightarrow \infty} p_n = 1 \\ \lim_{n \rightarrow \infty} n(1-p_n) = \infty \end{cases} \quad (5)$$

Moreover, suppose

$$\lim_{n \rightarrow \infty} \frac{\sqrt{k}}{a_{\frac{n}{k}}} \left\{ x_{p_n} - \left( b_{\frac{n}{k}} - a_{\frac{n}{k}} \ln \frac{n(1-p_n)}{k} \right) \right\} = 0 \quad (6)$$

is satisfied. If we choose  $\{k\}$  so that it satisfies  $\lim_{n \rightarrow \infty} \frac{n(1-p_n)}{k} = c$ , then

$$\sqrt{k} \left( \frac{\hat{x}_{p_n} - x_{p_n}}{a_{\frac{n}{k}}} \right) \xrightarrow{d} N(0, 1 + (\ln c)^2)$$

□

**Remark** Asymptotic optimal choice of  $k$

It follows from theorem 5 that the asymptotic variance is minimized when  $c = 1$ . Therefore asymptotic optimal choice of  $k$  can be expressed as

$$k = n(1-p_n) + c_n \tag{7}$$

where  $c_n = o(n(1-p_n))$ . In this case, it is obvious that

$$c = \lim_{n \rightarrow \infty} \frac{n(1-p_n)}{k} = \lim_{n \rightarrow \infty} \frac{n(1-p_n)}{n(1-p_n) + c_n} = \lim_{n \rightarrow \infty} \frac{1}{1 + \frac{c_n}{n(1-p_n)}} = 1$$

□

According to theorem 5, we can find the condition (6) is crucial for the asymptotic normality of the estimator  $\hat{x}_{p_n}$ . There are two important cases where the condition (6) are satisfied, which are exponential distribution and empirical quantile. Let's consider the case of exponential distribution first. If  $F$  is exponential distribution,

$$x_{p_n} = -\ln(1-p_n), \quad a_{\frac{n}{k}} = 1, \quad b_{\frac{n}{k}} = \ln \frac{n}{k}$$

Hence

$$x_{p_n} - b_{\frac{n}{k}} + a_{\frac{n}{k}} \ln \frac{n(1-p_n)}{k} = -\ln(1-p_n) - \ln \frac{n}{k} + \ln \frac{n(1-p_n)}{k} = 0, \quad \forall k.$$

Thus for exponential distribution, (6) holds for all  $k$ , meaning theorem 5 also holds. Next let's consider the case of empirical quantile. If we choose  $\{k\}$  so that  $k = n(1-p_n)$  holds, then  $\hat{x}_{p_n}$  coincides with empirical quantile and

$$x_{p_n} - b_{\frac{n}{k}} + a_{\frac{n}{k}} \ln \frac{n(1-p_n)}{k} = F^{-1}(p_n) - F^{-1}(p_n) + a_{\frac{n}{k}} \ln 1 = 0, \quad \forall k.$$

Hence we can apply theorem 5 to the empirical quantile.

The empirical quantile has another important property. As we mentioned above, asymptotic optimal choice of  $k$  can be expressed as (7), which empirical quantile satisfies. Thus empirical quantile is asymptotically efficient. There is, unfortunately, no guarantee that the estimators  $\hat{x}_{p_n}$  with asymptotic optimal choice of  $k$  satisfies the condition (6) in general. However we can show the

following theorem.

**Theorem 6**

Suppose  $F \in D(\Lambda)$  with  $x_F = \infty$ . Suppose further  $\{p_n\}$  satisfies (5). If we choose  $\{k\}$  so that satisfies (7) with  $c_n = O\left(\sqrt{n(1-p_n)}\right)$ , then

$$\sqrt{k} \left( \frac{\hat{x}_{p_n} - x_{p_n}}{a_{\frac{n}{k}}^*} \right) \xrightarrow{d} N(0, 1).$$

□

Thus this theorem indicates that  $\hat{x}_{p_n}$  with certain choice of  $k$ , which includes the empirical quantile as a special case, is asymptotic normal and optimal.

**5.2 Case 2 :  $F \in D(\Phi_\alpha)$  or  $F \in D(\Lambda)$  with  $x_F = \infty$**

In this subsection we consider the asymptotic distribution of the estimator  $\hat{x}_{p_n}^*$ .

By theorem 3, the exponential transformed estimator  $\hat{x}_{p_n}^*$  has the same properties as the estimator  $\hat{x}_{p_n}$ , we can easily show

**Theorem 7**

Suppose  $F \in D(\Phi_\alpha)$  or  $F \in D(\Lambda)$  with  $x_F = \infty$ . Suppose further  $\{p_n\}$  satisfies (5). If we choose  $\{k\}$  so that it satisfies (7), then

$$\frac{\sqrt{k}}{a_{\frac{n}{k}}^*} \log \left( \frac{\hat{x}_{p_n}^*}{x_{p_n}} \right) \xrightarrow{d} N(0, 1)$$

□

## 6 Optimal choice of $k$ in finite sample

In the previous chapter we found the asymptotic optimal choice of  $k$ . It, however, does not provide the way how to choose  $k$  in finite sample or in practice. In this section, we consider the optimal choice of  $k$  for  $\hat{x}_p$  in finite sample. More precisely we consider the use of bootstrap to choose optimal  $k$  automatically so that minimize the sample MSE.

One possible attractive way to choose  $k$  optimally is to use resampling methods, such as the bootstrap and the subsampling method. Both methods attempt to approximate finite sample distributions of some estimators using the empirical distribution of the estimators from resampling pseudo samples. Although these methods still need a large sample size, it is expected that the resulting approximation would be better than the one obtained by using asymptotic theory, because

they approximate not the asymptotic distribution, but finite sample distributions directly. Invented by Efron (1979), bootstrap is well known for its validity and higher order efficiency, provided that the finite distribution converges to a limit distribution with a certain uniformity. As for extreme order statistics, however, Bickel and Freedman (1981) actually use the extreme order statistics as one of the counter examples where the usual bootstrap would not work well, because of a certain lack of uniformity. Let  $X_1, \dots, X_n$  iid,  $X_i \sim U(0, \theta)$ . Then by extreme value theory,

$$\frac{n(X_{1,n} - \theta)}{\theta} \xrightarrow{d} \Psi_1.$$

Hence it is natural to bootstrap  $\frac{n(X_{1,n} - \theta)}{\theta}$  by  $\frac{n(X_{1,n}^* - X_{1,n})}{X_{1,n}}$ , where  $X_1^*, \dots, X_n^*$  is a bootstrap pseudo sample. Bickel and Freedman (1981), however, point out that this bootstrap does not work, because with probability 1, the conditional distribution of  $\frac{n(X_{1,n}^* - X_{1,n})}{X_{1,n}}$  does not have a weak limit, since  $\limsup n(X_{1,n}^* - X_{k,n}) = \infty$ , and  $\liminf n(X_{1,n}^* - X_{k,n}) = 0$  with probability 1 for each  $k$ . One intuitive reason why the usual bootstrap does not work well is that it is impossible to obtain pseudo samples with values greater than  $X_{1,n}$ .

In order to overcome this problem Zelterman (1993) considers a semiparametric bootstrap technique using EVT. His basic idea is based on one of Weissman's (1978) results below.

**Lemma 1** Weissman (1978)

Suppose  $F \in D(\Lambda)$ . For fixed  $k$ , as  $n \rightarrow \infty$ , the normalized spacings  $ia_n^{-1}(X_{i,n} - X_{i+1,n})$ ,  $i = 1, \dots, k$  are asymptotically jointly distributed as independent standard exponential random variables.  $\square$

According to this result spacings are asymptotically iid, so we can expect the bootstrap to work. In fact Zelterman (1993) shows that we can get a valid pseudo sample for the extreme order statistics using the following bootstrap:

1. Calculate the normalized spacings or the difference of order statistics

$$d_i = i(X_{i,n} - X_{i+1,n}), \quad i = 1, \dots, n-1.$$

2. Choose  $k^*$ .
3. Bootstrap using  $d_1, \dots, d_{k^*}$  to get  $d_1^*, \dots, d_{k^*}^*$ .
4. Construct a pseudo sample for the first  $k^*$  largest order statistics by

$$X_{j,n}^* = X_{k^*+1,n} + \sum_{i=j}^{k^*} \frac{d_i^*}{i}, \quad j = 1, \dots, k^*.$$

Using this Zelterman's bootstrap we might be able to choose  $k$  as the following sample MSE criterion:

$$\hat{k} = \operatorname{argmin} \left\{ k : \frac{1}{B} \sum_{i=1}^B [\hat{x}_p^{*,i}(k) - \tilde{x}_p]^2 \right\},$$

where  $\hat{x}_p^{*,i}(k)$  is the estimate of  $\hat{x}_p$  with  $k$  from the  $i$ th pseudo bootstrap sample and  $\tilde{x}_p$  is some consistent estimator. Although the Zelterman's bootstrap works for the extreme order statistic, his method is not applicable to our problems without modification. One reason for that is to conduct the Zelterman's bootstrap we also have to choose  $k^*$ , which indicates the number of the largest order statistics used for the bootstrap. We, however, might choose  $k$  and  $k^*$  simultaneously. In other words, we can use same number for  $k$  and  $k^*$ , but this simultaneous choice might introduce downward bias as we discussed below. Another problem is the choice of a consistent estimator  $\tilde{x}_p$ . To approximate the MSE we need a consistent estimator, which is usually given as the original estimate  $\hat{x}_p(k)$  or sample average of the estimates from the bootstrap samples  $\tilde{x}_p^*(k) = \frac{1}{B} \sum_{i=1}^B \hat{x}_p^{*,i}(k)$ . For our purpose, however, those consistent estimators might not work, because they induce downward bias again. To remedy those problems we propose the following modifications:

- Parametric bootstrap
- Bootstrap using the first  $k + 1$  largest observation, i.e.  $k^* = k + 1$
- Set  $n(1 - p) \leq k \leq n/3$
- Use empirical quantile as a consistent estimator.

One defect of the Zelterman's semiparametric bootstrap is that we are not able to get enough variation, if we choose small  $k$ . As a consequence we tend to choose small  $k$  if we use the bootstrap sample MSE criterion. According to Weissman's results, normalized spacings are asymptotically not only iid, but also standard exponential distributed. The parametric bootstrap, therefore, is also available and it might mitigate that problem. The second modification, namely using the  $k + 1$  largest observation might also mitigate the problem because it would provide more variation without much distortion. The third modification is to set minimum of  $k$  as  $\min\{1, n(1 - p)\}$  to prevent from choosing too small  $k$ . Also we set maximum of  $k$  as  $n/3$ , since theory does not hold with too large  $k$ . The last modification is about for a consistent estimator. Since we know the empirical quantiles are good estimators, we might use them as a consistent estimator. As a result we can consider the following six different types of bootstrap methods:

1. Semiparametric bootstrap with  $\tilde{x}_p = \hat{x}_p(k)$  and  $k^* = k + 1$

2. Semiparametric bootstrap with  $\tilde{x}_p = \tilde{x}_p^*(k)$  and  $k^* = k + 1$
3. Semiparametric bootstrap with  $\tilde{x}_p =$  empirical quantile and  $k^* = k + 1$
4. Parametric bootstrap with  $\tilde{x}_p = \hat{x}_p(k)$  and  $k^* = k + 1$
5. Parametric bootstrap with  $\tilde{x}_p = \tilde{x}_p^*(k)$  and  $k^* = k + 1$
6. Parametric bootstrap with  $\tilde{x}_p =$  empirical quantile and  $k^* = k + 1$

In order to investigate the validity of proposed six different bootstrap methods, we conduct the following simulations:

1. Generate iid samples with size 100 from normal and lognormal distributions and calculate estimates using empirical quantile and estimator  $\hat{x}_{p_n}$  with choosing  $k$  via six different bootstrap methods for  $p = .90, .95, .975, .99, .995, .999$ .
2. Repeat step 1, 1,000 times.
3. Calculate the ratio of sample MSE between empirical quantile and each estimator.

The calculated sample MSEs for normal distribution are shown in table 1 and for lognormal distribution are shown in table 2. Note that if the MSE ratio is less than 1, the estimator outperforms empirical quantile. According to the simulation results, all bootstrap methods seem to work very well, in particular for normal distribution. We need, however, more sophisticated simulations and theoretical justification to conclude these points and to obtain clearer results.

## 7 Concluding remarks

In this paper we consider the use of EVT in extreme quantile estimation, that uses an iid data sequence  $\{X_1, \dots, X_n\}$ . More precisely, section 3 introduced the basic EVT, while section 4 reviewed the derivation of the estimators  $\hat{x}_p$  and  $\hat{x}_p^*$  under the assumption that  $\{X_1, \dots, X_n\}$  are independently and identically distributed with  $F$  which belongs to the maximum domain of attraction of the Gumbel or the Fréchet distribution.

In section 5, we derived asymptotic distribution of the estimators  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$ , which also provided the asymptotic optimal choice of  $k$ . Our results showed that if  $F$  belongs to the domain of attraction of the Gumbel distribution with the infinite right endpoint and  $\{p_n\}$  satisfies (5), then the estimator  $\hat{x}_{p_n}$  with  $\{k\}$  so that satisfies (7) is asymptotically efficient and normally distributed. The corresponding theorem for the estimator  $\hat{x}_{p_n}^*$  was also provided. Following section 5, we further



investigated the optimal choice of  $k$  in finite sample using bootstrap in section 6. The results of simulations indicated that the proposed methods work well, but further studies and theoretical justification are required.

This paper opens up many possible topics for future investigation on the large quantile estimators  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$ , which are derived by using EVT. One of which is to investigate the asymptotic and finite-sample properties of the estimators  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$  more generally. Since we use the simplest estimators to estimate the norming constants  $a_{\frac{n}{k}}$  and  $b_{\frac{n}{k}}$ , derivation of ways to improve those estimators is another future research subject. Those methods developed by Csörgo, Deheuvels and Mason (1985) and Csörgo and Horváth (1987) are possibilities for that. Also since we assume that the data sequences are all iid, the effects of relaxing this assumption can be another possibility for future studies as well. Among others, Dress (2001) considers this problem for  $\hat{x}_{p_n}^{DEH}$  by allowing time dependence, while Chernozhukov (2005) develops regression quantile theory for  $\hat{x}_{p_n}^{DEH}$ .

Finally, in the last part of the paper we summarize the method for large quantile estimation using EVT. This method is a semiparametric method and can be used to estimate large quantiles that are usually difficult to estimate. There are a lot of possible ways to improve the estimators  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$ . For instance, through an optimal choice of  $k$  and/or improving the norming constants estimators. From those observations we conclude that the estimators  $\hat{x}_{p_n}$  and  $\hat{x}_{p_n}^*$  can be powerful tools for estimating large quantiles. Further researches are expected to be conducted in the near future.

## Appendix: Proof of Theorems

### Proof of Theorem 5

$$\begin{aligned} \sqrt{k} \left( \frac{\hat{x}_{p_n} - x_{p_n}}{a_{\frac{n}{k}}} \right) &= \ln \frac{n(1-p_n)}{k} \cdot \sqrt{k} \left( \frac{\hat{a}_{\frac{n}{k}}}{a_{\frac{n}{k}}} - 1 \right) + \sqrt{k} \left( \frac{\hat{b}_{\frac{n}{k}} - b_{\frac{n}{k}}}{a_{\frac{n}{k}}} \right) \\ &\quad + \frac{\sqrt{k}}{a_{\frac{n}{k}}} \left\{ x_{p_n} - \left( b_{\frac{n}{k}} - a_{\frac{n}{k}} \ln \frac{n(1-p_n)}{k} \right) \right\} \end{aligned}$$

Since it is obvious that  $\{k\}$  chosen as  $\lim_{n \rightarrow \infty} \frac{n(1-p_n)}{k} = c$  satisfies (4), by Theorem 4 and (6), we can get

$$\begin{aligned} \sqrt{k} \left( \frac{\hat{x}_{p_n} - x_{p_n}}{a_{\frac{n}{k}}} \right) &\xrightarrow{d} \ln c \cdot Z_1 + Z_2, \quad \text{where } (Z_1, Z_2)' \sim N(\mathbf{0}, \mathbf{I}_2). \\ &\sim N(0, 1 + (\ln c)^2) \end{aligned}$$

□

**Proof of Theorem 6**

It is enough to show (6) is satisfied. Define

$$f(t) = \frac{\bar{F}(b_t)}{F'(b_t)}, \quad a(t) = t \int_{b(t)}^{\infty} \bar{F}(s) ds,$$

then from Davis and Resnick (1984)

$$\frac{f(ts)}{a(t)} \rightarrow 1, \quad \text{uniformly on a compact neighborhood of } s = 1 \text{ as } t \rightarrow \infty. \quad (8)$$

We also have

$$\begin{aligned} b'_t &= \frac{\bar{F}(b_t)}{F'(b_t)t} \quad \left( \because \bar{F}(b_t) = \frac{1}{t}, \quad b'_t = \frac{1}{F'(b(t))t^2} \right) \\ &= \frac{f(t)}{t} \end{aligned} \quad (9)$$

Therefore,

$$\begin{aligned} \frac{\sqrt{k}}{a_{\frac{n}{k}}} (x_{p_n} - b_{\frac{n}{k}}) &= \frac{\sqrt{k}}{a_{\frac{n}{k}}} (b_{\frac{n}{k-c_n}} - b_{\frac{n}{k}}) \quad \left( \because p_n = 1 - \frac{k-c_n}{n} \right) \\ &= \sqrt{k} \int_{\frac{n}{k}}^{\frac{n}{k-c_n}} \frac{f(t)}{a_{\frac{n}{k}} \cdot t} dt \quad (\because (9)) \\ &= \sqrt{k} \int_1^{\frac{k}{k-c_n}} \frac{f(\frac{n}{k}s)}{a_{\frac{n}{k}} \cdot s} ds \\ &= \sqrt{k} \int_1^{\frac{k}{k-c_n}} \left\{ \frac{1}{s} + o(1) \right\} ds \\ &\quad \left( \because \lim_{n \rightarrow \infty} \frac{k}{k-c_n} = 1, \quad \lim_{n \rightarrow \infty} \frac{n}{k} = \infty, (8) \right) \\ &= \sqrt{k} \ln \frac{k}{k-c_n} \{1 + o(1)\} \\ &= -\sqrt{k} \ln \frac{k-c_n}{k} \{1 + o(1)\}. \end{aligned}$$

Note that by the Taylor theorem,

$$\begin{aligned} \sqrt{k} \ln \frac{k-c_n}{k} &= \sqrt{k} \ln \left( 1 - \frac{c_n}{k} \right) \\ &= \sqrt{k} \left( -\frac{c_n}{k} \right) + \sqrt{k} \cdot o\left( \frac{c_n}{k} \right) \\ &= O(1) + o(1) \quad \left( \because c_n = O\left( \sqrt{n(1-p_n)} \right), \quad k = n(1-p_n) + c_n \right) \\ &= O(1). \end{aligned}$$

Hence

$$\frac{\sqrt{k}}{a_{\frac{n}{k}}} (x_{p_n} - b_{\frac{n}{k}}) = -\sqrt{k} \ln \frac{k-c_n}{k} + o(1).$$

It follows that

$$\begin{aligned}
& \frac{\sqrt{k}}{a_{\frac{n}{k}}} \left\{ x_{p_n} - \left( b_{\frac{n}{k}} - a_{\frac{n}{k}} \ln \frac{n}{k} (1 - p_n) \right) \right\} \\
&= \frac{\sqrt{k}}{a_{\frac{n}{k}}} \left( x_{p_n} - b_{\frac{n}{k}} \right) + \sqrt{k} \ln \frac{n(1 - p_n)}{k} \\
&= -\sqrt{k} \ln \frac{k - c_n}{k} + o(1) + \sqrt{k} \ln \frac{k - c_n}{k} \\
&= o(1),
\end{aligned}$$

which completes the proof.  $\square$

### Proof of Theorem 7

It is the immediate consequence from theorem 6.  $\square$

## References

- [1] Beirlant, J., J.T. Teugels and P. Vynckier (1996a). Excess functions and estimation of the extreme-value index, *Bernoulli* **2**(4), 293–318.
- [2] Beirlant, J., J.L. Teugels and P. Vynckier (1996b). Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Amer. Statist. Assoc.* **91**, 1659–1667.
- [3] Bickel, P. J., and D.A. Freedman (1981). Some asymptotics for the bootstrap, *Ann. Statist.* **9**, 1196–1217.
- [4] Boos, D.D. (1984). Using extreme value theory to estimate large percentiles, *Technometrics* **26**, 33–39.
- [5] Chernozhukov, V. (2005) Extremal quantile regression, *Ann. Statist.* **33**, 806–839.
- [6] Csörgo, S., P. Deheuvels and D. Mason (1985). Kernel estimates of the tail index of a distribution, *Ann. Statist.* **13**, 1050–1077.
- [7] Csörgo, M. and L. Horváth (1987). On the optimality of estimating the tail index and a naive estimator, *Austral. J. Statist.* **29**(2), 166–178.
- [8] Danielsson, J. L. de Haan, L. Peng, and C.G. de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation, *J. Multivariate Analysis* **76**, 226–248.
- [9] Davis, R.A. and S.I. Resnick (1984). Tail estimates motivated by extreme value theory, *Ann. Statist.* **12**, 1467–1487.

- [10] Dekkers, A.L.M. and L. de Haan (1989). On the estimation of the extreme-value index and large quantile estimation, *Ann. Statist.* **17**, 1795–1832.
- [11] Dekkers, A.L.M., J.H.J. Einmahl and L. de Haan (1989). On the estimation of the extreme-value index and large quantile estimation, *Ann. Statist.* **17**, 1833–1855.
- [12] Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance, *Bernoulli* **9**, 617–657.
- [13] Efron, B. (1979). Bootstrap method: another look at the Jackknife, *Ann. Statist.* **7**, 1–26.
- [14] Embrechts, P., C. Klüppelberg and T. Mikosch (1997). *Modelling Extremal Events*, Springer, Verlag.
- [15] Fisher, R.A. and L.H.C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. Cambridge Phil. Soc.* **24**, 180–190.
- [16] Gail, M.H. and J.L. Gastwirth (1978). A scale-free goodness-of-fit test for the exponential distribution based of the Gini statistic, *J. Royal Statistical Society Ser. B*, **40**, 350–357.
- [17] Gnedenko, Par B. (1943). Sur la distribution limite du terme mximum d'une série aléatoire, *Ann. Math.* **44**, 423–453.
- [18] Hann, L. de (1994). Extreme value statistics, In: Galambos et al. (Eds.) *Extreme Value Theory and Applications*, 93–122.
- [19] Hann, L. de and H. Rootzén (1993). On the setimation of high quantiles, *J. Statistical Planning and Inference*, **35**, 1–13.
- [20] Pickands, III, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.* **3**, 119–131.
- [21] Wang, J.Z. (1995). Selection of the  $k$  largest order statistics for the domain of attraction of the Gumbel distribution, *J. Amer. Statist. Assoc.* **90**, no. 431, 1055–1061.
- [22] Weissman, I. (1978). Estimation of parameters and large quantiles based on the  $k$  largest observations, *J. Amer. Statist. Assoc.* **73**, 812–815.
- [23] Zelterman, D. (1993). A semiparametric bootstrap technique for simulating extreme order statistics, *J. Amer. Statist. Assoc.* **88**, no. 422, 477–485.

Table 1: Simulation results for normal distribution with  $n=100$ ,  $B=500$

quantile	0.9	0.95	0.975	0.99	0.995	0.999
bs1	<b>0.870</b>	<b>0.703</b>	<b>0.801</b>	<b>0.722</b>	<b>0.835</b>	<b>0.837</b>
bs2	<b>0.954</b>	<b>0.920</b>	<b>0.961</b>	<b>0.911</b>	<b>0.871</b>	<b>0.752</b>
bs3	<b>0.918</b>	<b>0.829</b>	<b>0.854</b>	<b>0.874</b>	<b>0.970</b>	<b>0.825</b>
bs4	<b>0.827</b>	<b>0.751</b>	<b>0.904</b>	<b>0.776</b>	<b>0.879</b>	<b>0.775</b>
bs5	<b>0.977</b>	<b>0.909</b>	<b>0.928</b>	<b>0.829</b>	<b>0.833</b>	<b>0.758</b>
bs6	<b>0.857</b>	<b>0.761</b>	<b>0.883</b>	<b>0.784</b>	<b>0.942</b>	<b>0.831</b>

Table 2: Simulation results for lognormal distribution with  $n=100$ ,  $B=500$

quantile	0.9	0.95	0.975	0.99	0.995	0.999
bs1	1.081	1.287	<b>0.751</b>	<b>0.800</b>	<b>0.299</b>	<b>0.695</b>
bs2	<b>0.849</b>	<b>0.975</b>	<b>0.716</b>	<b>0.805</b>	<b>0.310</b>	<b>0.731</b>
bs3	1.069	1.194	<b>0.804</b>	<b>0.938</b>	<b>0.391</b>	<b>0.703</b>
bs4	1.230	1.283	<b>0.778</b>	<b>0.780</b>	<b>0.291</b>	<b>0.746</b>
bs5	1.314	1.208	<b>0.715</b>	<b>0.816</b>	<b>0.305</b>	<b>0.782</b>
bs6	1.159	1.241	<b>0.818</b>	<b>0.897</b>	<b>0.446</b>	<b>0.994</b>

# 経験尤度法による極値パラメータの推定

東京大学経済学研究科 D1 大和田 孝

2006年4月25日

## 概要

金融データの分析では、近年、fat-tail な現象をモデル化する必要が一段と高まってきている。極値論 (Extreme Value Theory (EVT)) は、分布の裾の確率的挙動を記述する強力な分析手法である。極値論の利用に当たっては、極値パラメータを正確に推定しなければならないが、Hill(1975) や Pickands(1975) の推定法では、推定に使う観測値数の選び方や推定量のバイアスなど無視できない問題が多い。これを踏まえ本稿では、経験尤度法 (Empirical Likelihood) に基づく推定手法を提案する。推定の結果、真の極値パラメータが 0 に近い正の値であるときには、経験尤度法に基づく手法は、これまでの手法より優れたパフォーマンスを示すことが分かった。さらに、TOPIX の日次対数収益率を用いた実証分析では、分布の fat-tail 性を正確に捉えることに成功したほか、パーセンタイルも正確に推定できた。

## 1 Introduction

近年、金融リスク管理の現場では、突発的かつ大規模な現象をどのようにモデル化するかが重要な課題となっている。例えば、保険商品を価格付けする際、商品価格は災害・事故などの各種の rare events のモデル化の方法に強く依存する。また、rare events の適切なモデル化に成功すれば、リスク管理の現場で通常用いられる Value at risk (VaR) もより有効な指標になるかもしれない。こうした中、極値論 (Extreme Value Theory (EVT)) は、fat-tail な経済現象の分析に役立つと言われており、本稿では rare events のモデル化への極値論の適用を検討する。極値論では、未知パラメータをどのように推定するかがポイントとなるが、本稿では新たな方法として経験尤度法 (Empirical Likelihood) に基づく推定量を構成する。その後、シミュレーション実験を通して本稿の手法と従来の手法のパフォーマンス比較を行い、最後に TOPIX の日次対数収益率を用いた実証分析を行う。

## 2 極値論 (Extreme Value Theory (EVT))

極値論では、以下の2つのデータが主要な分析対象となる。

### 1. 最大値の系列

例 (投資家の) 月ごとの最大損失額、年間最大降水量など。

### 2. ある閾値 $u$ を超過した観測値

例

$(X_1, X_2, X_3, X_4, X_5) = (1, 5, 6, 2, 3)$  とする。このとき、閾値として  $u = 4$  にとると、 $(X_2, X_3)$  が「超過値」として分析対象となる。

### 2.1 $M_n = \max\{X_1, X_2, \dots, X_n\}$ の確率的性質

最初に上記 1. のケースを説明する。 $X_1, X_2, \dots, X_n$  を共通の分布関数  $F$  をもつ独立な確率変数とする。

$$X_1, X_2, \dots, X_n, i.i.d \sim F$$

ここで、 $(X_1, \dots, X_n)$  の最大値を  $M_n$  と定義する。すなわち、 $M_n = \max\{X_1, \dots, X_n\}$  である。次の定理は、適当な数列  $a_n > 0, b_n \in \mathbf{R}$  で基準化された最大値  $M_n$  の  $n \rightarrow \infty$  における極限分布を与える。詳しい証明については、Embrechts et al.(1997) を参照されたい。この定理の結果は、「最大値  $M_n$  に関する中心極限定理」との解釈も可能である。

#### **Theorem 2.1.1 (Fisher-Tippett theorem)**

$X_1, X_2, \dots, X_n$  を独立同一な確率変数とし、共通の分布関数を  $F$  とおく。このとき、基準化数列  $a_n > 0, b_n \in \mathbf{R}$  とある分布関数  $G$  が存在して、

$$\begin{aligned} P\left(\frac{M_n - b_n}{a_n} \leq z\right) &= P(X_1 \leq a_n z + b_n, \dots, X_n \leq a_n z + b_n) \\ &= \{F(a_n z + b_n)\}^n \longrightarrow G(z), \quad \text{as } n \rightarrow \infty \text{ for all } z: \mathbf{R} \text{ 上の連続点} \end{aligned}$$

が成立すれば、極限分布  $G$  は以下の3つのいずれかに等しい。すなわち、元の分布  $F$  が

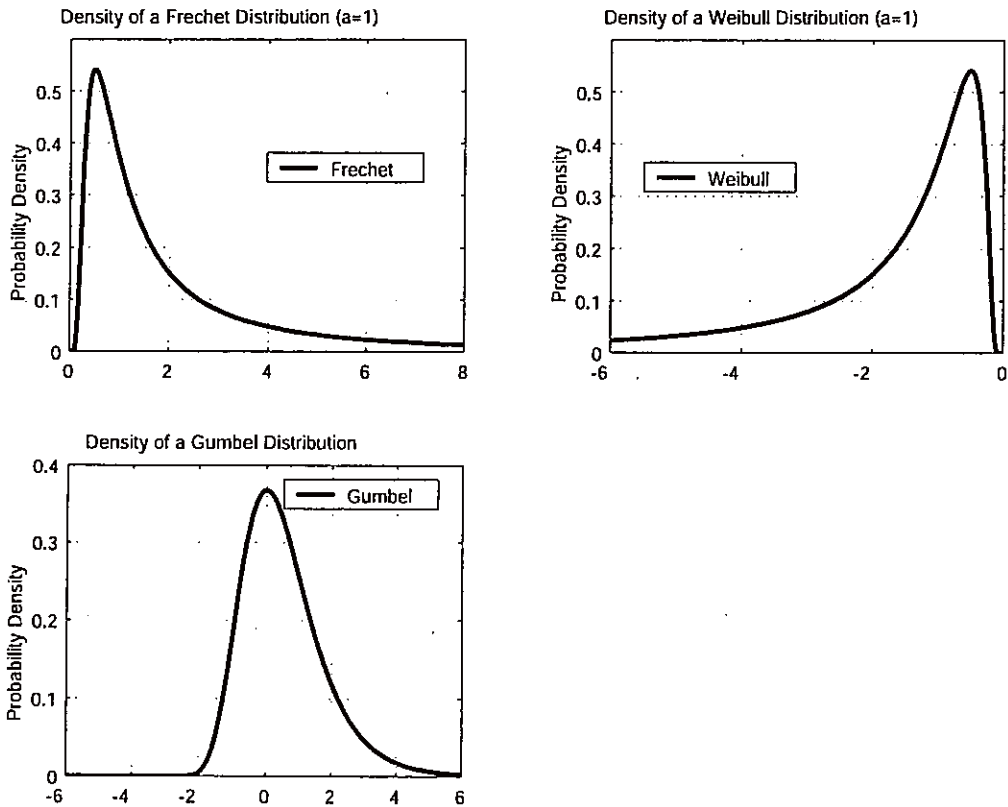
何であれ、 $M_n$  の極限分布は Fréchet、Weibull、Gumbel のいずれかの分布に一致する。

$$\text{Fréchet : } \Phi_\alpha(z) = \begin{cases} 0, & z \leq 0 \\ \exp\{-z^{-\alpha}\}, & z > 0, \alpha > 0 \end{cases} \quad (1)$$

$$\text{Weibull : } \Psi_\alpha(z) = \begin{cases} \exp\{-(-z)^\alpha\}, & z \leq 0 \\ 1, & z > 0, \alpha > 0 \end{cases} \quad (2)$$

$$\text{Gumbel : } \Lambda(z) = \exp\{-e^{-z}\}, \quad z \in \mathbf{R} \quad (3)$$

□



なお、上記の3つの分布は一本の式で表現可能である。この分布クラスは、一般極値分布 (generalized extreme value distribution) と呼ばれ、集合  $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$  上で定義される。



$$H_{\xi;\mu,\sigma}(z) = \begin{cases} \exp \left[ - \left( 1 + \xi \frac{z - \mu}{\sigma} \right)^{-1/\xi} \right], & \text{if } \xi \neq 0 \\ \exp \left[ - \exp \left( - \frac{z - \mu}{\sigma} \right) \right], & \text{if } \xi = 0 \end{cases} \quad (4)$$

但し、各パラメータは  $\xi \in \mathbf{R}$ ,  $\mu \in \mathbf{R}$ ,  $\sigma > 0$  を満たしており、 $\mu$  と  $\sigma$  はそれぞれ位置パラメータ、スケールパラメータと呼ばれる。また、 $\xi = 0$  のときの分布関数は  $H_{\xi;\mu,\sigma}$ ,  $\xi \neq 0$  の  $\xi \rightarrow 0$  における極限分布とも解釈できる。一般極値分布と Fréchet, Weibull, Gumbel 分布の対応関係は以下の通り。

- $\xi > 0$  のときは、 $\xi = 1/\alpha$ ,  $\mu = 1$ ,  $\sigma = 1/\alpha$  とおくと、 $H_{\xi;\mu,\sigma} = \Phi_\alpha$ : Fréchet 分布が得られる。
- $\xi < 0$  のときは、 $\xi = -1/\alpha$ ,  $\mu = -1$ ,  $\sigma = 1/\alpha$  とおくと、 $H_{\xi;\mu,\sigma} = \Psi_\alpha$ : Weibull 分布が得られる。
- $\xi = 0$  のときは、 $\mu = 0$ ,  $\sigma = 1$  とおくと、 $H_{\xi;\mu,\sigma} = \Lambda$ : Gumbel 分布が得られる。

### Example

#### Example 2.1.2 (一様分布)

$X_1, X_2, \dots$  を一様分布  $U(0, 1)$  に従う独立な確率変数とする。このとき、 $F(x) = x$ ,  $0 \leq x \leq 1$  であり、 $z < 0$  に対して  $n > -z$  にとって、 $a_n = 1/n$ ,  $b_n = 1$  とおけば、

$$\begin{aligned} P \left( \frac{M_n - b_n}{a_n} \leq z \right) &= \{F(n^{-1}z + 1)\}^n \\ &= \left( 1 + \frac{z}{n} \right)^n \\ &\rightarrow e^z, \quad \text{as } n \rightarrow \infty \end{aligned}$$

がいえる。 $a_n, b_n$  で基準化された最大値  $M_n$  は、 $\xi = -1$  なる Weibull 分布に分布収束する。

□

#### Example 2.1.3 (指数分布)

$X_1, X_2, \dots$  を標準指数分布  $Exp(1)$  に従う独立な確率変数とすれば、 $F(x) = 1 - e^{-x}$ ,  $x > 0$  がいえる。このとき、 $a_n = 1$ ,  $b_n = \log n$  とれば、

$$\begin{aligned} P\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \{F(z + \log n)\}^n \\ &= [1 - \exp\{-(z + \log n)\}]^n \\ &= [1 - n^{-1}e^{-z}]^n \\ &\rightarrow \exp\{-e^{-z}\}, \quad \text{as } n \rightarrow \infty \end{aligned}$$

である。極限分布は Gumbel 分布となり、極値パラメータ  $\xi$  は 0 となる。

□

#### Example 2.1.4 (パレート分布)

$X_1, X_2, \dots$  を独立同一に  $F(x) = 1 - (\kappa/(\kappa + x))^\alpha$ ,  $x > 0$ ,  $\alpha, \kappa > 0$  なるパレート分布に従う確率変数とする。  $a_n = \kappa n^{1/\alpha}$ ,  $b_n = 0$  とすると、 $z > 0$  に対して

$$\begin{aligned} P\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \{F(\kappa n^{1/\alpha} z)\}^n \\ &= \left\{1 - \left(\frac{1}{1 + n^{1/\alpha} z}\right)^\alpha\right\}^n \\ &\rightarrow \exp\{-z^{-\alpha}\}, \quad \text{as } n \rightarrow \infty \end{aligned}$$

が成り立つ。従って、極限分布は  $\xi = 1/\alpha$  の Fréchet 分布である。

□

#### Example 2.1.5 (コーシー分布)

$X_1, X_2, \dots$  を独立かつ、 $f(x) = (\pi(1 + x^2))^{-1}$ ,  $x \in \mathbf{R}$  なる密度関数のコーシー分布に従う確率変数とし、 $a_n = n/\pi$ ,  $b_n = 0$  に選べば、 $z > 0$  に対して、

$$\begin{aligned}
P\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \{F(nz/\pi)\}^n \\
&= \left\{ \int_{-\infty}^{nz/\pi} \frac{dx}{\pi(1+x^2)} \right\}^n \\
&= \left\{ \frac{1}{\pi} \left( \tan^{-1} \frac{n}{\pi} z + \frac{\pi}{2} \right) \right\}^n \\
&\rightarrow \exp\{-z^{-1}\}, \quad \text{as } n \rightarrow \infty
\end{aligned}$$

が成立し、最大値  $M_n$  は  $\xi = 1$  の Fréchet 分布に分布収束することが分かる。

□

$\xi$  の大小は、元の分布  $F$  の裾の厚さと密接な関係がある。すなわち、 $F$  がコーシー分布やパレート分布のような fat-tail な分布であるとき、 $\xi$  は大きな値となることが多い (Fréchet 分布に対応)。一方、右端  $x_F = \sup\{x \in \mathbf{R} : F(x) < 1\}$  が有限値をとる分布では、 $\xi$  は負の値となり、極限分布は Weibull 分布となる。Gumbel 分布 ( $\xi = 0$ ) については、分布の裾が厚いとも薄いともいえない、いわば中間のケースに該当する。

実際の分析では、データセットを適当に分割して各ブロックから最大値を抽出してから、最大値の系列に一般極値分布を当てはめればよいことが Theorem 2.1.1 から示唆される。応用上は、データセットを 1 年ごと、もしくはひと月ごとに区切ることが多い。一般極値分布のパラメータ推定問題に関しては、これまでに多くの手法が提案されており、最尤推定量 (Maximum Likelihood estimator (MLE)) (Prescott and Walden(1980))、the method of probability-weighted moments(Hosking et al.(1983)) などが一例として挙げられる。。

## 2.2 閾値 $u$ を超過した確率変数の性質

「最大値」のほかに、「閾値  $u$  を超過した観測値」も極値論では重要な分析対象となる。超過値については、「資産運用していて損失額がある基準額を超えたとき、その損失額はどのような分布に従うか」、「超過損失額の平均値はどのくらいか」など様々な問題設定での分析が考えられる。また、最大値の分析では、1つのブロックに 2 個以上の大きな観測値が存在する場合、最も大きな観測値以外は全て分析対象から捨てられてしまうため、十分大きな観測値でも除外され得るとの意味で無駄の多いアプローチともいえる。このた

め、超過値の分析では、最大値の分析と比較して少なくとも実証上はより有効な推定量が得られる可能性がある。まず、超過値を数学的に扱うため「超過分布関数」を定義する。

**Definition 2.2.1 超過分布関数 (Excess Distribution Function)**

分布関数  $F$  をもつ確率変数  $X$  を考える。このとき、 $u < x_F = \sup\{x \in \mathbf{R} : F(x) < 1\}$  に対して、 $X$  の超過分布関数を次のように定義する。

$$F_u(x) = P(X \leq x + u \mid X > u), \quad x \geq 0 \quad (5)$$

**Remark**

条件付確率の定義から、

$$F_u(x) = \frac{F(x+u) - F(u)}{1 - F(u)} \quad (6)$$

がいえる。

□

実証上は、分析者にとって元の分布  $F$  は未知であり、従って  $F_u(x)$  も未知となるため、極値パラメータの推定にあたっては、 $F_u(x)$  を一般パレート分布で近似することとなる。

**Theorem 2.2.2**

$X_1, X_2, \dots, X_n$  を共通の分布関数  $F$  をもつ独立な確率変数とする。 $a_n > 0, b_n \in \mathbf{R}$  が存在し、 $F$  が Theorem 2.1.1 の意味で極限分布として  $H_{\xi;\mu,\sigma}, \xi \in \mathbf{R}, \mu \in \mathbf{R}, \sigma > 0$  をもつと仮定する。

$$\{F(a_n z + b_n)\}^n \rightarrow H_{\xi;\mu,\sigma} \quad \text{as } n \rightarrow \infty$$

このとき、 $\beta(u) > 0$  が存在して以下が成り立つ。

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0 \quad (7)$$

ここで、 $x_F$  は分布の右端 (i.e.  $x_F = \sup\{x \in \mathbf{R} : F(x) < 1\}$ ) であり、 $G_{\xi,\beta}(x)$  は

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{if } \xi = 0 \end{cases} \quad (8)$$

$$x \in D(\xi, \beta) = \begin{cases} [0, \infty) & \text{if } \xi \geq 0, \\ [0, -\beta/\xi] & \text{if } \xi < 0. \end{cases}$$

と定義される。 $G_{\xi, \beta}(x)$  は一般パレート分布 (generalized Pareto distribution (GPD)) と呼ばれる。なお、 $G_{0, \beta}$  については、 $\xi \rightarrow 0$  における  $G_{\xi, \beta}$ ,  $\xi \neq 0$  の極限とも解釈できる。

□

### **Remark**

上記の定理は、十分大きな  $u$  に対して  $F_u(x)$  が一般パレート分布にて近似可能なことを示している。なお、定理の詳しい証明は Embrechts et al.(1997) を参照。関係式 (9) は、次のセクションで  $\xi$  を推定する際に用いられる。

$$F_u(x) \approx G_{\xi, \beta(u)}(x), \quad \text{for all } 0 < x < x_F - u \quad (9)$$

### **Example**

#### **Example 2.2.3 (一様分布)**

$F(x) = x$ ,  $0 \leq x \leq 1$  なる一様分布  $U(0, 1)$  については、任意の閾値  $u > 0$  と  $0 \leq x \leq 1 - u$  に対して、

$$\begin{aligned} F_u(x) &= \frac{F(x+u) - F(u)}{1 - F(u)} \\ &= \frac{x+u-u}{1-u} \\ &= \frac{x}{1-u} = G_{-1, 1-u}(x) \end{aligned}$$

が成立する。これは、 $\xi = -1$ 、 $\beta(u) = 1 - u$  なる一般パレート分布に対応している。

□

#### **Example 2.2.4 (指数分布)**

$F$  が標準指数分布  $Exp(1)$  に従うものとする。このとき、 $F(x) = 1 - e^{-x}$ ,  $x > 0$  であるため、 $F_u(x)$  は任意の閾値  $u > 0$  に対して、厳密に  $\xi = 0$ ,  $\beta = 1$  なる一般パレート分布に従う。

$$\begin{aligned} F_u(x) &= \frac{F(x+u) - F(u)}{1 - F(u)} \\ &= \frac{1 - e^{-(x+u)} - (1 - e^{-u})}{1 - (1 - e^{-u})} \\ &= 1 - e^{-x} = G_{0,1}(x) \end{aligned}$$

□

以下では、関係式 (9) を扱いやすい形に変形してから、 $p$ -パーセンタイルの推定値  $\hat{x}_p$  を定式化する。最初に、(6) を (9) に代入すると、 $\xi \neq 0$  に対して、

$$\begin{aligned} \frac{F(x+u) - F(u)}{1 - F(u)} &\approx 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} \\ \Leftrightarrow \frac{1 - F(x+u)}{1 - F(u)} &\approx \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} \end{aligned} \quad (10)$$

がいえる。超過データの個数を  $N_u$  とし、「全サンプル数に対する  $N_u$  の比率」が  $1 - F(u)$  の自然な推定量であることに着目すると、

$$1 - \widehat{F}(u) = \frac{N_u}{n} \quad (11)$$

(11) を (10) に代入すると、

$$F(x+u) \approx 1 - \frac{N_u}{n} \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} \quad (12)$$

が成り立つ。最後に、(12) で  $F(x) = p$  としてから、 $(\xi, \beta)$  を推定値  $(\hat{\xi}, \hat{\beta})$  で置き換え、 $x$  について解くと  $x_p$  の推定値が得られる。

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\xi}} \left[ \left( \frac{n}{N_u} (1-p) \right)^{-\hat{\xi}} - 1 \right] \quad (13)$$

### 2.3 様々な推定手法

このセクションでは、「閾値  $u$  を超過した観測値」に基づく極値パラメータの推定手法を幾つか紹介する。 $X_1, \dots, X_n$  を共通の分布関数  $F$  をもつ独立な確率変数とする。さらに  $F$  が Theorem 2.1.1. の意味で極値分布  $H_{\xi; \mu, \sigma}$  をもつとき、順序統計量  $X_{k,n} \leq X_{k-1,n} \leq \dots \leq X_{1,n}$  ( $X_{1,n}$  は最大の観測値) を利用した以下の3つの推定法が知られている。

**Method 1: Pickands's Estimator for  $\xi \in \mathbf{R}$  (Pickands(1975))**

$$P_{k,n} = \frac{1}{\log 2} \log \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \quad (14)$$

**Method 2: Hill's Estimator for  $\xi > 0$  (Hill(1975))**

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log X_{j,n} - \log X_{k,n} \quad (15)$$

**Method 3: The Dekkers-Einmahl-de Haan Estimator for  $\xi \in \mathbf{R}$  (Dekkers, Einmahl und de Haan(1989))**

$$D_{k,n} = 1 + H_n^{(1)} + \frac{1}{2} \left( \frac{\{H_n^{(1)}\}^2}{H_n^{(2)}} - 1 \right)^{-1} \quad (16)$$

但し、

$$H_n^{(1)} = \frac{1}{k} \sum_{j=1}^k (\log X_{j,n} - \log X_{k+1,n}), \quad H_n^{(2)} = \frac{1}{k} \sum_{j=1}^k (\log X_{j,n} - \log X_{k+1,n})^2$$

閾値  $u$  が  $u = X_{k,n}$  に設定されていると考えると、これら3つの推定量は「閾値  $u$  を超過した観測値」から構成されているとみなせる。特に  $k$  に対して幾つかの条件 ( $k(n) \rightarrow \infty, k(n)/n \rightarrow 0, \text{ as } n \rightarrow \infty$ ) を課せば、これら推定量は一致性、漸近正規性など望ましい性質をもつことが知られている。詳しい議論については、上で紹介した3つの論文を参照されたい。また、上記以外の手法として、Smith (1987) は最尤法の使用を提

案した。  $X_1, \dots, X_k$  を適当に設定した閾値  $u$  に対する  $k$  個の超過値とするとき、一般パレート分布の密度関数が  $f(x) = (1/\beta)(1 + \xi x/\beta)^{-1/\xi-1}$ ,  $\xi \neq 0$  であることから、対数尤度関数は  $1 + \beta^{-1}\xi X_j > 0$ ,  $j = 1, \dots, k$  のとき、

$$l(\xi, \beta; \mathbf{X}) = -k \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^k \log \left(1 + \xi \frac{X_j}{\beta}\right) \quad (17)$$

となる。ある  $j = 1, \dots, k$  で  $1 + \beta^{-1}\xi X_j \leq 0$  のときは、 $l(\xi, \beta; \mathbf{X}) = -\infty$  とおく。また  $\xi = 0$  のとき、対数尤度関数は次の形をとる。

$$l(\beta; \mathbf{X}) = -k \log \beta - \frac{1}{\beta} \sum_{j=1}^k X_j \quad (18)$$

以上で尤度関数が定義され、最尤推定量がこの尤度関数を最大化するパラメータの組  $(\hat{\xi}_n, \hat{\beta}_n)$  として定義される。なお、 $\xi \leq -1/2$  のときは正則条件が満たされず、最尤推定できないことには注意が必要である。また、最尤推定量については以下の漸近的結果も知られている。

$$\sqrt{n} \left( \hat{\xi}_n - \xi, \frac{\hat{\beta}_n}{\beta} - 1 \right) \xrightarrow{d} N(0, M^{-1}), \quad \text{as } n \rightarrow \infty,$$

但し、

$$M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & -1 \\ -1 & 2 \end{pmatrix}$$

### 3 経験尤度法による極値パラメータの推定

このセクションでは、Owen(1988, 1990) が提唱した経験尤度法 (Empirical Likelihood) によって、未知パラメータ  $(\xi, \beta)$  の推定を行う。一般パレート分布では必ずしも 1 次・2 次モーメントが存在するとは限らない (詳しくは Embrechts et al.(1997) 参照) ので、「分布のモーメント」に関する条件を推定に用いることはできないが、代わりに経験分布関数が利用できる。最初に「経験尤度関数」を定義する。

$$L_n(F) = \prod_{k=1}^n (F(X_k) - F(X_{k-})) = \prod_{k=1}^n p_k$$



但し、 $F(\cdot)$  は分布関数で、 $p_k$  ( $k = 1, \dots, n$ ) は観測値  $X_k$  が得られる確率とする。このモデル設定では、全ての観測値が推定に使用されていることに注意されたい。 $p_k \geq 0$  と  $\sum_{k=1}^n p_k = 1$  以外の条件を課さなければ、経験尤度関数  $L_n(F)$  は明らかに、 $p_k = 1/n$  ( $k = 1, \dots, n$ ) のとき最大値をとる。続いて、経験尤度比関数を

$$R_n(F) = \prod_{k=1}^n np_k$$

で定義し、下記の条件の下で  $R_n(F)$  を最大化する  $\hat{\theta}_n = (\hat{\xi}_n, \hat{\beta}_n)$  を最大経験尤度推定量 (maximum empirical likelihood estimator、以下 MELE と記述) と定義する。

$$\left\{ p_k \geq 0, k = 1, \dots, n \quad \sum_{k=1}^n p_k = 1, \sum_{k=1}^n p_k \left[ I_{\{X_k > u + s_l\}} - \frac{N_u}{n} \left( 1 + \xi \frac{s_l}{\beta} \right)^{-1/\xi} \right] = 0, l = 1, \dots, m \right\} \quad (19)$$

この制約条件の意味を説明しておく。(12) から十分大きな  $u$  に対して下記が成り立つことが分かる。

$$1 - P(X \leq u + s_l) - \frac{N_u}{n} \left( 1 + \xi \frac{s_l}{\beta} \right)^{-1/\xi} = 0, \quad s_l > 0 \quad l = 1, \dots, m \quad (20)$$

両辺の期待値をとると、

$$E \left[ I_{\{X > u + s_l\}} - \frac{N_u}{n} \left( 1 + \xi \frac{s_l}{\beta} \right)^{-1/\xi} \right] = 0, \quad l = 1, \dots, m \quad (21)$$

となる。但し、 $I_{\{X > a\}} = 1$  if  $X > a$ ,  $= 0$  if  $X \leq a$  である。 $p_k$  ( $k = 1, \dots, n$ ) が観測値  $X_k$  ( $k = 1, \dots, n$ ) が得られる確率であることに注意すると、(21) は

$$\sum_{k=1}^n p_k \left[ I_{\{X_k > u + s_l\}} - \frac{N_u}{n} \left( 1 + \xi \frac{s_l}{\beta} \right)^{-1/\xi} \right] = 0, \quad l = 1, \dots, m \quad (22)$$

と表される。続いて、 $m \times 1$  ベクトル

$$\mathbf{g}(X_k, \theta) = \left[ I_{\{X_k > u+s_1\}} - \frac{N_u}{n} \left(1 + \xi \frac{s_1}{\beta}\right)^{-1/\xi}, \dots, I_{\{X_k > u+s_m\}} - \frac{N_u}{n} \left(1 + \xi \frac{s_m}{\beta}\right)^{-1/\xi} \right]'$$

を定義した後、ラグランジアンを構成する。

$$L_n(\theta) = \sum_{k=1}^n \log np_k - \mu \left[ \sum_{k=1}^n p_k - 1 \right] - n\lambda' \sum_{k=1}^n p_k \mathbf{g}(X_k, \theta)$$

ここで、 $\mu, \lambda = (\lambda_1, \dots, \lambda_m)'$  とともにラグランジュ乗数である。 $L_n(\theta)$  を  $p_k$  で微分して方程式

$$\frac{\partial L_n(\theta)}{\partial p_k} = \frac{1}{p_k} - \mu - n\lambda' \mathbf{g}(X_k, \theta) = 0$$

を解くと、以下の解が求まる。

$$\hat{\mu} = n$$

$$\hat{p}_k = \frac{1}{n} \frac{1}{1 + \lambda' \mathbf{g}(X_k, \theta)}$$

ここで  $\lambda = \lambda(\theta)$  は

$$0 = \sum_{k=1}^n \hat{p}_k \mathbf{g}(X_k, \theta) = \frac{1}{n} \sum_{k=1}^n \frac{\mathbf{g}(X_k, \theta)}{1 + \lambda' \mathbf{g}(X_k, \theta)}$$

の解となる。このとき、対数尤度関数  $l_n(\theta)$  を最大にするパラメータが  $\theta$  の MELE である。

$$l_n(\theta) = \log \prod_{k=1}^n k \hat{p}_k = - \sum_{k=1}^n \log(1 + \lambda'(\theta) \mathbf{g}(X_k, \theta))$$

## 4 シミュレーション実験

モンテカルロ実験で MELE のパフォーマンスを確かめる。始めに、下記の 4 つの分布に従う独立同一な確率変数をそれぞれ 10000 個発生させる。

元の分布	密度関数 $f$ または分布関数 $F$	$\xi$ の真の値
コーシー分布	$f(x) = (\pi(1+x^2))^{-1}, x \in \mathbf{R}$	1
$\alpha = 3/2$ のパレート分布	$F(x) = 1 - \{x/(1+x)\}^{3/2}, x > 0$	2/3
$\alpha = 5$ のパレート分布	$F(x) = 1 - \{x/(1+x)\}^5, x > 0$	0.2
標準指数分布 Exp(1)	$f(x) = e^{-x}, x > 0$	0

ここで、閾値  $u$  として、(19) で  $u = X_{p,10000}$ ,  $p = 2000, 1500, 1000$  の 3 パターンのいずれかに設定する。但し、 $X_{k,n}$  は  $n$  個のサンプルにおける第  $k$  順序統計量である。また、 $s_1, s_2, \dots$  を下表の通り定める。各閾値  $u$  に対して、平均値、最大値、最小値、 $RMSE = \sqrt{\sum_{i=1}^{100} (\hat{\xi}_i - \xi_0)^2 / 100}$  を計算する。但し、 $\hat{\xi}_i$  は  $i$  番目の推定値、 $\xi_0$  は真の値である。シミュレーション結果については、図表 1-4 (16~19 ページ) を参照されたい。シミュレーション結果によれば、 $\xi_0$  が大きくなるにつれて RMSE も大きくなる、すなわち元の分布の裾が厚くなるにつれ、MELE の推定精度は落ちるといえる。元の分布がコーシー分布 ( $\xi = 1$ ) のときは、各閾値の選び方に対して、 $\xi$  は明らかに過小推定されており、RMSE も元の分布がコーシー分布以外るときと比べて大きくなってしまふ。これに対して、標準指数分布 ( $\xi = 0$ ) や  $\alpha = 5$  のパレート分布 ( $\xi = 0.2$ ) では、 $\xi$  の推定値に殆どバイアスはみられない。

#### 各閾値に対する制約条件

閾値	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
制約条件	$s_1 = X_{1900,10000} - u$	$s_1 = X_{1400,10000} - u$	$s_1 = X_{950,10000} - u$
	$s_2 = X_{1800,10000} - u$	$s_2 = X_{1300,10000} - u$	$s_2 = X_{900,10000} - u$
	$\vdots$	$\vdots$	$\vdots$
	$s_{19} = X_{100,10000} - u$	$s_{14} = X_{100,10000} - u$	$s_{19} = X_{50,10000} - u$

続いて、RMSE を基準に MELE と 10 ページで紹介した Method 1-3 の推定精度の比較を行う。MELE で  $u = X_{p,10000}$  ( $p = 2000, 1500, 1000$ ) に設定することは、Method 1 では  $k = p/4$ 、Method 2-3 では  $k = p$  とすることに相当する。すでに指摘したように、MELE は比較的小さな  $\xi$  に対して高い推定精度を有するので、ケース 1 (標準指数分布 ( $\xi = 0$ )) やケース 2 ( $\alpha = 5$  のパレート分布 ( $\xi = 0.2$ )) では 4 つの推定量の中で、MELE の RMSE が最も小さくなる。一方、大きな  $\xi$  に関しては、Hill's (Method 2) や DEdH (Method 3) の方がパフォーマンスが良い。このことから、実証分析では分布の裾が「極端には」厚くなく、「若干」fat-tail な現象に MELE を用いるとよいことが示唆される。大部分の金融データは、「若干の」fat-tail 性がみられるケースに相当するため、MELE は実証上は他の推定量より優れているともいえる。

### ケース 1 : 指数分布 ( $\xi = 0$ ) の RMSE

	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
MELE	0.0314 (1)	0.0311 (1)	0.0520 (1)
Pickands (Method 1)	0.0858 (3)	0.0985 (3)	0.1153 (3)
Hill (Method 2)	0.4247 (4)	0.3753 (4)	0.3237 (4)
DEdH (Method 3)	0.0646 (2)	0.0570 (2)	0.0521 (2)

( ) は推定精度の順位。

### ケース 2 : $\alpha = 5$ のパレート分布 ( $\xi = 0.2$ ) の RMSE

	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
MELE	0.0349 (1)	0.0333 (1)	0.0549 (1)
Pickands (Method 1)	0.0918 (3)	0.1010 (3)	0.1223 (3)
Hill (Method 2)	0.3332 (4)	0.2847 (4)	0.2348 (4)
DEdH (Method 3)	0.0754 (2)	0.0658 (2)	0.0582 (2)

( ) は推定精度の順位。

ケース 3 :  $\alpha = 3/2$  のパレート分布 ( $\xi = 2/3$ ) の RMSE

	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
MELE	0.0433 (3)	0.0434 (3)	0.0639 (3)
Pickands (Method 1)	0.0996 (4)	0.1034 (4)	0.1335 (4)
Hill (Method 2)	0.0139 (1)	0.0162 (1)	0.0208 (1)
DEdH (Method 3)	0.0270 (2)	0.0307 (2)	0.0388 (2)

( ) は推定精度の順位。

ケース 4 : コーシー分布 ( $\xi = 1$ ) の RMSE

	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
MELE	0.0761 (2)	0.0663 (3)	0.0748 (3)
Pickands (Method 1)	0.1152 (4)	0.1198 (4)	0.1291 (4)
Hill (Method 2)	0.0990 (3)	0.0566 (2)	0.0368 (1)
DEdH (Method 3)	0.0447 (1)	0.0404 (1)	0.0455 (2)

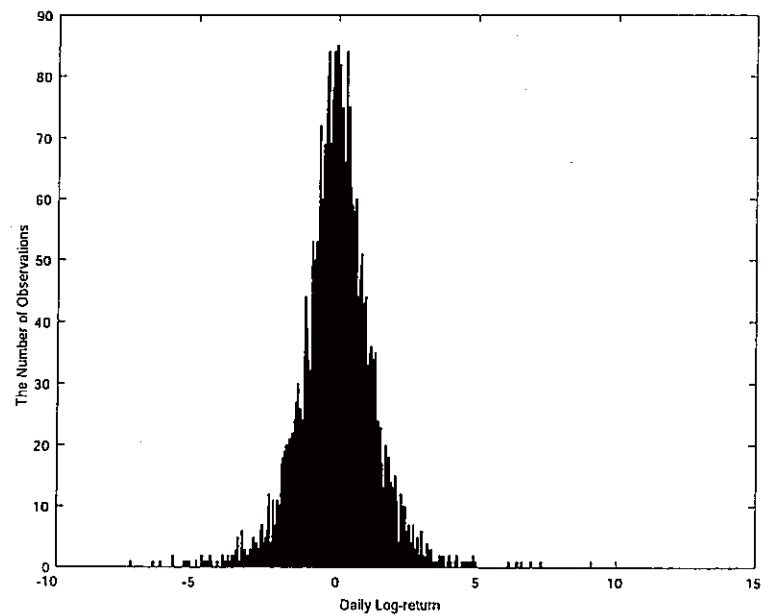
( ) は推定精度の順位。

## 5 実証分析

このセクションでは、1990~2005年の TOPIX の日次対数収益率による MELE の実証分析を行う。始めに収益率に関する基礎統計量をみると、平均値と歪度がほぼ 0 に等しいことから、収益率の分布はほぼ左右対称であることが窺われる。さらに、尖度が 3 を越えていることから、収益率の分布は正規分布より裾が厚い。このことから以下では、分布の裾に注目した分析を行うこととする。

### TOPIX 日次対数収益率の基礎統計量

観測数	3800
平均値	-0.0157
最大値	9.1158
最小値	-7.3646
標準偏差	1.2728
歪度	0.0996
尖度	6.5098



裾の分析に入る前にまず、誤って収益率の分布に正規分布を仮定した場合に何が起こるかみておく。正規分布の平均  $\mu$  と分散  $\sigma^2$  は最尤法から、

$$\tilde{\mu} = \frac{1}{3800} \sum_{i=1}^{3800} z_i = -0.0157$$

$$\tilde{\sigma}^2 = \frac{1}{3800} \sum_{i=1}^{3800} (z_i - \tilde{\mu})^2 = 1.6196$$

と推定される。ここで、 $z_i$  は  $i$  番目の観測値。このとき、各パーセンタイルは以下のように推定される。

$$\begin{aligned}\tilde{x}_{0.999} &= 3.09\bar{\sigma} + \tilde{\mu} = 3.9168 \\ \tilde{x}_{0.99} &= 2.33\bar{\sigma} + \tilde{\mu} = 2.9496 \\ \tilde{x}_{0.01} &= -2.33\bar{\sigma} + \tilde{\mu} = -2.9810 \\ \tilde{x}_{0.001} &= -3.09\bar{\sigma} + \tilde{\mu} = -3.9482\end{aligned}$$

しかし、この場合にはパーセンタイルの推定値には明確なバイアスが存在する。すなわち、全サンプルの 0.58% (22 個) の観測値しか  $\tilde{x}_{0.99}$  を上回っていないほか、全サンプルの 0.55% (21 個) の観測値しか  $\tilde{x}_{0.01}$  を下回っておらず、収益率分布の fat-tail 性が上手く捉えきれていないことが分かる。 $\tilde{x}_{0.99}$  はいわゆるバリューアットリスク (VaR) とも解釈されるため、こうした誤りはリスク管理の観点からは重大な問題であるといえよう。

そこで、収益率の分布が正規分布に従うとの仮定を外し、代わりに極値分布の適用を考えてみる。始めに分布の右端に注目し、(19) で  $u = 1.1294$  (観測値の上位 15 パーセンタイルに相当)、 $u = 1.4295$  (観測値の上位 10 パーセンタイルに相当)、 $u = 1.9989$  (観測値の上位 5 パーセンタイルに相当) の 3 パターンに設定した後、極値パラメータ  $(\xi, \beta)$  とパーセンタイル  $x_{0.99}$ 、 $x_{0.999}$  の推定値を求める。

続いて分布の左端に注目する。このときは、元のデータを  $z_i$ ,  $i = 1, \dots, 3800$  とすると、データセット  $-z_i$ ,  $i = 1, \dots, 3800$  に対して同様の分析を行えばよいことに注意する。この場合、(19) の  $u$  を  $u = -1.1301$  (観測値の下位 15 パーセンタイルに相当)、 $u = -1.4859$  (観測値の下位 10 パーセンタイルに相当)、 $u = -1.9518$  (観測値の下位 5 パーセンタイルに相当) とすれば、 $(\hat{\xi}, \hat{\beta})$  と  $\hat{x}_{0.01}$ 、 $\hat{x}_{0.001}$  が全く同様に計算できる。推定結果は下表の通り。

### 分布の右端の分析

	$u = 1.1294$ (上位 15 パーセンタイル点)	$u = 1.4295$ (上位 10 パーセンタイル点)	$u = 1.9989$ (上位 5 パーセンタイル点)
$\hat{\xi}$	0.0821	0.0505	0.1227
$\hat{\beta}$	0.7481	0.7954	0.7758
$\hat{x}_{0.99}$	3.3982	3.3717	3.3793
$\hat{x}_{0.999}$	5.7664	5.5534	5.8942

### 分布の左端の分析

	$u = -1.1301$ (下位 15 パーセンタイル点)	$u = -1.4859$ (下位 10 パーセンタイル点)	$u = -1.9518$ (下位 5 パーセンタイル点)
$\hat{\xi}$	0.0436	0.1224	0.0297
$\hat{\beta}$	0.7897	0.7135	0.8912
$\hat{x}_{0.01}$	-3.4000	-3.3837	-3.4210
$\hat{x}_{0.001}$	-5.5525	-5.8993	-5.6488

結果をみると、 $\hat{\xi}$  は、閾値  $u$  の選び方によって若干変動しているとはいえ、変動幅は十分許容できる程度であるほか、パーセンタイルの推定値も相当安定している。例えば、閾値  $u$  を上下 10 パーセンタイルにとった場合、TOPIX の日次対数収益率は 100 営業日ごとに 1 回 3.3717% を超えることが期待され、同期間に 1 度だけ -3.3837% を下回ることが予想される。また、全ての推定結果で  $\xi > 0$  となっていることから、収益率の分布は正規分布よりも裾が厚いことが確認される（正規分布では  $\xi = 0$ 。詳しくは Embrechts et al.(1997) の第 3 章を参照）。



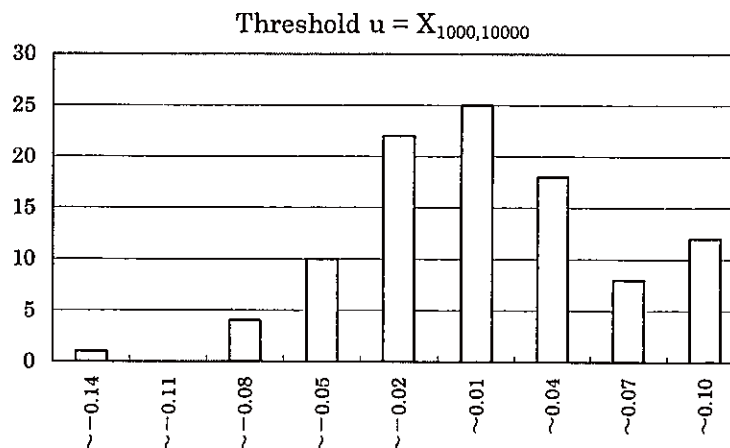
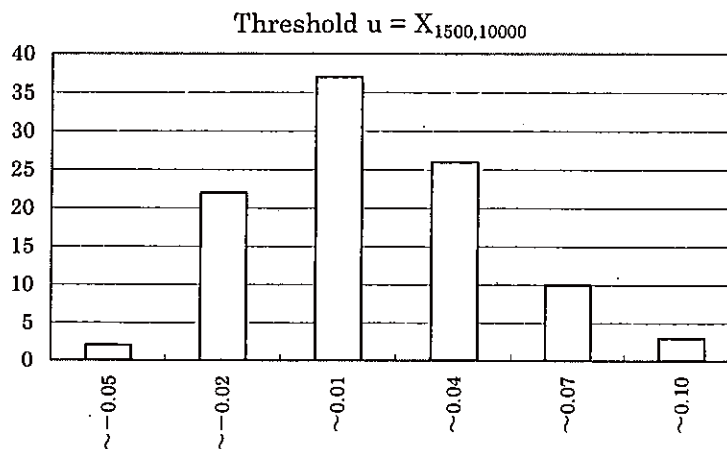
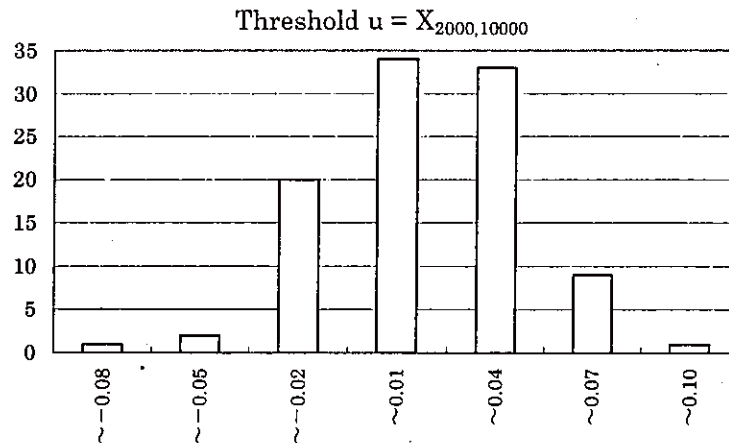
## 参考文献

- [1] Dekkers, A.L.M. and Haan, L. de. (1989). On the estimation of the extreme-value index and large quantile estimation, *The Annals of Statistics*, 17, 1795-1832
- [2] Embrechts, P., Kluppelberg and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag
- [3] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, 3, 1163-1174
- [4] Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251-261
- [5] Owen, A.B. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional, *Biometrika*, 75, 237-249
- [6] Owen, A.B. (1990). Empirical Likelihood Ratio Confidence Regions, *The Annals of Statistics*, 18, 90-120
- [7] Pickands, J. (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, 3, 119-131
- [8] Prescott, P. and Walden, A.T. (1983). Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples, *Journal of the Statistical Computational Simulation*, 16, 241-250
- [9] Smith, R.L. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, 15, 1174-1207

図表 1. 標準指数分布 ( $\xi = 0$ )

Simulation Result 1: Exponential ( $\xi = 0$ )

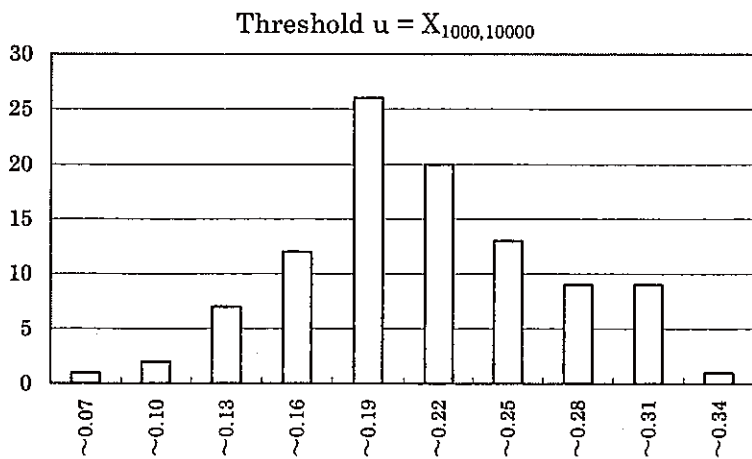
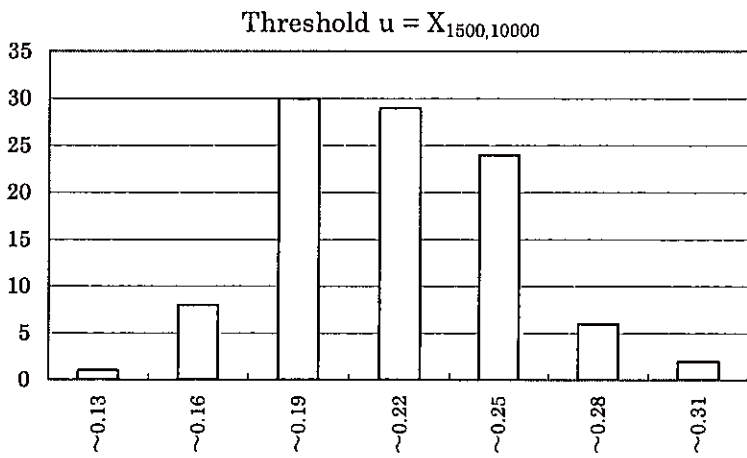
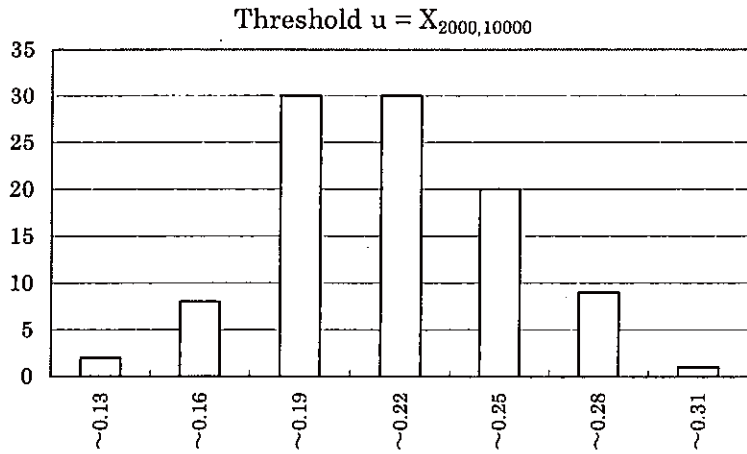
Threshold	$u=X_{2000,10000}$	$u=X_{1500,10000}$	$u=X_{1000,10000}$
Average	0.0019	0.0036	-0.0018
Max	0.0986	0.0914	0.0997
Min	-0.0875	-0.0700	-0.1655
RMSE	0.0314	0.0311	0.0520



図表 2. パレート分布 ( $\xi = 0.2$ )

Simulation Result 2: Pareto ( $\xi = 0.2$ )

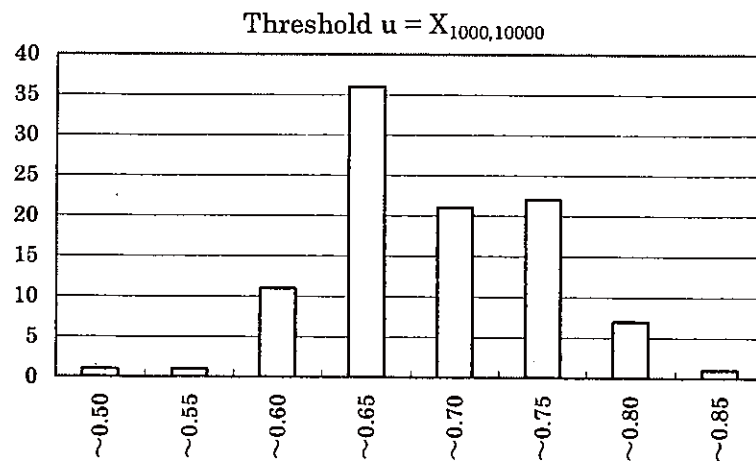
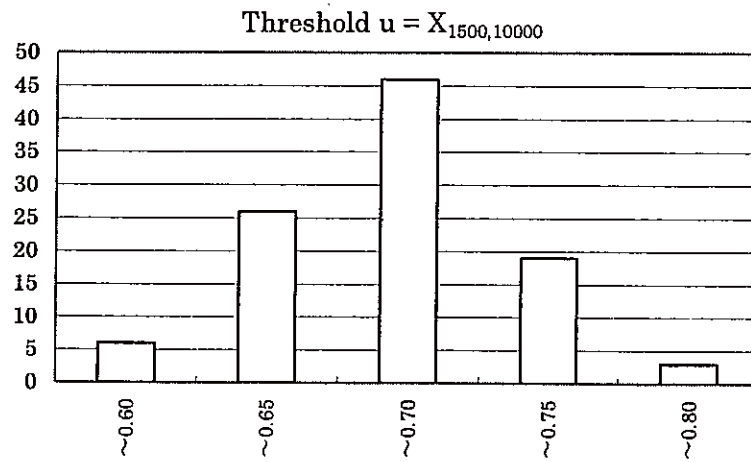
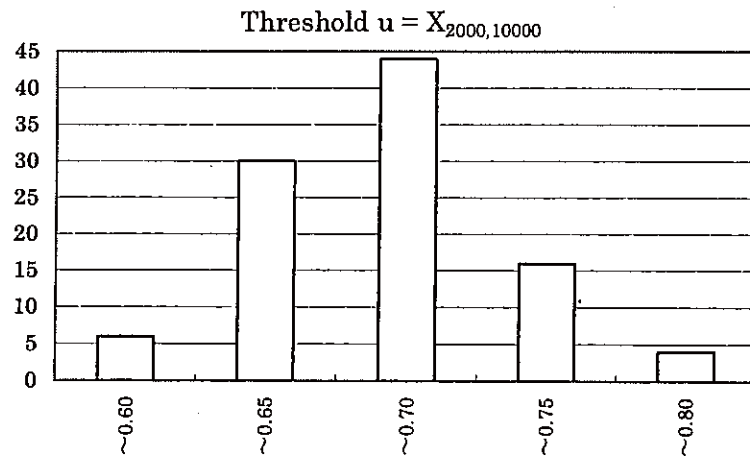
Threshold	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
Average	0.2015	0.2032	0.1986
Max	0.2993	0.3029	0.3167
Min	0.1041	0.1270	0.0216
RMSE	0.0349	0.0333	0.0549



図表 3. パレート分布 ( $\xi = 2/3$ )

Simulation Result : Pareto ( $\xi = 2/3$ )

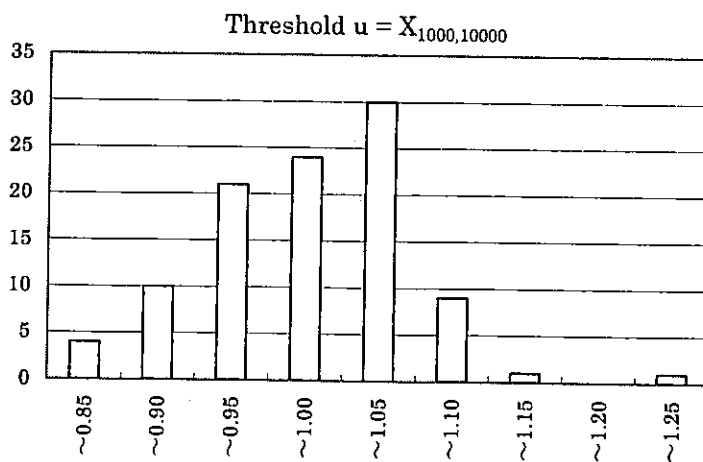
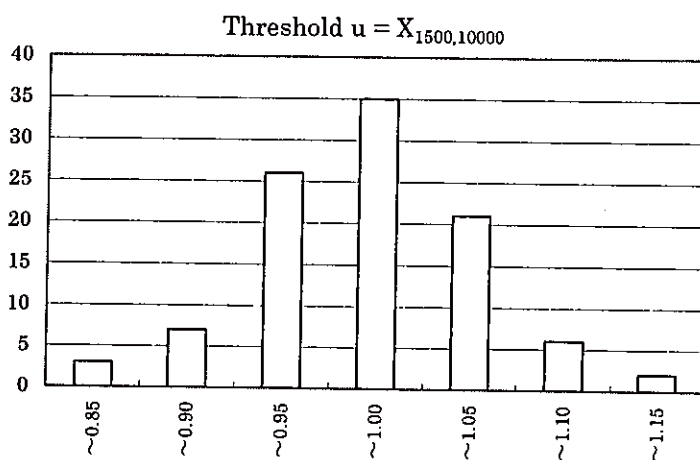
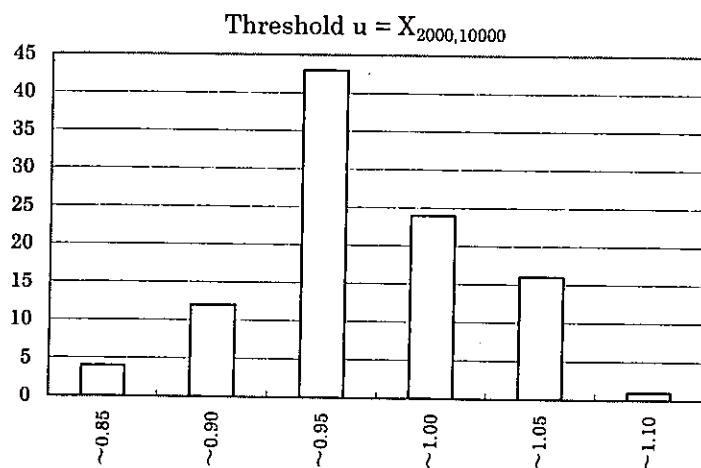
Threshold	$u = X_{2000,10000}$	$u = X_{1500,10000}$	$u = X_{1000,10000}$
Average	0.6669	0.6697	0.6636
Max	0.7654	0.7954	0.8403
Min	0.5530	0.5666	0.4657
RMSE	0.0433	0.0434	0.0639



図表 4. コーシー分布 ( $\xi = 1$ )

Simulation Result 4: Cauchy ( $\xi = 1$ )

Threshold	$u=X_{2000,10000}$	$u=X_{1500,10000}$	$u=X_{1000,10000}$
Average	0.9429	0.9668	0.9763
Max	1.0939	1.1233	1.2182
Min	0.8189	0.8415	0.8078
RMSE	0.0761	0.0663	0.0748



# コピュラの保険への応用例

岡賢一

2006年2月

## 1 はじめに

本稿ではコピュラの保険への応用例を Frees and Valdez(1998) を中心とするサーベイによって紹介する。コピュラとは多次元分布を構成するために用いられる関数のことで、 $U_1, \dots, U_p$  を  $p$  個の一様分布  $U(0, 1)$  に従う確率変数としたとき

$$C(u_1, \dots, u_p) = Pr\{U_1 \leq u_1, \dots, U_p \leq u_p\}$$

を満たす関数  $C$  のことである<sup>1</sup>。さらに任意の周辺分布を  $F_1(x_1), \dots, F_p(x_p)$  と書くと

$$C(F_1(x_1), \dots, F_p(x_p)) = F(x_1, \dots, x_p) \quad (1)$$

とすることで、「周辺分布が  $F_1(x_1), \dots, F_p(x_p)$ 、コピュラが  $C$  である多次元分布」を構成できる。すなわちコピュラを一つに固定すれば任意の周辺分布を持った多次元分布を容易に構成できるわけである。従って変数間の依存関係が重要性を持つような状況において、コピュラの使用が有益であると考えられる。そこで以下ではそのような例を、保険への応用例を中心にみていくことにする。

## 2 多変量生存解析

連合生命保険や連合生命年金の価格を求める上で親子・夫婦・兄弟といった複数の生命の同時生存時間を推定する必要がある。しかし親子・夫婦・兄弟といった関連の深い生命の生存時間については「broken heart syndrome」「common risk factor」といった要因により、依存性が存在することが知られている。「broken heart syndrome」とは夫婦の一方が亡くなった直後にもう一方の死亡率が大きく上昇するという経験的に観測される現象のことであり、「common risk factor」とは家族間の遺伝的な類似性や共同生活者間の生活習慣の類似性などのことである。いずれも生命間の依存性の要因になる可能性があるといえるだろう。そこで以下では、Frailty モデルと呼ばれるコピュラを用いた多変量生存解析モデルを見ていく。

まず、一変量で記号を設定する。各記号について添え字  $i$  をつけた場合、個体  $i$  に関する記号であることを意味する。 $T$  を生存時間とし、生存関数  $S(t)$ 、ハザード関数  $h(t)$  を

$$S(t) = Pr\{T > t\} = 1 - F(t)$$

<sup>1</sup>コピュラの厳密な定義・性質は、例えば Nelsen(1998) に詳しい。

$$h(t) = -\frac{\partial \log S(t)}{\partial t} = \frac{f(t)}{S(t)}$$

と定義する。さらに Cox の比例ハザードモデルを考えるとハザード関数は

$$h(t, Z) = e^{\beta Z} b(t)$$

となる。ここで  $Z$  は説明変数、 $\beta$  は未知パラメータ、 $b(t)$  はベースハザードである。このとき

$$\gamma = e^{\beta Z}$$

$$B(t) = \exp\left(-\int_0^t b(s) ds\right)$$

とすると生存関数  $S(t|\gamma)$  は

$$S(t|\gamma) = \exp\left(-\int_0^t h(s, Z) ds\right) = B(t)^\gamma$$

と書ける。ここで  $Z$ 、すなわち  $\gamma$  を未知としたものが Frailty モデルである。 $\gamma$  が大きいほど生存関数の値が小さくなることから  $\gamma$  が「もろさ (frailty)」を表すことになる。 $\gamma$  が未知であるので、生存関数  $S(t)$  は  $\gamma$  についての期待値をとって  $S(t) = E_\gamma[S(t|\gamma)]$  とすることで得られる。

次にこれを複数の生命で考える。この場合も  $\gamma$  を未知とすることから、多変量 Frailty モデルは生命間の依存性の要因を特定したくない場合、あるいは依存性の要因がわからない場合を想定しているといえる。 $T_1, \dots, T_p$  を  $p$  個の個体の生存時間とし、 $\gamma$  が与えられた下で互いに独立であるとする。すなわち

$$\begin{aligned} Pr\{T_1 > t_1, \dots, T_p > t_p | \gamma\} &= Pr\{T_1 > t_1 | \gamma\} \cdots Pr\{T_p > t_p | \gamma\} \\ &= S_1(t_1 | \gamma) \cdots S_p(t_p | \gamma) \\ &= B_1(t_1)^\gamma \cdots B_p(t_p)^\gamma \end{aligned}$$

と仮定し、依存性の要因を  $\gamma$  のみで表すわけである。このとき多変量生存関数は、一変量の場合と同様に  $\gamma$  に関して期待値をとって

$$Pr\{T_1 > t_1, \dots, T_p > t_p\} = E_\gamma\{[B_1(t_1) \cdots B_p(t_p)]^\gamma\} \quad (2)$$

と定義される。

この (2) 式について、Hougaard(1986) は  $\gamma$  に安定分布を仮定して具体的に評価している。 $\gamma$  が指数  $\alpha$  の安定分布に従うとすると、 $E_\gamma[e^{-s^\alpha}] = \exp(-s^\alpha)$  となることを踏まえ (2) 式を変形すると

$$\begin{aligned} Pr\{T_1 > t_1, \dots, T_p > t_p\} &= E_\gamma[\exp(\gamma \log\{B_1(t_1) \cdots B_p(t_p)\})] \\ &= \exp(-\{-\log B_1(t_1) - \cdots - \log B_p(t_p)\}^\alpha) \\ &= \exp(-\{[-\log S_1(t_1)]^{1/\alpha} + \cdots + [-\log S_p(t_p)]^{1/\alpha}\}^\alpha) \quad (3) \end{aligned}$$

とすることができる。ここで最後の等式は

$$S(t) = E_\gamma[S(t|\gamma)] = E_\gamma[B(t)^\gamma] = E_\gamma[e^{-(-\gamma \log B(t))}] = \exp(-(-\log B(t))^\alpha)$$

となることを用いている。この(3)式はコピュラを用いて記述された式と解釈することができる。さらに Hougaard(1986) は  $p = 2$  のとき周辺分布にワイブル分布を仮定し (すなわち  $i = 1, 2$  について  $B_i(t) = \exp(-a_i t^{b_i})$ 、 $S_i(t|\gamma) = \exp(-a_i \gamma t^{b_i})$  を仮定し)、(3)式を

$$Pr\{T_1 > t_1, T_2 > t_2\} = \exp\left(-[a_1 t_1^{b_1} + a_2 t_2^{b_2}]^\alpha\right)$$

と具体的に求めている。

### 3 Competing Risk

主に医学統計などの分野で研究されている competing risk 問題であるが、保険の対象となる人や物についても様々な要因で死亡・破損する可能性があるといえるので competing risk の議論が応用できると考えられる。そこで以下では competing risk の議論をコピュラの枠組みで見ていく。

#### 3.1 Competing Risk について

例えば人はガンや心臓病や事故など様々な要因で亡くなるわけだが、そのような要因として  $p$  個の要因を考え、その個別の各要因のみに関する架空の生存時間を  $T_1, \dots, T_p$  と書くことにする。これらは実際に観測することはできない。実際にはこれらのうち一つの要因で亡くなるわけであるので、実際の生存時間  $T = \min(T_1, \dots, T_p)$  及びどの要因で亡くなったかということが観測できることになる。

この生存時間について、各  $T_1, \dots, T_p$  に関する依存性の存在を仮定し生存関数を求める。記号については2節と同様のものを用い、さらに2節と同様に各  $T_1, \dots, T_p$  は  $\gamma$  が与えられた下で互いに独立とする (すなわち依存性の要因は  $\gamma$  のみとする)。すると

$$\begin{aligned} Pr\{T > t|\gamma\} &= Pr\{\min(T_1, \dots, T_p) > t\} \\ &= Pr\{T_1 > t|\gamma\} \cdots Pr\{T_p > t|\gamma\} \\ &= B_1(t)^\gamma \cdots B_p(t)^\gamma \end{aligned}$$

となり、 $\gamma$  についての期待値をとることで

$$Pr\{T > t\} = E_\gamma[\{B_1(t) \cdots B_p(t)\}^\gamma]$$

と書ける。さらにここでも2節と同様に  $\gamma$  に指数  $\alpha$  の安定分布を仮定することで

$$Pr\{T > t\} = \exp\left(-\{[-\log S_1(t)]^{1/\alpha} + \cdots + [-\log S_p(t)]^{1/\alpha}\}^\alpha\right)$$

と書け、この式はコピュラを用いて記述された式と解釈することができる。この場合も、 $p = 2$  のとき周辺分布にワイブル分布を仮定すると

$$Pr\{T > t\} = \exp\left(-[a_1 t^{b_1} + a_2 t^{b_2}]^\alpha\right)$$

と具体的な評価が得られる。



### 3.2 Competing Riskの保険への応用

生命保険契約や生命年金契約では、死亡の他に解約(失効等を含める)による脱退によって契約が消滅することが多い。そこで Valdez(2001)は、これら二つの脱退要因に関する生存関数を考え、「逆選択」と呼ばれる現象が実際に起こっているかということの検証に用いている。生命保険契約における逆選択とは、死力<sup>2</sup>の低い人(つまり健康な人)ほど解約してしまい、時間が経つにつれて死力の高い人、すなわち保険金を払う必要が生じる可能性の高い人が被保険者として残ってしまうという現象のことである。また生命年金における逆選択とは、死力の高い人ほど解約してしまい、多くの年金を支給しなければならない健康な人ばかりが被保険者として残ってしまうという現象のことである。何れの場合も保険会社にとって不利益をもたらす現象といえるだろう。従って、逆選択が実際に生じているのかということは、保険会社にとって重要な問題となる。

この逆選択の有無を正確に検証するためには、死亡と解約という二つの脱退要因の依存性を考えることが必要になる。厳密な議論は省略するが、もし死力と解約の起こりやすさに正の相関があれば健康な人ほど被保険者として残るわけであり、生命年金契約における逆選択が疑われる。一方負の相関があれば、不健康な人ほど被保険者として残るわけであり、生命保険契約における逆選択が生じている可能性があるわけである。従って、直感的にはあるが、死亡と解約の依存性を考慮に入れる必要があることがわかるだろう。

そこで Valdez(2001)は、周辺分布として死亡年齢にゴムパーツ分布<sup>3</sup>、解約年齢に指数分布を仮定し、コピュラとして Franc コピュラを用いて生存関数を推定している。さらに逆選択を数学的に定義し、推定された生存関数を用いて導出した統計量により、逆選択の有無を検定する方法を提案している。そしてその統計量を用いた実証分析により、データによっては逆選択が生じている可能性があるということを示唆している。

## 4 Loss-Expense データの分析

### 4.1 同時分布の推定

ここでは Frees and Valdez(1998)Section4 に従い、損害保険会社における損害保障支払額<sup>4</sup>とそれに付随する調整費用<sup>5</sup>の分析例を紹介する。損害保険会社の各クレームは損害保障額 (loss、 $X_{11}, \dots, X_{1n}$  とする) とそれに付随する調整費用 (ALAE、 $X_{21}, \dots, X_{2n}$  とする) からなる。ここで調整費用とは、弁護士への支払や調査費用からなる費用のことである。これらの変数は正の相関を持つことが経験的に知られており、Frees and Valdez(1998) で用いられたデータ ( $n = 1500$ ) では相関係数が 0.41 となっている。これは大規模な事故ほど調査などのための費用がかさむということの意味しており、直感的にも妥当な数字であるといえるだろう。この依存関係を踏まえ、これらの変数の同時分布を適切に求めることがここでの目的となる。

Frees and Valdez(1998) では分析に入る前にデータに関する注意として、loss データにおける

<sup>2</sup>死亡脱退に関するハザード率のことである。

<sup>3</sup>例えば「生命保険数学(上巻)」(二見隆、1992)でゴムパーツの法則として紹介されている。

<sup>4</sup>「損保数理」(日本アクチュアリー会、2000)の1-5ページで「保険料に対する保険金」と呼ばれている金額のことである。

<sup>5</sup>「損保数理」(日本アクチュアリー会、2000)の1-5ページで「損害の調査および立証のためまたはそれに付随する費用(これを損害調査費という)」と呼ばれている費用のことである。

センサリングの問題が挙げられている。これは保険金の支払額に上限 (policy limit) が設けられていることに起因するセンサリングである。Frees and Valdez(1998) で用いられた 1500 件のクレームデータでは 34 件で支払限度額と実際の支払額が一致しており、実際にセンサリングが観察されている。尤度関数を求める際などに注意が必要となる。

コピュラを用いた分析では、まず各変数の周辺分布を特定する必要がある。Frees and Valdez(1998) ではいずれの変数についてもパレート分布

$$F(x) = 1 - \left( \frac{\lambda}{\lambda + x} \right)^\theta$$

が用いられており、経験分布との比較によりフィットのよさが確かめられている。

次に最適なコピュラを選択する。最適なコピュラを選択は以下の手順 (Genest and Rivest(1993)) で進められる。

1. 候補となるコピュラの族を決める。ここでは

- Gumbel-Hougaard コピュラ:  $C(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$
- Frank コピュラ:  $C(u, v) = \frac{1}{\alpha} \log \left( 1 + \frac{(e^{\alpha u} - 1)(e^{\alpha v} - 1)}{e^{\alpha} - 1} \right)$
- Cook-Johnson コピュラ:  $C(u, v) = \exp\{-[(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha}\}$

を候補とする。いずれのコピュラも未知パラメータは一つである。

2. 各コピュラについて相関の指標として知られるケンドールの  $\tau$  を求める。ケンドールの  $\tau$  はコピュラを用いて

$$\tau = 4 \iint C(u, v) dC(u, v) - 1$$

と表現することができ<sup>6</sup>、候補となっているコピュラについてこのケンドールの  $\tau$  を求めると

- Gumbel-Hougaard コピュラ:  $\frac{\alpha}{\alpha+2}$
- Frank コピュラ:  $1 - \frac{4}{\alpha} \left\{ \frac{1}{-\alpha} \int_0^{-\alpha} \frac{t}{e^{-1} - 1} dt - 1 \right\}$
- Cook-Johnson コピュラ:  $1 - \alpha^{-1}$

となる。

3. 与えられたデータについて

$$\tau_n = \frac{1}{n C_2} \sum_{i < j} \text{sign}[(X_{1i} - X_{1j})(X_{2i} - X_{2j})]$$

によって、経験的なケンドールの  $\tau_n$  を求める。

<sup>6</sup>詳しくは Nelsen(1997)。他の順位相関の指標であるスピアマンの  $\rho$  もコピュラを用いて

$$\rho = 12 \iint \{C(u, v) - uv\} dudv$$

と表現することができ、よく用いられる。これらの指標は確率変数の単調増加関数による変換について不変性を持っており、確率変数のスケールに影響されない。

4. 手順2で各コピュラについて求めた  $\tau$  と、手順3でデータから求めた  $\tau_n$  を用い

$$\tau = \tau_n$$

とすることで各コピュラに関して  $\alpha$  の推定値  $\alpha_n$  を求める。

5.  $Z_i = F(X_{1i}, X_{2i})$  という確率変数を考え、その分布関数を  $K(z) = Pr\{Z_i \leq z\}$  とする。この  $K(z)$  は各コピュラについて

$$K(z) = z - \frac{\phi(z)}{\phi'(z)}$$

と求められることが知られている。ここで  $\phi(z)$  は generator と呼ばれる関数で、各コピュラについて

- Gumbel-Hougaard コピュラ:  $\phi(z) = z^{-\alpha} - 1$
- Frank コピュラ:  $\phi(z) = \log \frac{e^{\alpha z} - 1}{e^{\alpha} - 1}$
- Cook-Johnson コピュラ:  $\phi(z) = (-\log z)^\alpha$

となることが知られている。

6. 手順5で求めた各コピュラの  $K(z)$  に、手順4で求めた  $\alpha$  の推定値  $\alpha_n$  を代入する。これにより、各コピュラについて  $K(z)$  の具体的な形状がパラメトリックに得られたわけである。

7. データからノンパラメトリックに  $K(z)$  を求め、それを  $K_n(z)$  とする。  $K(z)$  をノンパラメトリックに求める具体的な手順は

- (a)  $Z_i$  の観測値を  $z_i = \{X_{1j} < X_{1i} \text{かつ } X_{2j} < X_{2i} \text{となる } (X_{1j}, X_{2j}) \text{の数}\} / (n-1)$  として  $i = 1, \dots, n$  について求める。
- (b)  $K_n(z) = \{z \text{以下となる } Z_i \text{の割合}\}$  とする。  
である。

8. 手順6で求めた  $K(z)$  と手順7で求めた  $K_n(z)$  の q-q プロットを行い、  $K(z)$  及び  $K_n(z)$  に対応する2本の線が最も近いコピュラを最適なコピュラとして選択する。

以上が最適なコピュラを選択する方法である。Frees and Valdez(1998)ではこの方法により、Gumbel-Hougaard コピュラが選択されている。

ここまでの議論でコピュラ・周辺分布が決まったので、(1)式により用いる2次元分布  $F(x_1, x_2)$  が特定される。そこで次に最尤法を用いて改めてこの2次元分布のパラメータを推定し直す。コピュラのパラメータは1つ、周辺分布であるパレート分布のパラメータは2つであり、2次元分布であるので合計で5つの未知パラメータを推定することになる。尤度関数を求めるにあたっては、センサーされたデータについて分布関数の偏微分を用いて尤度関数への貢献を調節する必要がある<sup>7</sup>。

Frees and Valdez(1998)ではデータからパラメータの推定がなされ、その推定値をもとにスピアマンの  $\rho$  が31%と計算されている。これは損害保障支払額とそれに付随する調整費用に正の相関があるという経験的な事実と整合的な結果といえるだろう。さらに Frees and Valdez(1998)では各コピュラについて AIC も計算されており、AIC から Gumbel-Hougaard コピュラの使用が支持されている。最適なコピュラ選択には、上で用いた Genest and Rivest(1993)の方法や

<sup>7</sup>詳しくは Frees and Valdez(1998)Appendix A 参照。

AICによる方法など様々な手法がこれまでに検討されているが絶対的な選択基準は未だないようである。

## 4.2 推定された同時分布の利用

前節により損害保障支払額 ( $X_1$ ) とそれに付随する調整費用 ( $X_2$ ) の2次元分布が推定された。ここではそれを用いた再保険会社の負担額の予測という応用例を紹介する。推定に用いたデータの契約を有する会社が再保険会社と契約し、支払を分担という状況を考える。そのとき再保険会社の負担額  $g(X_1, X_2)$  を

$$g(X_1, X_2) = \begin{cases} 0 & \text{if } X_1 < R \\ X_1 - R + \frac{X_1 - R}{X_1} X_2 & \text{if } R \leq X_1 < L \\ L - R + \frac{L - R}{L} X_2 & \text{if } X_1 \geq L \end{cases}$$

とする。ここで  $R$  はエクセスポイント、 $L$  は支払限度額である。この式は、 $R$  以下の少額支払については再保険会社は支払を分担しない、 $R$  と  $L$  の間の支払については  $R$  を超える分を再保険会社が負担し、調整費用についても損害保障支払額の分担割合と同等の割合で分担する、 $L$  を超える支払については上限である  $L$  と調整費用を と同様の方法で分担する、ということを表している。すなわち超過損害額再保険を想定しているわけである。

このとき再保険会社の期待負担額は  $E[g(X_1, X_2)]$  となる。そこで Frees and Valdez(1998) では前節で求めた2次元分布からの乱数を100000回発生させ<sup>8</sup>、この期待負担額  $g^*$  と標準誤差  $se(g^*)$  を

$$g^* = \frac{1}{100000} \sum_{j=1}^{100000} g(x_{1j}, x_{2j})$$

$$se(g^*) = \sqrt{\frac{\frac{1}{100000} \sum_{j=1}^{100000} g(x_{1j}, x_{2j})^2 - g^{*2}}{100000}}$$

としてシミュレーションにより計算している。ここで  $R$  及び  $L$  は様々に設定されており、 $R$  及び  $L$  の期待負担額への感応度が調べられている。その結果、支払限度額  $L$  が高いほど期待負担額は高くなる、 $R/L$  が高いほど期待負担額が低くなるという結論を得ている。

さらにこれらの変数が独立であるという仮定の下でも同様のシミュレーションがなされており<sup>9</sup>、 $R/L$  及び  $L$  が高くなるほど依存性を考慮したモデルに比べ期待負担額を過小に評価してしまうというシミュレーション結果を得ている。

## 4.3 その後の発展

この損害保障支払額とそれに付随する調整費用の例は、Frees and Valdez(1998)以降他のいくつかの論文でも研究されている。Klugman and Parsa(1999)では周辺分布として様々な分布を尤度により比較し、損害保障支払額に inverse paralogistic 分布、調整費用に inverse Burr 分布を用

<sup>8</sup>乱数の発生方法については、Frees and Valdez(1998)Section3に詳しい。

<sup>9</sup>これは独立コピュラ  $C(u, v) = uv$  を用いるということである。

いている。さらにコピュラとして扱いの容易さから Franc コピュラを用い、二変量 goodness-of-fit について議論している。Cebrian et al.(2003) では周辺分布に指数分布を用い、極値コピュラによる分析を行っている。その結果、変数間の依存性を考慮しない場合再保険会社にかかるプレミアムが誤って推定されてしまうという Frees and Valdez(1998) と同様の結論を得ている。さらに Denuit et al.(2006) では損害保障支払額のセンサリングを考慮した上でコピュラの選択を行い (Frees and Valdez(1998) では、コピュラの選択はセンサリングを無視して行われている)、ノンパラメトリックに推定した Archimedean コピュラを用いた分析を行っている。

## 5 ソルベンシー評価への応用

保険の規制・監督に関する国際基準の設立を目指し、2004年に国際アクチュアリー会 (IAA) によって「A Global Framework for Insurer Solvency Assessment」が公表された。その中で、異なるリスク間の依存構造を扱う手法としてコピュラの使用が有益であるということが述べられている。このように近年コピュラのソルベンシー評価への応用が注目されている。

1984年以降銀行の規制・監督に関する国際基準の設立の議論が盛り上がり、1988年にバーゼル委員会により BIS 規制 (バーゼル ) が策定された。さらにその後も BIS 規制見直し作業が行われ、2006年末から新 BIS 規制 (バーゼル ) の適用が開始される。このような銀行業監督の国際基準策定の動きの影響を受け、国際的な保険業監督・規制に関する検討を使命とする保険監督者国際機構 (IAIS) が 1994年に設立された。そして IAIS は「ソルベンシー評価問題」を最優先課題の一つとして採択し、国際アクチュアリー会 (IAA) への協力を要請することになる。この要請を受けて 2002年、IAA 保険監督委員会に作業部会が設置され、ソルベンシー確保に必要なファンド総額を計量する手法、国際的なリスクベースのソルベンシー資本システムの基礎、リスクとその相互依存関係の最良測定方法などに関する議論がなされた。その結果 2004年に公表されたのが、上述の「A Global Framework for Insurer Solvency Assessment」である。

ここで IAIS によると、ソルベンシーとは「いつでも全ての契約に関わる負債を支払うことができる能力」とされている。しかし多くの長期契約を保有する保険業においては、環境の変化による保険事故率の上昇、インフレによる保険金支払額の増加、通常は発生しない巨大災害の発生といった保険金支払を困難にする状況が存在する。そのような場合であっても保険金を確実に支払える健全性が保険会社には求められており、そのために適切な額の資本を持つことが必要となる。では、各保険会社が有している資本はどの程度適切な額であるといえるだろうか。これがソルベンシー評価問題である。

この問題に対処するためには、保険会社において通常の予測を超えて生じうるリスクの額を見積もる必要がある。例えば損害保険会社の場合、リスクは 保険リスク、 予定利率リスク、 資産運用リスク、 経営管理リスクというように分類することができ、それぞれさらに細分類される。このように分類された各リスクについて損失分布を求め、それに基づいて通常の予測を超えて生じうるリスクの額を見積もるわけだが、会社としてのリスクを算出するためにはこれら各リスクの損失分布を結びつけなくてはならない。その過程でコピュラの利用が有益となる可能性があるわけである。

「A Global Framework for Insurer Solvency Assessment」では、全社的要求 (Total Company

Requirement) の項目でコピュラについて

It is more likely necessary to construct dependency models that reflect observed and expected dependencies without formalizing the structure of those dependencies with cause-effect models. The theory of copulas provides a comprehensive modelling tool that can reflect dependencies in a very flexible way.

While structural dependencies are modelled directly in an internal model, empirical dependencies are most easily modelled using specific dependency models called copulas. Copulas describe the relationship between the quantiles of distributions of different risks. They can be selected in order to recognize so-called "tail-dependencies" where dependencies only occur, or only appear, in extreme circumstances. Appendix gives an overview of some technical aspects of dependency modelling using copulas.

と言及している。さらに Appendix でコピュラの簡単な紹介がなされており、その中でコピュラが必要となる理由について

The model for  $x$ (=overall risk of the company, which can be described as  $x = \sum_j x_j$ , where  $x_j$  are risk components) is completely specified if we assume a multivariate Normal setting in which each component has a univariate normal distribution and all dependencies are expressed through correlations. However, insurance claim data immediately show shortcomings of this assumption as,

- loss distribution are usually skewed and heavy tailed(i.e., the downside risk due to large losses is substantial),
- dependency between risks usually increases in the tails(i.e., various lines of business may look almost independent in "normal" situations, but they are strongly correlated in the tails as occurred with September 11, 2001).

と書かれている(一部改)。すなわち、コピュラを用いてリスク間の依存性をより柔軟に反映したモデルが望まれているわけである。しかし一方で、具体的なコピュラを選択方法や周辺分布の設定方法などについての記述はない。従って、保険リスク分析に適したコピュラ利用手順の検討が今後の課題となるだろう。

## 参考文献

- [1] Cebrian, A.C., Denuit, M. and Lambert, P.(2003). Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database. *Belgian Actuarial Bulletin*, 3, 33-41.
- [2] Denuit, M. and Purcaru, O.(2004). How to deal with correlated risks in actuarial science? *Proceedings of the 2nd Actuarial and Financial Mathematics Day*, Brussels, Belgium.

- [3] Denuit, M., Purcaru, O. and Keilegom, I.V.(2006). Bivariate Archimedean copula modeling for loss-ALAE data in non-life insurance. *Journal of Actuarial Practice*, forthcoming
- [4] Free, E.W. and Valdez, E.A.(1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2, 1-25.
- [5] Genest, C. and Rivest, L.P.(1993). Statistical inference procedure for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88, 1034-1043.
- [6] Hougaard, P.(1986). A class of multivariate survival. *Biometrika*, 73, 671-678.
- [7] Klugman, S.A. and Parsa, R.(1999). Fitting bivariate loss distribution with copulas. *Insurance: Mathematics and Economics*, 24, 139-148.
- [8] Nelsen, R.B.(1999). *An Introduction to Copulas*. Springer-Verlag, New York.
- [9] Valdez, E.A.(2001). Bivariate analysis of survivorship and persistency. *Insurance: Mathematics and Economics*, 29, 357-373.

# コピュラの性質とその応用\*

広瀬光久†

2006年2月14日

## 1 はじめに

多変量データを統計的に扱うにあたり、その同時分布を推定することは、全ての分析の根本に位置する最も基本的な問題である。特に金融データを分析する際、資産の収益率分布に多変量正規分布を仮定することが多く、その根本部分を再考することは非常に重要であると考えられる。

例えば、一次元の金融データにおいてはその収益率分布が正規分布に従っていないということがよく知られている。正規分布よりも裾の厚い、かつ歪んだ分布であるということがしばしば議論され、この実際の形状を表現するために多くのパラメトリックな分布が提案されているようである。

つまり、これと同様のことを多次元の分布についても考えようというのが本論文の問題意識である。

しかし一次元の場合とは異なり、多変量分布の多くは（たとえ2次元であっても）複雑で扱いにくいものばかりである。そのため本論文ではコピュラ (copula) と呼ばれる概念を導入し、多変量の分布を簡単に噛み砕くことから始める。これは同時分布を「周辺分布と依存構造」に分解することができる非常に便利なツールであり、論文の前半はこのコピュラを理論的に解明することが目的となる。そして多変量正規分布に内在する「正規型依存構造」の意味と問題点をはっきりとさせた上で、新たなパラメトリック分布を導入し、当初の問題意識のもと実際の金融データへの応用を試みることになる。

なお、本論文における統計解析・シミュレーションの全ては R2.2.0 によって行う。

## 2 コピュラとは

### 2.1 定義

ここでは Carmona (2004) に従い以下のように定義する。<sup>\*1</sup>

**定義** コピュラとは一様分布する確率変数群の同時分布関数である。

このコピュラが広く使われる理由として、以下の定理の存在が大きい。

**定理 1 (Sklar)** 確率変数  $(X_1, \dots, X_n)$  において、同時分布関数  $F$ 、周辺分布関数を  $F_1, \dots, F_n$  とすれば、

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad \text{for all } x_1, \dots, x_n \in \bar{\mathbf{R}} \quad (2.1)$$

を満たすコピュラ  $C$  が常に存在し、特に確率変数  $(X_1, \dots, X_n)$  が連続なら  $C$  は唯一つである<sup>\*2</sup>。逆に  $C$  がコピュラであるなら、(2.1) によって定義された  $F$  は  $F_1, \dots, F_n$  を周辺分布に持つ同時分布関数である。

\* 本論文を書くにあたって、いろいろ御指導して下さった国友直人先生に感謝したい。

† 東京大学経済学部 4 年。

<sup>\*1</sup> サブコピュラを含んだより厳密な定義に関しては、例えば Nelsen (1999) を参照。本論文では連続な確率変数を主に扱うため、サブコピュラの定義は特に必要ないと判断した。

<sup>\*2</sup> 正確には「 $\text{Ran} F_1 \times \dots \times \text{Ran} F_n$  を定義域とするサブコピュラが一意に存在する」ということになる (Ran は値域)。



(証明は Nelsen (1999) を参照)

このようにして定まる  $C$  を,  $(X_1, \dots, X_n)$  のコピュラと呼ぶ.  $F$  が連続ならば, (2.1) に  $x_1 = F_1^{-1}(u_1), \dots, x_n = F_n^{-1}(u_n)$  を代入し

$$\begin{aligned} C(u_1, \dots, u_n) &= F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \\ &= P(F_1(X_1) \leq u_1, \dots, F_n(X_n) \leq u_n) \quad \text{for all } u_1, \dots, u_n \in [0, 1]. \end{aligned} \quad (2.2)$$

ゆえに  $C$  は  $(F_1(X_1), \dots, F_n(X_n))$  の同時分布関数である. すなわち,  $(X_1, \dots, X_n)$  を  $(F_1(X_1), \dots, F_n(X_n))$  と変換し周辺分布  $F_1, \dots, F_n$  の影響を完全に排除することで<sup>\*3</sup>, 純粋な依存構造のみをコピュラ  $C$  として抽出したということになる. このように, 多変量同時分布関数は (2.1) 式によって依存構造 (=コピュラ) と周辺分布とに完全に分解することができるのである.

また  $C$  が絶対連続であると仮定すると, コピュラの密度関数  $c$  は (2.2) 式より

$$c(u_1, \dots, u_n) = \frac{f(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))}{f_n(F_1^{-1}(u_n)) \cdots f_n(F_n^{-1}(u_n))} \quad (2.3)$$

と計算される.

## 2.2 一般的性質

本論を理解する上で必要な, いくつかの重要な性質を列挙する. 各定理の証明は Nelsen (1999), Cherubini et al. (2004) を参照. なお本論文では, これ以降簡単のため 2 次元のコピュラのみを扱うこととする.

**定理 2**  $(X_1, X_2)$  を分布関数  $F$  に従う連続な確率変数, そのコピュラを  $C$  とすると

$$\text{ケンドールの } \tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1$$

が成り立つ. すなわち  $\tau$  は  $C$  に対して一意に定まる<sup>\*4</sup>.

**定理 3** (Fréchet-Hoeffding の不等式) 任意のコピュラ  $C$  に対し,

$$\max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2) \quad \text{for all } u_1, u_2 \in [0, 1]$$

が成り立ち,  $\max(u_1 + u_2 - 1, 0)$ ,  $\min(u_1, u_2)$  もまた  $[0, 1] \times [0, 1]$  においてコピュラである<sup>\*5</sup> (以後,  $C^-, C^+$  と書く).

<sup>\*3</sup> 連続な確率変数  $X_i$  をその分布関数  $F_i$  で  $F_i(X_i)$  と変換すれば, もとの分布  $F_i$  に関係なく  $[0, 1]$  の一様分布に従うことはよく知られた事実である.

<sup>\*4</sup> これに対し, 通常のピアソンの相関係数が

$$\rho = \frac{\int \int (C(F_1(x_1), F_2(x_2)) - F_1(x_2)F_2(x_2)) dx_1 dx_2}{\sqrt{\int (x_1 - E[X_1])^2 dF_1(x_1)} \sqrt{\int (x_2 - E[X_2])^2 dF_2(x_2)}}$$

となり周辺分布の影響を受けてしまうことは非常に興味深い. なお共分散の表現は Hoeffding (1994) による (結果のみ載っているが部分積分を用いて簡単に導出できる).

<sup>\*5</sup>  $n$  次元においても

$$\max(u_1 + u_2 + \dots + u_n - n + 1, 0) \leq C(u_1, u_2, \dots, u_n) \leq \min(u_1, u_2, \dots, u_n)$$

が成立するが,  $n \geq 3$  のとき左辺はコピュラとはならないので注意が必要である (Cherubini et al. (2004) に具体例がある). その結果, 定理 4 の 1. も 3 次元以上では存在しなくなるが, 3 つ以上の変数において全て粗の (順位) 相関係数  $-1$  がありえないことを考えると, 直感的にも納得のいく事実であろう.

定理4 連続な確率変数  $(X_1, X_2)$  において, 以下が成立する.

1.  $(X_1, X_2)$  がコピュラ  $C^-$  を持つ  $\Leftrightarrow P(F_1(X_1) + F_2(X_2) = 1) = 1$   
 $\Leftrightarrow X_2$  はほとんど確実に  $X_1$  の狭義単調な減少関数である  
 $\Leftrightarrow \tau = -1$ .
2.  $(X_1, X_2)$  がコピュラ  $C^+$  を持つ  $\Leftrightarrow P(F_1(X_1) = F_2(X_2)) = 1$   
 $\Leftrightarrow X_2$  はほとんど確実に  $X_1$  の狭義単調な増加関数である  
 $\Leftrightarrow \tau = 1$ .
3.  $(X_1, X_2)$  がコピュラ  $C^\perp(u_1, u_2) = u_1 u_2$  を持つ  $\Leftrightarrow X_1, X_2$  は独立.

定理5 あるコピュラ  $C$  を与えたとき, 以下で定義される関数  $\bar{C}$  もコピュラとなる.

$$\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \quad (2.4)$$

これをサバイバルコピュラと呼ぶ.

$\bar{C}$  は  $(\frac{1}{2}, \frac{1}{2})$  について  $C$  を「対称に変換した」分布である.  $C$  に従う変数  $(U_1, U_2)$ ,  $\bar{C}$  に従う変数  $(\bar{U}_1, \bar{U}_2)$  とすると

$$\begin{aligned} P(U_1 > 1 - u_1, U_2 > 1 - u_2) &= u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2) \\ &= \bar{C}(u_1, u_2) = P(\bar{U}_1 \leq u_1, \bar{U}_2 \leq u_2), \end{aligned}$$

また  $C$  が絶対連続であるなら  $\bar{C}$  も絶対連続となり\*<sup>6</sup>, 密度関数  $\bar{c}$  も

$$\bar{c}(u_1, u_2) = c(1 - u_1, 1 - u_2) \quad (2.5)$$

と対称である.

### 2.3 コピュラのパラメトリック族

次に, 本論文で用いたパラメトリックコピュラを紹介する.\*<sup>7</sup>より詳細な解説は, Cherubini et al. (2004), Embrechts et al. (2001) を見よ. 特に, ケンドールの  $\tau$  の解析的な導出は後者を参照. 分布収束については付録 B.2 を参照.

#### (1) 正規コピュラ

Sklar の定理より, 二次元正規分布に対するコピュラが唯一存在する.  $\Phi_\rho$  を相関係数  $\rho$  ( $-1 \leq \rho \leq 1$ ) の二次元標準正規分布,  $\Phi$  を標準正規分布関数とすれば, (2.2) 式より

$$\begin{aligned} C_\rho^N(u_1, u_2) &= \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{2\rho st - s^2 - t^2}{2(1-\rho^2)}\right) ds dt. \end{aligned}$$

これを正規コピュラと呼ぶ. 直感的にも明らかのように,  $\rho$  は以下のような性質を持つ相関パラメータである.

- $\tau = \frac{2}{\pi} \arcsin \rho$ . \*<sup>8</sup>
- $\rho \rightarrow 1$  のとき  $C^+$ ,  $\rho \rightarrow -1$  のとき  $C^-$  にそれぞれ分布収束する. また  $\rho = 0$  のとき  $C^N = C^\perp$  である.

\*<sup>6</sup> (2.13) を積分してやれば簡単に示せる.

\*<sup>7</sup> その他のコピュラについては, Joe (1997), Nelsen (1999) などが詳しい.

\*<sup>8</sup> 二次元正規分布における  $\tau$  を計算 (例えば竹内 (1975) を参照) し, コピュラに対する  $\tau$  の一意性 (定理2) を用いて証明することも可能であろう.

(2) t コピュラ

同様に二次元 t 分布におけるコピュラを定義する。  $t_{\rho, \nu}$  を相関係数  $\rho$  ( $-1 \leq \rho \leq 1$ ), 自由度  $\nu$  ( $\geq 0$ ) の二次元標準 t 分布,  $t_\nu$  を自由度  $\nu$  ( $\geq 0$ ) の t 分布関数とし

$$\begin{aligned} C_{\rho, \nu}^t(u_1, u_2) &= t_{\rho, \nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} ds dt. \end{aligned}$$

これを t コピュラと呼ぶ\*9。正規コピュラとほぼ同様に,

- $\tau = \frac{2}{\pi} \arcsin \rho$ .
- $\rho \rightarrow 1$  のとき  $C^+$ ,  $\rho \rightarrow -1$  のとき  $C^-$  にそれぞれ分布収束する。ただし  $\rho = 0$  のとき  $C^\perp$  にはならない。

(3) ガンベル・コピュラ

$\beta \geq 1$  のとき, 以下で定義される  $C_\beta^{Gu}$  をガンベル・コピュラと呼ぶ。

$$C_\beta^{Gu}(u_1, u_2) = \exp\left(-\{(-\log u_1)^\beta + (-\log u_2)^\beta\}^{\frac{1}{\beta}}\right).$$

この  $\beta$  がガンベル・コピュラにおける相関パラメーターであり,

- $\tau = 1 - \beta^{-1}$ .
- $\beta \rightarrow \infty$  のとき  $C^+$  に分布収束し,  $\beta = 1$  のとき  $C^\perp$  となる。

(4) 生存ガンベル, 混合ガンベル・コピュラ

$-1 \leq \alpha \leq 1, \beta \geq 1$  としてガンベル・コピュラを以下のように拡張する。

$$\begin{aligned} C_{\alpha, \beta}^{MG}(u_1, u_2) &= \alpha C_\beta^{Gu}(u_1, u_2) + (1 - \alpha) \bar{C}_\beta^{Gu}(u_1, u_2) \\ &= \alpha \exp\left(-\{(-\log u_1)^\beta + (-\log u_2)^\beta\}^{\frac{1}{\beta}}\right) \\ &\quad + (1 - \alpha) \left\{u_1 + u_2 - 1 + \exp\left(-\{(-\log(1 - u_1))^\beta + (-\log(1 - u_2))^\beta\}^{\frac{1}{\beta}}\right)\right\}. \end{aligned}$$

これを混合ガンベル・コピュラと呼ぶことにする\*10。

- 任意の  $\alpha$  について,  $\beta \rightarrow \infty$  のとき  $C^+$  に分布収束し,  $\beta = 1$  のとき  $C^\perp$  となる。

また  $\bar{C}_\beta^{Gu}$  を生存ガンベル・コピュラと呼ぶことにする。生存ガンベルについては (3) における 2 つの性質がそのまま成立する (証明は容易)。

(補足) 正規コピュラと t コピュラ

相関係数  $\rho$  の二次元正規分布に内在する依存構造が「正規コピュラ」  $C_\rho^N$  であり, 相関係数  $\rho$ , 自由度  $\nu$  の二次元 t 分布に内在する依存構造が「t コピュラ」  $C_{\rho, \nu}^t$  である。

具体的には, 以下のような同値関係が成り立つ。

- $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  が平均  $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ , 共分散行列  $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix}$  の二次元正規分布に従う。

$$\Leftrightarrow \begin{cases} \text{周辺分布 } X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2) \\ \text{かつ, コピュラ } C_\rho^N \text{ を持つ。} \end{cases}$$

- $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$  が自由度  $\nu$ , 相関係数  $\rho$  の二次元標準 t 分布に従う。

$$\Leftrightarrow \begin{cases} \text{周辺分布 } X_1 \sim t_\nu, X_2 \sim t_\nu \\ \text{かつ, コピュラ } C_{\rho, \nu}^t \text{ を持つ。} \end{cases}$$

\*9 t コピュラに関する包括的な議論としては, Demarta and McNeil (2004) がある。

\*10 本来なら  $C = \alpha C_\beta^{Gu}(u_1, u_2) + (1 - \alpha) \bar{C}_\beta^{Gu}(u_1, u_2)$  を混合ガンベルと呼ぶべきであろうが, 分析・解釈の簡単化のため  $\beta = \gamma$  とした。Fantazzini (2004) では,  $C = \frac{1}{2} C_\beta^{Gu}(u_1, u_2) + \frac{1}{2} \bar{C}_\beta^{Gu}(u_1, u_2)$  を mixed gumbel としている。また, このようにして作られた関数がコピュラの定義を満たすことについては Nelsen (1999) Exercise 2.3 を参照。

特に t 分布の場合、2つの周辺分布の自由度とコピュラの自由度が全て等しくないと、二次元 t 分布とはならないので注意が必要であろう。逆に言えば、二次元 t 分布は周辺分布・コピュラの3つの自由度が全て等しいという制約を暗黙のうちに課してしまっていることになる。2.4でも述べる通り、二次元 t 分布ではなく t コピュラを用いることのメリット（の一部）はこの点にある。

## 2.4 コピュラ概念を導入するメリット

多変量同時分布を周辺分布と依存構造（コピュラ）に完全に分解できるようになった結果…

モデリングの柔軟性が格段に上昇した。単純に考えて、周辺分布の種類×コピュラの種類だけ分布を作ることが可能である。また、このようにして作った [周辺分布×依存構造] の分布は扱いやすく直感的理解もしやすいであろう。分析可能な分布の範囲は大きく広がったと言える。

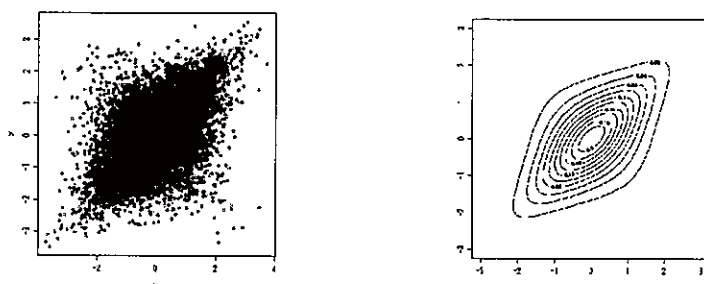
例えば「2次元 t 分布」「2次元正規分布」は良いとしても、「周辺分布が標準正規分布で、依存構造が自由度3の2次元 t 分布と同じ」などという分布は従来の枠組みではなかなか扱うことが出来なかった。しかしコピュラと Sklar の定理を用いることで、その分布関数も

$$F(x_1, x_2) = C_{\rho, 3}^t(\Phi(x_1), \Phi(x_2)) \quad (2.6)$$

と簡単に、直感的にもわかりやすく記述することが出来る。

ちなみにこの場合、周辺の正規分布も「t 分布」とみなすことは可能であるが、t コピュラと自由度が異なっているため (2.6) はもはや2次元 t 分布のような綺麗な（楕円形の）形状をしていない（図1. 散布図と等高線プロット.）。この点については Demarta and McNeil (2004) も参照。

(図1)



また本論文で対象とする、非正規性や fat tail という金融データへの応用を考えると、二変量正規分布に対する批判として“周辺分布の非正規性”と“依存構造の非正規性”を別々に扱い、別々に検証することができるのは大きなメリットである。例えばあるポートフォリオの収益率分布が（正規分布と比較して）fat tail 性を持ってしまうということは、

1. 各資産の周辺分布自体が fat tail 性を持っている
2. 各資産の「裾における依存関係」が多変量正規分布に比べて強くなってしまっている

という2つの理由（のみ）が考えられるはずで、この2番目を分析する際に、2.3で導入したパラメトリックコピュラが強力な武器となるのである。

正規コピュラに比べ「裾における依存関係」がより強いコピュラを使えば、このような分布をモデリングすることも十分可能となる。次節ではこの「裾における依存」について説明する。

### 3 分布の裾における依存構造

#### 3.1 裾における確率

各コピュラの裾における依存構造を見るため、 $U_1, U_2$  がともにある一定の閾値  $v$  を超える確率  $P(U_1 > v, U_2 > v)$  と、超えない確率  $P(U_1 \leq v, U_2 \leq v)$  (7 ページ図 2 の斜線部) が  $v$  によってどのように推移するかを調べる。正規コピュラ、 $t$  コピュラについては 100 万個の乱数を発生させ近似的な計算を行った。その結果が図 3 である (正規コピュラと  $t$  コピュラは左右対称なので上側のみを示した)。どのコピュラであっても最終的には 0 に向かうが、その収束する速さには大きな違いがあるように見える。裾における依存関係が強いほど、なかなか値が 0 に収束しないと考えることができるだろう。

例えば、図 3 の一番下にある各コピュラの比較を見れば、「右上の裾」ではガンベル、 $t$  (自由度 3)、正規の順に裾の依存関係は弱くなり、「左下の裾」では  $t$  (自由度 3)、正規、ガンベルの順に弱くなると判断することができる。

なお相関の基準として用いたケンドールの  $\tau$  について、その根拠となるいくつかの望ましい性質については Cherubini et al. (2004) を参照。

#### 3.2 Tail Dependence

ここで以下の概念を定義する (詳細は Joe (1997), Coles (2001) を見よ)

**定義** 次式 (3.1), (3.2) を満たす  $\lambda_U, \lambda_L$  を、それぞれ上側 Tail Dependence, 下側 Tail Dependence と呼ぶ。

$$\lambda_U = \lim_{v \uparrow 1} P(F_1(X_1) > v | F_2(X_2) > v) = \lim_{v \uparrow 1} P(F_2(X_2) > v | F_1(X_1) > v) \quad (3.1)$$

$$\lambda_L = \lim_{v \downarrow 0} P(F_1(X_1) \leq v | F_2(X_2) \leq v) = \lim_{v \downarrow 0} P(F_2(X_2) \leq v | F_1(X_1) \leq v). \quad (3.2)$$

特に  $(X_1, X_2)$  のコピュラを  $C$  とし、 $U_1 = F_1(X_1)$ ,  $U_2 = F_2(X_2)$  と置けば、

$$\lambda_U = \lim_{v \uparrow 1} P(U_1 > v | U_2 > v) = \lim_{v \uparrow 1} P(U_2 > v | U_1 > v) \quad (3.3)$$

$$\lambda_L = \lim_{v \downarrow 0} P(U_1 \leq v | U_2 \leq v) = \lim_{v \downarrow 0} P(U_2 \leq v | U_1 \leq v) \quad (3.4)$$

となり、コピュラを用いて

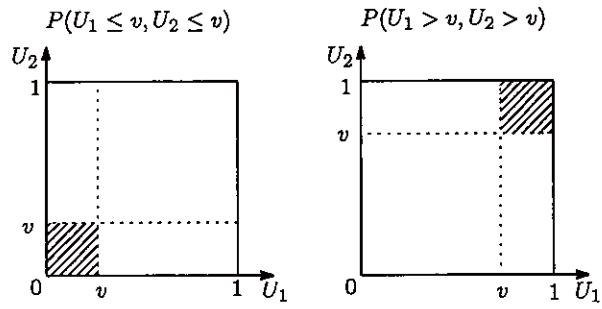
$$\lambda_U = \lim_{v \uparrow 1} \frac{P(U_1 > v, U_2 > v)}{1 - v} = \lim_{v \uparrow 1} \frac{1 - 2v + C(v, v)}{1 - v} \quad (3.5)$$

$$\lambda_L = \lim_{v \downarrow 0} \frac{P(U_1 \leq v, U_2 \leq v)}{v} = \lim_{v \downarrow 0} \frac{C(v, v)}{v} \quad (3.6)$$

となる。

すなわち  $\lambda_U, \lambda_L$  は、3.1 で扱った「閾値  $v$  を超える確率」が、どの程度の速さで 0 に収束するかを数値化した 1 つの指標であると思なすことができる。また、(3.5)(3.6) から明らかなように、周辺分布の影響を受けずコピュラに対して一意に定まる概念であり、2 変量分布の裾における依存構造を把握する際の重要なツールとなる。

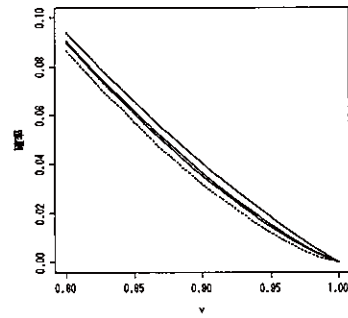
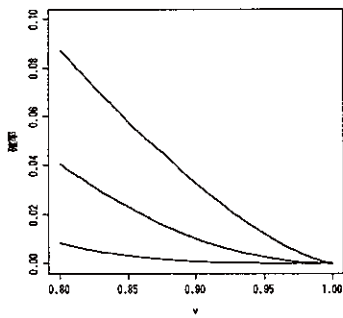
(図2)



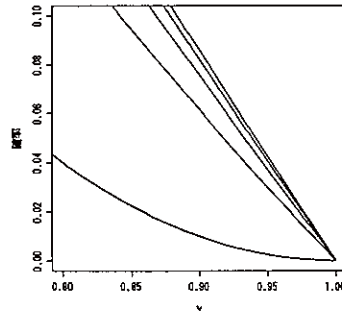
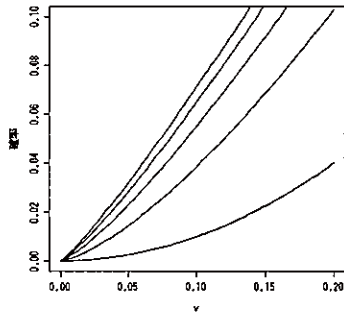
(図3)

正規コピュラの  $\rho$  による違い (上から順に 0.5, 0, -0.5)

tコピュラの自由度  $\nu$  による違い (上から順に 3, 6, 9,  $\infty$ )



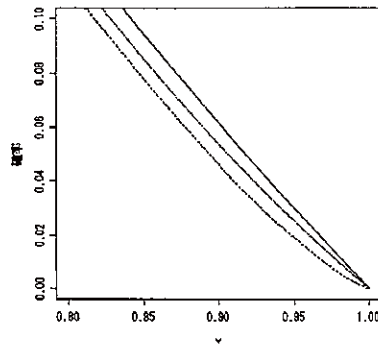
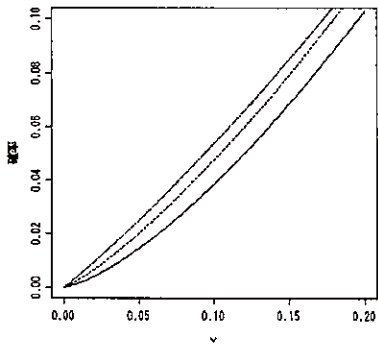
ガンベル・コピュラの  $\beta$  による違い (両側とも上から順に 5, 4, 3, 2, 1)



正規コピュラ, tコピュラ, ガンベル・コピュラの比較 ( $\tau = 0.5$  で固定, tコピュラの自由度 3)

上から順に t, 正規, ガンベル

上から順にガンベル, t, 正規



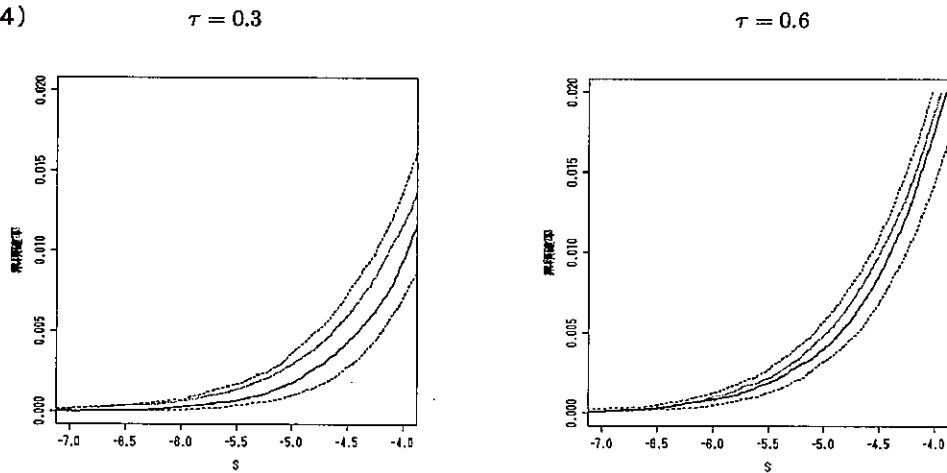
(例)

$(X_1, X_2)$  の裾における依存構造が、和の分布  $S = X_1 + X_2$  の裾にどのような影響を及ぼすか実際に検証する。正規コピュラ, tコピュラ (自由度3), ガンベル・コピュラ, 生存ガンベル・コピュラの4つのコピュラについて10万個の乱数を発生させ、分布の下側の裾を描写した (図4)。なおコピュラの上側の裾における影響だけを見るため、全体の相関は  $\tau = 0.3$ ,  $\tau = 0.6$  と統一し、周辺分布には標準正規分布を仮定した (つまり正規コピュラの場合は二変量標準正規分布に従う  $(X_1, X_2)$  の和の分布となる)。詳しくは次節の表にまとめるが、このときの下側 Tail Dependence は

	正規	t	ガンベル	生存ガンベル
$\tau = 0.3$	0	0.2876	0	0.3755
$\tau = 0.6$	0	0.5512	0	0.6805

である。

(図4)



図は上から順に生存ガンベル, t, 正規, ガンベルである。結果を見ると、予想通り裾における依存構造の違いが和の分布の fat tail となって表れている。また

- 相関が強いと、相対的に Tail Dependence の差異による影響が小さくなり、裾の違いも小さくなる。
  - ガンベル・コピュラ, 正規コピュラともに下側 Tail Dependence は0だが、厳密にはその裾に差がある。
- すなわち  $\lambda$  も裾の依存構造における一つの側面を反映しているに過ぎない。

という事実も同時に確認された。なお、2つ目の点については、Coles (2001), 山井・吉羽 (2002) の議論とも整合している。

## 4 各パラメトリック・コピュラの性質 (まとめ)

以上これまでの性質を表1にまとめ、各コピュラの散布図を付録C (本論文末尾) に示した。Tail Dependence の計算は比較的容易なので省略する (例えば Embrechts et al. (2001) を参照)。

(表1)

コピュラ	パラメータ	ケンドール $\tau$	上側 Tail Dependence	下側 Tail Dependence	特殊ケース
正規	$-1 \leq \rho \leq 1$	$\frac{2}{\pi} \arcsin \rho$	0	0	$C^-, C^\perp, C^+$
t	$-1 \leq \rho \leq 1, \nu > 0$	$\frac{2}{\pi} \arcsin \rho$	$2 t_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}} \right)$	$2 t_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{(1+\rho)}} \right)$	$C^-, C^+$
ガンベル	$\beta \geq 1$	$1 - \frac{1}{\beta}$	$2 - 2^{1/\beta}$	0	$C^\perp, C^+$
生存ガンベル	$\beta \geq 1$	$1 - \frac{1}{\beta}$	0	$2 - 2^{1/\beta}$	$C^\perp, C^+$
混合ガンベル	$-1 \leq \alpha \leq 1, \beta \geq 1$	$1 - \frac{1}{\beta} ?$	$\alpha(2 - 2^{1/\beta})$	$(1 - \alpha)(2 - 2^{1/\beta})$	$C^\perp, C^+$

注: 混合ガンベルの  $\tau$  は変数を様々な値にシミュレーションし、 $1 - \frac{1}{\beta}$  にほぼ等しいことが確認されたが、厳密には異なると思われる。

明らかではあるが各コピュラ間の関係として、

- 混合ガンベル  $C_{\alpha,\beta}^{MG}$  は、 $\alpha = 1, -1$  でそれぞれガンベル  $C_{\beta}^{G^u}$ 、生存ガンベル  $\bar{C}_{\beta}^{G^u}$  となる。
- tコピュラ  $C_{\rho,\nu}^t$  は  $\nu \rightarrow \infty$  で正規コピュラ  $C_{\rho}^N$  に分布収束する\*11。

が成り立つ。

表1、付録Cを見ればわかる通り、(周辺分布ではなく)依存構造に正規性を仮定する問題点としては、大きく次の2つが考えられる。

- (1) Tail Dependence が両側とも0である。
- (2)  $C = \bar{C}$  という性質が成り立ち\*12、Tail Dependence も含め、依存構造に強い対称性を仮定してしまっている。

特に(1)は金融リスク管理において、あまりにも楽観的すぎると言わざるをえないだろう。(2)も実際の分布にこのような強い仮定があてはまるという保証はなく、例えば株価の動きなど、下方的ショックに対する価格の相関が通常より強いというのは経験的にもよく知られた事実である。

これに対する比較とし、次の実証分析では主にtコピュラ、混合ガンベル・コピュラを利用することになる。

tコピュラを用いるメリットとして、正規コピュラと同様の性質を持つ相関パラメータ  $\rho$  を備えながら、自由度  $\nu$  によって裾の依存構造を(直感的に)調節することができることがあげられる。正規コピュラの手堅さを持たせ一般化したモデルであるとみなすことが可能で、解釈のしやすさ、正規コピュラとの比較のしやすさという点でtコピュラは大きな強みを持っているだろう。

単純に考えて、推定したtコピュラの自由度が十分小さければ、少なくとも正規コピュラの仮定は否定することができるはずである。\*13実証的にも、Mashal and Zeevi (2002), Schmidt (2004) など、金融データに対してはtコピュラを用いた推定をする方がより望ましいということが主張されている。ただしこのtコピュラも、(2)について正規コピュラと同様の問題点を抱えているということは注意しなければならない。

一方の混合ガンベル・コピュラは、非対称かつTail Dependenceを持った依存構造であり、(2),(1)両方に対してある程度の改善が予想される。左右のバランスを示す  $\alpha$  が  $\frac{1}{2}$  からズレることで、より非対称なコピュラとなる。

しかし当然のことながら(パラメータは2個なので)、非対称性を許容したぶん全体の相関  $\tau$  とTail Dependence (の和)に一定の関係が存在してしまっており、一長一短である。むしろ正規コピュラとネストで無い分、従来型との比較と言う点でtコピュラよりもやや扱いにくいと言えるかもしれない。また本論文では特に意識する必要が無かったが、負の相関を全く捉えることができないという点も(混合)ガンベル・コピュラの短所であろう。

次節ではこれまでに解説したコピュラ概念と各パラメトリック・コピュラの性質をふまえ、二次元正規分布に代わるモデルとして現実のデータへの応用とその推定法を見る。

## 5 実証分析

### 5.1 データ

金融データへの応用という観点から、為替レートの収益率分布を例として分析する。日本円に換算した各国通貨価値(1ドル120円、1ポンド200円など)の週次対数変化率\*14を各変数とし、全てのサンプルは独立同一分布に従うと仮定した。

期間は1990年1月5日から2005年12月30日(データ数834)。対象通貨は、米ドル(USD)、オーストラリアド

\*11 ガンマ関数が出てこないため、1次元におけるt分布→正規分布の収束よりも証明は簡単である。

\*12 Nelsen (1999) 2.7節を見よ。

\*13 ただし厳密に  $H_0: \nu = \infty$  という帰無仮説を立てることは不可能であり、 $H_0: \nu > 1000$  などと  $H_0$  を恣意的に決めざるをえない。Mashal and Zeevi (2002) を参照。

\*14 データはOANDA.comから取得し、各金曜日を実際のデータとした。データを週次としたのは、一般的にも知られている通り、日次データのボラティリティに曜日効果が大きく(あるいは相関構造自体に曜日効果が存在するという点も考えられる。今後の研究課題としたい)、日次の場合、収益率分布の正確な推定がかなり複雑になってしまうと考えられるためである。



ル (AUD), 英ポンド (GBP), スイスフラン (CHF), 独マルク (DEM), ユーロ (EUR) の 6 種で, (USD, AUD) など 2 種類ごとの同時分布に対するコピュラを推定する。なお, 独マルク, ユーロについては通貨統合の影響により, それぞれ 1999 年 1 月 1 日まで, 1999 年 1 月 8 日以降を分析の対象とした。

## 5.2 通常の最尤法による推定

最尤法を用いて, 各コピュラのパラメーター推定を行う。(2.1) より, 2 次元同時分布の対数尤度関数は

$$\sum_i \log f(x_{i1}, x_{i2}) = \sum_i \log(c(F_1(x_{i1}), F_2(x_{i2}))) + \sum_i \log f_1(x_{i1}) + \sum_i \log f_2(x_{i2}) \quad (5.1)$$

となり, これを周辺分布・コピュラの全パラメーターについて最大化すればよい<sup>\*15</sup>。コピュラの最尤推定値を求めるためには周辺分布の特定化を行わなければならないが, 今回は分布が左右対称に近いと考え, 以下の 3 種類

$$\text{正規分布 } f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

$$\text{ロジスティック分布 } f(x) = \frac{\exp\left(\frac{x-\mu}{\phi}\right)}{\phi(1 + \exp\left(\frac{x-\mu}{\phi}\right))^2}$$

$$\text{ラプラス分布 } f(x) = \frac{1}{2\phi} \exp\left(-\frac{|x - \mu|}{\phi}\right)$$

を仮定した。<sup>\*16</sup>ロジスティック分布 (尖度 4.2), ラプラス分布 (尖度 6) は正規分布に比べて裾の厚い形状をしており, 周辺分布の fat tail 性にもある程度の対応が可能である。周辺分布が真の分布に近ければ, そのぶん変換したデータ  $F_1(x_1), F_2(x_2)$  も真のコピュラに近い分布をすれば, その結果コピュラの推定値もより真の値に近いものになると考えられる。2.4 でも述べたが, このようなモデリングの柔軟性こそがコピュラ概念を用いる最大のメリットである。

またコピュラは正規, t, 混合ガンベルの 3 種類を選択した (密度関数は付録 B.1 を見よ)。以上の [周辺分布×コピュラ] の組み合わせについて, ①米ドル・オーストラリアドル, ②英ポンド・スイスフラン, ③米ドル・独マルク (99 年 1 月 1 日まで), ④米ドル・ユーロ (99 年 1 月 8 日以降) の各 2 次元同時分布を推定したものが表 2 である。各通貨 1 行目と 2 行目が [周辺分布×コピュラ] の組み合わせを表しており, 特に 1 番左の列 [正規×正規] は, 通常よく用いられる二変量正規分布と同じモデルである。

モデルの簡易的な評価基準として AIC, BIC を仮定すると (最小の値は太字にしてある), それぞれ最もフィットの良い組み合わせは, ① [ロジスティック×混合ガンベル], ② [ロジスティック×t], ③ [ロジスティック×t], ④ [ロジスティック×正規] となる。以下この結果をもう少し詳しく見ていく。

### (1) 正規コピュラのフィット

まず④については, t コピュラの自由度が収束しないなど正規コピュラがデータに最もよくフィットしている。(マーケットの安定した) ある一定の期間だけに限定すれば, 正規型の依存構造という仮定もそこまで間違ったものではないのかもしれない。しかしその他の①~③においては, どの周辺分布を仮定しても正規コピュラのフィットが最も劣っており, t コピュラとの比較として自由度  $\nu$  の値を見ても十分小さい値をとっていることがわかる。<sup>\*17</sup> やはり多くのデータに関して, 裾の依存関係は正規コピュラよりもはるかに強いとすることができるだろう。t コピュラを使うことのメリットが早くも実証されたことになる。

<sup>\*15</sup> ただし密度関数が複雑でパラメータ数も多いことから, 最大化する際の計算負荷はかなり大きくなる。より簡便かつ現実的な方法としては, 第 2,3 項 (周辺分布) を先に最大化し, その結果を所与としてから第 1 項を最大化する「IFM 法」がある。詳しい方法, 推定量の性質などは Joe (1997) を参照。

<sup>\*16</sup> これらの詳しい性質は養谷 (2003) を参照。

<sup>\*17</sup> たとえば①の [正規×t] の自由度  $\nu$  における 99% 漸近的信頼区間は (2.096, 14.548) となり, 「十分小さい」とみなせる。

(表2) 各周辺分布を仮定したときの推定結果

- ・推定値下のカッコ内は漸近的標準誤差 (Hessian から計算した)。
- ・ $\mu_1, \mu_2, \sigma_1, \sigma_2, \phi_1, \phi_2$  とその標準誤差の単位はパーセントである。
- ・Kendall's  $\tau$ , 上側 Tail Dependence (上側 T.D.), 下側 Tail Dependence (下側 T.D.) は, 推定したパラメータから表1の結果を用いて計算したもの。サンプルから直接計算した推定量  $\hat{\tau}$  がそれぞれ1番上に載せてある。

① USD,AUD ( $\hat{\tau} = 0.444$ )

周辺分布	正規分布			ロジスティック分布			ラプラス分布		
	正規	t	混合ガンベル	正規	t	混合ガンベル	正規	t	混合ガンベル
コピュラ									
$\mu_1$	-0.0106 (0.0226)	-0.0168 (0.0226)	-0.0170 (0.0227)	0.0256 (0.0205)	0.0212 (0.0204)	0.0184 (0.0206)	0.0441 (0.0194)	0.0398 (0.0193)	0.0376 (0.0195)
$\mu_2$	-0.0146 (0.0268)	-0.0188 (0.0271)	-0.0106 (0.0274)	0.0343 (0.0252)	0.0346 (0.0250)	0.0323 (0.0252)	0.0604 (0.0232)	0.0613 (0.0231)	0.0580 (0.0233)
$\sigma_1, \phi_1$	0.6521 (0.0138)	0.6563 (0.0140)	0.6592 (0.0139)	0.3452 (0.0068)	0.3441 (0.0067)	0.3459 (0.0069)	0.4852 (0.0142)	0.4833 (0.0133)	0.4815 (0.0135)
$\sigma_2, \phi_2$	0.7746 (0.0170)	0.7909 (0.0180)	0.7952 (0.0175)	0.4239 (0.0095)	0.4233 (0.0093)	0.4251 (0.0093)	0.5938 (0.0183)	0.5936 (0.0173)	0.5907 (0.0173)
コピュラ 推定値	$\rho: 0.650$ (0.020)	$\rho: 0.676$ (0.022) $\nu: 8.322$ (2.417)	$\alpha: 0.694$ (0.079) $\beta: 1.988$ (0.069)	$\rho: 0.647$ (0.019)	$\rho: 0.647$ (0.022) $\nu: 5.699$ (1.618)	$\alpha: 0.656$ (0.084) $\beta: 1.871$ (0.062)	$\rho: 0.654$ (0.020) $\nu: 3.126$ (0.648)	$\rho: 0.644$ (0.025) $\alpha: 0.652$ (0.098)	$\beta: 1.874$ (0.063)
Kendall's $\tau$	0.451	0.473	0.497	0.448	0.448	0.465	0.454	0.446	0.466
上側 T.D.	0	0.211	0.405	0	0.271	0.362	0	0.397	0.360
下側 T.D.	0	0.211	0.178	0	0.271	0.190	0	0.397	0.192
対数尤度	6113.349	6118.453	6135.169	6171.705	6181.499	6186.976	6141.932	6160.399	6160.430
AIC	-12224.70	-12232.91	-12266.34	-12341.41	-12359.00	<b>-12369.95</b>	-12281.86	-12316.80	-12316.86
BIC	-12223.78	-12231.06	-12264.50	-12340.49	-12357.16	<b>-12368.11</b>	-12280.94	-12314.96	-12315.02

② GBP,CHF ( $\hat{\tau} = 0.509$ )

周辺分布	正規分布			ロジスティック分布			ラプラス分布		
	正規	t	混合ガンベル	正規	t	混合ガンベル	正規	t	混合ガンベル
コピュラ									
$\mu_1$	-0.0079 (0.0245)	-0.0054 (0.0245)	-0.0056 (0.0258)	0.0197 (0.0222)	0.0224 (0.0221)	0.0231 (0.0222)	0.0633 (0.0209)	0.0693 (0.0208)	0.0715 (0.0211)
$\mu_2$	-0.0025 (0.0243)	-0.0069 (0.0241)	-0.0063 (0.0250)	0.0218 (0.0229)	0.0165 (0.0228)	0.0164 (0.0226)	0.0635 (0.0211)	0.0613 (0.0209)	0.0633 (0.0211)
$\sigma_1, \phi_1$	0.7084 (0.0150)	0.7123 (0.0151)	0.7473 (0.0165)	0.3748 (0.0076)	0.3735 (0.0075)	0.3757 (0.0077)	0.5278 (0.0155)	0.5231 (0.0144)	0.5218 (0.0147)
$\sigma_2, \phi_2$	0.7022 (0.0148)	0.7013 (0.0145)	0.7150 (0.0154)	0.3837 (0.0079)	0.3819 (0.0076)	0.3801 (0.0078)	0.5408 (0.0159)	0.5339 (0.0147)	0.5298 (0.0147)
コピュラ 推定値	$\rho: 0.732$ (0.016)	$\rho: 0.743$ (0.017) $\nu: 10.261$ (3.153)	$\alpha: 0.548$ (0.078) $\beta: 2.342$ (0.087)	$\rho: 0.720$ (0.016)	$\rho: 0.721$ (0.018) $\nu: 6.515$ (1.686)	$\alpha: 0.489$ (0.085) $\beta: 2.050$ (0.069)	$\rho: 0.729$ (0.016) $\nu: 3.675$ (0.783)	$\rho: 0.718$ (0.020) $\alpha: 0.436$ (0.098)	$\beta: 2.059$ (0.070)
Kendall's $\tau$	0.522	0.533	0.573	0.512	0.513	0.512	0.520	0.510	0.514
上側 T.D.	0	0.223	0.359	0	0.304	0.292	0	0.424	0.262
下側 T.D.	0	0.223	0.296	0	0.304	0.305	0	0.424	0.338
対数尤度	6216.495	6231.601	6225.583	6266.732	6279.763	6277.827	6236.672	6256.747	6253.762
AIC	-12430.99	-12459.20	-12447.17	-12531.46	<b>-12555.53</b>	-12551.65	-12471.34	-12509.49	-12503.52
BIC	-12430.07	-12457.36	-12445.32	-12530.54	<b>-12553.68</b>	-12549.81	-12470.42	-12507.65	-12501.68

(つづく)

(表2つづき)

③ USD,DEM (99年1月1日以前) ( $\hat{\tau} = 0.338$ )

周辺分布	正規分布			ロジスティック分布			ラプラス分布		
	正規	t	混合ガンベル*	正規	t	混合ガンベル	正規	t	混合ガンベル
$\mu_1$	-0.0218 (0.0324)	-0.0288 (0.0310)	-0.0314	0.0198 (0.0285)	0.0095 (0.0282)	0.0126 (0.0284)	0.0533 (0.0271)	0.0413 (0.0269)	0.0468 (0.0271)
$\mu_2$	-0.0210 (0.0310)	-0.0119 (0.0293)	-0.0169	0.0163 (0.0285)	0.0242 (0.0274)	0.0225 (0.0276)	0.0567 (0.0273)	0.0619 (0.0262)	0.0592 (0.0267)
$\sigma_1, \phi_1$	0.7015 (0.0203)	0.6867 (0.0186)	0.7268	0.3606 (0.0102)	0.3595 (0.0099)	0.3598 (0.0100)	0.5029 (0.0202)	0.5052 (0.0192)	0.4992 (0.0190)
$\sigma_2, \phi_2$	0.6714 (0.0192)	0.6505 (0.0172)	0.6772	0.3570 (0.0099)	0.3484 (0.0090)	0.3453 (0.0091)	0.5004 (0.0200)	0.4891 (0.0178)	0.4820 (0.0178)
コピュラ 推定値	$\rho : 0.618$ (0.028)	$\rho : 0.570$ (0.038) $\nu : 5.767$ (1.823)	$\alpha : 0.620$ $\beta : 1.715$	$\rho : 0.556$ (0.031)	$\rho : 0.523$ (0.039) $\nu : 4.522$ (1.308)	$\alpha : 0.542$ (0.141) $\beta : 1.555$ (0.065)	$\rho : 0.553$ (0.033) $\nu : 2.642$ (0.618)	$\rho : 0.502$ (0.043) $\alpha : 0.479$ (0.163) $\beta : 1.561$ (0.066)	
Kendall's $\tau$	0.424	0.386	0.417	0.375	0.350	0.357	0.373	0.335	0.359
上側 T.D.	0	0.217	0.311	0	0.240	0.237	0	0.339	0.211
下側 T.D.	0	0.217	0.191	0	0.240	0.201	0	0.339	0.230
対数尤度	3454.854	3469.691	3466.969	3491.532	3502.130	3501.508	3473.829	3487.656	3484.704
AIC	-6907.71	-6935.38	-6929.94	-6981.06	<b>-7000.26</b>	-6999.02	-6945.66	-6971.31	-6965.41
BIC	-6906.79	-6933.54	-6928.10	-6980.14	<b>-6998.42</b>	-6997.17	-6944.74	-6969.47	-6963.57

④ USD,EUR (99年1月8日以降) \*\* ( $\hat{\tau} = 0.345$ )

周辺分布	正規分布		ロジスティック分布		ラプラス分布	
	正規	混合ガンベル	正規	混合ガンベル	正規	混合ガンベル
$\mu_1$	0.0042 (0.0305)	0.0077 (0.0311)	0.0203 (0.0301)	0.024 (0.0300)	0.0227 (0.0289)	0.0376 (0.0290)
$\mu_1$	0.0058 (0.0362)	0.0008 (0.0368)	0.0292 (0.0346)	0.0268 (0.0347)	0.0607 (0.0342)	0.0749 (0.0344)
$\sigma_1, \phi_1$	0.5822 (0.0183)	0.5917 (0.0196)	0.3301 (0.0098)	0.3294 (0.0098)	0.4630 (0.0205)	0.4568 (0.0197)
$\sigma_1, \phi_1$	0.6914 (0.0227)	0.6998 (0.0237)	0.3835 (0.0128)	0.3843 (0.0127)	0.5358 (0.0249)	0.5362 (0.0242)
コピュラ 推定値	$\rho : 0.552$ (0.036)	$\alpha : 0.512$ (0.175) $\beta : 1.531$ (0.077)	$\rho : 0.548$ (0.036)	$\alpha : 0.421$ (0.198) $\beta : 1.495$ (0.069)	$\rho : 0.549$ (0.038)	$\alpha : 0.329$ (0.235) $\beta : 1.502$ (0.069)
Kendall's $\tau$	0.372	0.347	0.369	0.331	0.370	0.334
上側 T.D.	0	0.219	0	0.173	0	0.136
下側 T.D.	0	0.209	0	0.238	0	0.278
対数尤度	2724.352	2714.012	2726.542	2720.294	2705.054	2702.467
AIC	-5446.70	-5424.02	<b>-5451.08</b>	-5436.59	-5408.11	-5400.93
BIC	-5445.78	-5422.18	<b>-5450.16</b>	-5434.75	-5407.19	-5399.09

(注)

それぞれの対数尤度関数を書いたのち、Rのoptimを用いて最大化を行った。  
パラメータ数が多いぶん局所的最適解も無数にあり、計算結果には若干の誤差があると思われる。

\* ③の[正規×混合ガンベル]の組み合わせは、(制約条件を満たすように)パラメータを変数変換した上での最適解しか求まらなかった。そのため標準誤差は面倒なので計算していない。

\*\* ④は各tコピュラの自由度が収束しなかった。

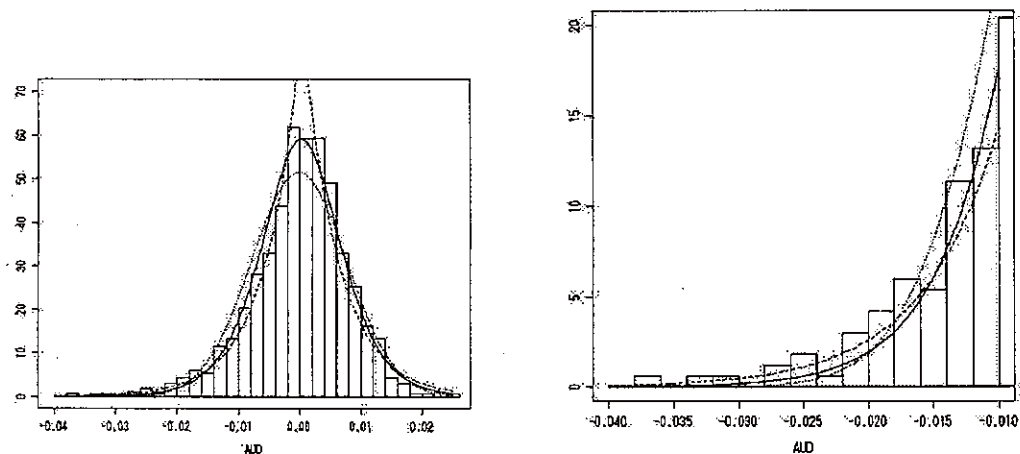
(2) 周辺分布のフィット

周辺分布についても、モデル選択基準に照らせば、正規分布を仮定したモデルのフィットがどれも相対的に芳しくない。ロジスティック分布を使ったモデルが最も優れているという結論を得ることができる。すなわち2変量正規分布を仮定することは、周辺分布とコピュラの2重の意味で望ましくないということになる。

また周辺分布については、仮定した分布の違いによって各コピュラの推定値がバラついてしまったということも注意しなければならない。データ全体の相関とリンクした変数  $\rho, \beta$  はかなり安定しているが、 $\nu, \alpha$  といった分析上重要な変数がブレてしまっている。例えば①のtコピュラにおける自由度は8.322, 5.699, 3.126となっており、Tail Dependenceの値もそのまま変化しているのがわかる。つまりコピュラのパラメータを正しく推定できるかどうかは、周辺分布のフィットにかなり依存してしまっているといえる。<sup>\*18</sup>

試しにオーストラリアドルの分布を例に、ヒストグラムと推定した周辺分布の密度関数を重ねてみる(図5)。対数尤度を元にしたモデル選択を行った結果、確かに全体的なロジスティック分布のあてはまりは良いが、左裾についてはラプラス分布よりも精度が悪くなっていることがわかる。この結果ロジスティック分布を用いた場合、(実際よりも左裾が薄い分布を用いて  $\hat{F}(X)$  とデータを変換してしまうため)コピュラの左下の裾におけるデータの信頼性も悪く、裾における依存構造の推定があまり正確でないことが予想される。またラプラス分布も左裾についてのフィットは良いが、右裾を含めた全体のあてはまりがロジスティック分布よりも劣ってしまっている。

(図5)



左端における裾の厚い順にラプラス、ロジスティック(実線)、正規分布

このように周辺分布のモデリングミスは極力避けなければならない問題である一方、パラメトリックなモデルで改善していくのは困難かつ際限が無い。例えば今回の場合なら、一般化パレート分布など裾の厚い分布を用いる、左右の歪んだ分布を用いる、2つの周辺分布に異なったパラメトリック族を仮定できるようにするなど、様々なモデルを考えねばならず、計算もますます複雑となることが予想される。

そのため、特に興味の対象として依存構造の分析に重点をおく場合、次に示すようにノンパラメトリックな周辺分布を用いるのが直感的にもベストである。

<sup>\*18</sup> 正確には、真の周辺分布と推定した周辺分布の違いにより、コピュラの推定量にバイアスが生まれてしまうということである。具体例としては Durrleman et al.(2001)を参照。現実には真の分布は未知であるため、観測されたデータに対するフィットの良さを、真の分布に対するフィットの良さと「思っ」て代替的に用いることになる。

### 5.3 周辺分布に経験分布関数を用いた推定

ここでは周辺分布にパラメトリックな仮定をおかず、経験分布関数  $\hat{F}$  を用いて最尤推定することを考える。 $\hat{U}_i = \hat{F}_i(X_i)$  とデータを変換し、この  $(\hat{U}_1, \hat{U}_2)$  のサンプルを用いて  $\sum_i \log c(\hat{u}_{i1}, \hat{u}_{i2})$  を最大化する。ただし正規コピュラ・tコピュラはサンプルに1という値が出ると最尤推定できないので(B.1の密度関数式を見よ)、通常  $n$  で割るところを  $n+1$  とし、経験分布関数は  $\hat{F}(x) = \frac{1}{n+1} \#\{i: 1 \leq i \leq n, x_i \leq x\}$  を用いることにする。 $\hat{F}$  が漸近的に  $F$  に収束することを考えれば、5.2 で見たような周辺分布におけるモデル選択の誤りがない分、最尤推定値も真の値に近いものとなるだろう。<sup>\*19</sup>

5.2 からさらに⑤米ドル・英ポンド、⑥英ポンド・オーストラリアドル、⑦米ドル・スイスフラン、⑧オーストラリアドル・スイスフランの組合せを追加し、コピュラも(通常の)ガンベル、生存ガンベルを新たに加えて推定した。その結果が表3である。前と同様、最小のAIC,BICは太字にしてある。

#### (1) 周辺分布の影響、表2との比較(①~④)

tコピュラの自由度  $\nu$  はロジスティック分布を仮定した場合とかなり近くなったが、一方で混合ガンベルの非対称性を表す  $\alpha$  の値は大きく変化した。相対的には  $\alpha$  はラプラス分布を用いた場合に近いが、ラプラス分布に比べれば今回の  $\nu$  の値はやや大きい。やはりラプラス、ロジスティックとも周辺分布の構造全てを捉え切れていない分、実際のデータを100%反映した場合に比べズレが生じているだろう。<sup>\*20</sup> また正規分布を仮定したモデルは  $\alpha, \nu$  とも最も遠い値になってしまった。

結局どの周辺分布を仮定しても完璧ではない。やはり安定して良い推定を行うためには経験分布関数を用いる方がベストであり簡単である。

#### (2) 正規コピュラのフィット

④と⑥についてAIC,BICが最小の値となった。しかしその他の組合せについては良い結果が出ておらず、tコピュラの自由度を見ても十分に小さい値となっていることがわかる。表2とほぼ同様の結果であり、やはり正規型依存構造の仮定は楽観的すぎると考えられる。

#### (3) tコピュラと混合ガンベル・コピュラ

混合ガンベルは  $\alpha$  が大きく0.5を下回り、Tail Dependenceの値も相対的に下側に大きくなった。これはデータの依存構造(この場合  $\hat{U}_1, \hat{U}_2$  の分布)に非対称性があるということを反映しており、その分、対称型のtコピュラについてはフィットが悪くなっているだろう。現実の依存構造に非対称性があるぶん、tコピュラよりも(もちろん正規コピュラよりも)混合ガンベル・コピュラの方が望ましい場合があることが確認された。

また④⑥の場合に限ってtコピュラの方がフィットは良いが、これはtコピュラが裾の依存  $\nu$  を全体の相関  $\rho$  と分けて扱えるため、Tail Dependenceが小さい場合も  $\nu$  を大きくすることで十分対応できたからであろう。このため、①~⑧全体について言えば、総じてtコピュラの方が混合ガンベルよりも「安定して」良い結果を出していると言えるかもしれない。

#### (4) 通常のガンベル・コピュラと生存ガンベル・コピュラ

非対称性が強すぎるせいどちらもそこまでフィットは良くないが、生存ガンベルの方がガンベルよりはるかに当てはまりがよく、ここでも下側Tail Dependenceの相対的な大きさが影響していると考えられる。

<sup>\*19</sup> ただし厳密には  $(\hat{U}_1, \hat{U}_2)$  も真のコピュラとは異なった分布に従うはずである。また  $\hat{F}(\cdot)$  が各サンプルに依存しているため  $(\hat{U}_1, \hat{U}_2)$  のサンプルは独立とはならず(Mashal and Zeevi (2002))、通常最尤推定の理論がそのまま通用しないことになるだろう。だがここでは簡単のため  $(\hat{U}_1, \hat{U}_2)$  が近似的には独立同一に真のコピュラ  $C$  に従うと仮定し、通常最尤推定が妥当であることにして、コピュラ的应用面のみに着目する。

<sup>\*20</sup> 繰り返すが、決して  $\hat{F}$  が真の分布であるとは言っていない。 $\hat{F}$  が漸近的には真の分布に収束するという議論を踏まえた上での、ある種近似的な話である。

(表3) 経験分布を仮定したときの推定結果

- ・推定値下カッコ内は標準的な最尤法が通用すると仮定したときの漸近的標準誤差 (Hessian から計算した)。
- ・Kendall's  $\tau$ , 上側 T.D., 下側 T.D.,  $\hat{\tau}$  についても表2と同様。

① USD,AUD ( $\hat{\tau} = 0.444$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho: 0.64513$ (0.01706)	$\rho: 0.64532$ (0.02153) $\nu: 4.54519$ (1.00061)	$\alpha: 0.38750$ (0.08845) $\beta: 1.83324$ (0.05443)	$\beta: 1.74639$ (0.04897)	$\beta: 1.77725$ (0.04990)
Kendall's $\tau$	0.446	0.447	0.455	0.427	0.437
上側 T.D.	0	0.319	0.209	0.513	0
下側 T.D.	0	0.319	0.331	0	0.523
対数尤度	220.997	236.232	237.445	211.248	226.826
AIC	-439.993	-468.464	<b>-470.889</b>	-420.495	-451.651
BIC	-439.072	-466.622	<b>-469.047</b>	-419.574	-450.730

② GBP,CHF ( $\hat{\tau} = 0.509$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho: 0.72091$ (0.01353)	$\rho: 0.72024$ (0.01662) $\nu: 6.17971$ (1.68519)	$\alpha: 0.34504$ (0.08419) $\beta: 2.04846$ (0.06085)	$\beta: 1.94102$ (0.05465)	$\beta: 1.99180$ (0.05605)
Kendall's $\tau$	0.513	0.512	0.512	0.485	0.498
上側 T.D.	0	0.315	0.206	0.571	0
下側 T.D.	0	0.315	0.391	0	0.584
対数尤度	301.853	312.279	312.643	281.24	302.212
AIC	-601.706	-620.557	<b>-621.287</b>	-560.486	-602.424
BIC	-600.785	-618.715	<b>-619.444</b>	-559.565	-601.503

③ USD,DEM (99年1月1日以前) ( $\hat{\tau} = 0.338$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho: 0.53247$ (0.02948)	$\rho: 0.51483$ (0.03832) $\nu: 4.39269$ (1.30586)	$\alpha: 0.30368$ (0.13817) $\beta: 1.54952$ (0.05853)	$\beta: 1.49346$ (0.05431)	$\beta: 1.53418$ (0.05599)
Kendall's $\tau$	0.357	0.344	0.355	0.330	0.348
上側 T.D.	0	0.242	0.132	0.409	0
下側 T.D.	0	0.242	0.304	0	0.429
対数尤度	75.792	83.270	84.194	71.299	81.260
AIC	-149.584	-162.540	<b>-164.389</b>	-140.598	-160.520
BIC	-148.913	-161.198	<b>-163.046</b>	-140.036	-159.599

④ USD,EUR (99年1月8日以降) ( $\hat{\tau} = 0.345$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho: 0.54361$ (0.03278)	$\rho: 0.54358$ (0.03279) $\nu: 6181181^*$ (0.42897)	$\alpha: 0.27666$ (0.16757) $\beta: 1.50443$ (0.06283)	$\beta: 1.45845$ (0.05895)	$\beta: 1.48525$ (0.05927)
Kendall's $\tau$	0.366	0.366	0.335	0.314	0.327
上側 T.D.	0	0	0.115	0.392	0
下側 T.D.	0	0	0.300	0	0.405
対数尤度	61.620	61.620	58.093	49.344	56.508
AIC	<b>-121.239</b>	-119.239	-112.186	-96.689	-111.016
BIC	<b>-120.677</b>	-118.114	-111.062	-96.126	-110.453

\* 自由度が収束しなかったため、局所的最適解の1つを参考として載せた。

(つづく)

(表3つづき)

⑤ USD,GBP ( $\hat{\tau} = 0.423$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho : 0.63792$ (0.01739)	$\rho : 0.62384$ (0.02365) $\nu : 4.32604$ (0.98850)	$\alpha : 0.31302$ (0.09391) $\beta : 1.77622$ (0.05162)	$\beta : 1.70518$ (0.04749)	$\beta : 1.74961$ (0.04876)
Kendall's $\tau$	0.440	0.429	0.437	0.414	0.428
上側 T.D.	0	0.314	0.164	0.498	0
下側 T.D.	0	0.314	0.359	0	0.514
対数尤度	214.530	227.948	228.659	201.907	221.971
AIC	-427.059	-451.895	<b>-453.319</b>	-401.813	-441.942
BIC	-426.138	-450.053	<b>-451.476</b>	-400.892	-441.021

⑥ GBP,AUD ( $\hat{\tau} = 0.359$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho : 0.55240$ (0.02118)	$\rho : 0.55051$ (0.02266) $\nu : 20.8435$ (16.9957)	$\alpha : 0.28333$ (0.10796) $\beta : 1.56037$ (0.04434)	$\beta : 1.49917$ (0.04044)	$\beta : 1.53477$ (0.04145)
Kendall's $\tau$	0.373	0.371	0.359	0.333	0.348
上側 T.D.	0	0.020	0.125	0.412	0
下側 T.D.	0	0.020	0.316	0	0.429
対数尤度	149.126	150.001	149.159	127.032	145.463
AIC	<b>-296.252</b>	-296.003	-294.318	-252.064	-288.926
BIC	<b>-295.330</b>	-294.160	-292.475	-251.142	-288.005

⑦ USD,CHF ( $\hat{\tau} = 0.300$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho : 0.47965$ (0.02421)	$\rho : 0.46559$ (0.02989) $\nu : 5.80545$ (1.60698)	$\alpha : 0.24913$ (0.11210) $\beta : 1.45635$ (0.04036)	$\beta : 1.40242$ (0.03749)	$\beta : 1.44531$ (0.03877)
Kendall's $\tau$	0.318	0.308	0.313	0.287	0.308
上側 T.D.	0	0.160	0.097	0.361	0
下側 T.D.	0	0.160	0.293	0	0.385
対数尤度	106.846	116.069	118.393	95.789	115.608
AIC	-211.692	-228.137	<b>-232.786</b>	-189.578	-229.216
BIC	-210.771	-226.295	<b>-230.944</b>	-188.657	-228.295

⑧ AUD,CHF ( $\hat{\tau} = 0.281$ )

	正規	t	混合ガンベル	ガンベル	生存ガンベル
推定値	$\rho : 0.43546$ (0.02594)	$\rho : 0.43342$ (0.02906) $\nu : 8.94510$ (3.47423)	$\alpha : 0.30776$ (0.12024) $\beta : 1.39739$ (0.03889)	$\beta : 1.34961$ (0.03586)	$\beta : 1.37669$ (0.03657)
Kendall's $\tau$	0.287	0.285	0.284	0.259	0.274
上側 T.D.	0	0.076	0.110	0.329	0
下側 T.D.	0	0.076	0.248	0	0.346
対数尤度	85.854	89.990	92.006	75.353	88.391
AIC	-169.708	-175.981	<b>-180.012</b>	-148.706	-174.783
BIC	-168.787	-174.139	<b>-178.169</b>	-147.784	-173.862

## 5.4 推定結果のまとめ

繰り返しになる部分が多いが、以上の結果について簡単にまとめる。

- (1) 全体として、tコピュラに比べ正規コピュラのデータへのあてはまりは良くないことが多い。また、たとえ真の分布が正規コピュラであっても、tコピュラは（ネストである分）自由度が大きくなって正規コピュラに近くなるので、（特にリスク管理という点では）誤ってtコピュラを用いてしまうことより、誤って正規コピュラを用いてしまうことの方がはるかに重大であると言える。標準的手法としては正規コピュラよりもtコピュラを仮定する方が望ましい。
- (2) 非対称性を持った混合ガンベル・コピュラの方がtコピュラよりもデータによくフィットする可能性がある。すなわち現実の依存構造に強い非対称性がある場合、必ずしもtコピュラは絶対ではない。
- (3) 混合ガンベル・コピュラも、全体の相関と Tail Dependence に強い連関性があり、その分tコピュラよりあてはまりが悪くなる場合がある。特に、独立である場合を除き Tail Dependence の存在を常に仮定してしまっている<sup>\*21</sup>ことが、今回の実証では問題点として目立った。
- (4) 周辺分布の仮定を誤ると、コピュラのパラメータ推定にバイアスがかかってしまう可能性がある。<sup>\*22</sup> 今回のような（裾の厚い）金融データの周辺分布に正規分布を仮定してしまうと、コピュラの推定値はデータと大きくズれてしまうことになる。
- (5) 特に二変量正規分布は、周辺分布とコピュラの両方について2重にモデリングミスをしてしまう可能性が高い。

## 5.5 リスク管理への応用

最後に、コピュラを用いたリスク管理の具体例を簡単に取り上げる。ここでは周辺分布を仮定した場合のみを扱い、完全なパラメトリックモデルの範囲内で比較を行うことにする。

例として2種類の外貨  $P_0$  円、 $Q_0$  円分を一週間所持したときの下方リスクを想定し、VaR と期待ショートフォールの値を計測する。0期における資産の所持比率として  $P_0 = Q_0$  を仮定すれば、このときの資産合計  $S_t = P_t + Q_t$  円 ( $t = 0, 1$ ) の対数収益率は

$$\log\left(\frac{S_1}{S_0}\right) = \log\left(\frac{P_1}{S_0/2} + \frac{Q_1}{S_0/2}\right) - \log 2 = \log\left(\exp\left(\log\left(\frac{P_1}{P_0}\right)\right) + \exp\left(\log\left(\frac{Q_1}{Q_0}\right)\right)\right) - \log 2$$

と表される ( $\log(\frac{P_1}{P_0})$ ,  $\log(\frac{Q_1}{Q_0})$  は今まで扱っていた変数である)。

通貨は① USD, AUD と② GBP, CHF を用い、二変量正規分布を表す [I : 正規分布×正規コピュラ] と、5.2 で推定した [II : ラプラス分布×その時の最小 AIC コピュラ] を比較する。ここでロジスティックではなくラプラス分布を用いたのは、左裾だけを考えればラプラス分布の方があてはまりが良いという 5.2 での議論による。コピュラに基づいて推定したモデルの中でも、今回のリスク計測にとって最もベストなモデルであろう。

推定した分布から ( $\log(\frac{P_1}{P_0})$ ,  $\log(\frac{Q_1}{Q_0})$ ) の乱数を 10 万個発生させ、VaR と期待ショートフォールの値を計算した (詳細は Carmona (2004) による)、<sup>\*23</sup> 結果が表 4 である。

<sup>\*21</sup> 表 1 の結果を参照。

<sup>\*22</sup> これについては Durrleman et al.(2001) も参照。真の分布を仮定した上での分析をしている。

<sup>\*23</sup> 計算速度の都合上 10 万個にしたが、精度をよくするために、もっと多くの乱数が必要（もしくは別の計算方法を行うべき）だと考えられる。なお、ここではこの数値計算上の誤差については無視することにする。



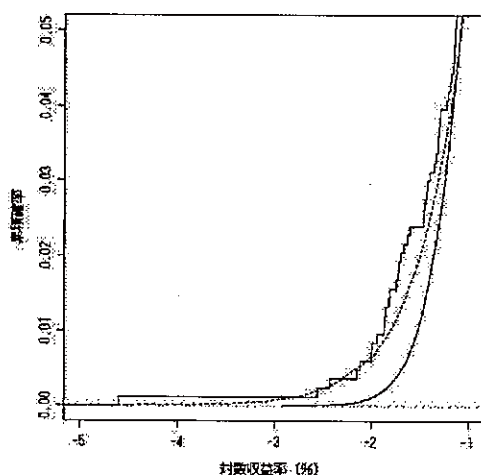
(表4)

VaR		「超過回数」は計算した VaR より損失の上回ったサンプルの数。					
95%	VaR	I		II		超過回数の理論値 (データ数× $\alpha$ )	
		超過回数	VaR	超過回数	VaR		
① USD,AUD	1.073	46	1.060	48		41.7	
② GBP,CHF	1.076	44	1.036	46		41.7	
99%		I		II		超過回数の理論値 (データ数× $\alpha$ )	
	VaR	超過回数	VaR	超過回数	VaR		
① USD,AUD	1.532	20	1.819	11		8.34	
② GBP,CHF	1.502	13	1.797	7		8.34	
99.9%		I		II		超過回数の理論値 (データ数× $\alpha$ )	
	VaR	超過回数	VaR	超過回数	VaR		
① USD,AUD	2.039	5	2.863	1		0.834	
② GBP,CHF	2.033	4	3.079	1		0.834	
期待ショートフォール		「historical」は実際のデータの「下側 $\alpha\%$ 点以下の標本平均」。					
		95%			99%		
		I	II	historical	I	II	historical
① USD,AUD		1.3508	1.5622	1.6519	1.7494	2.4852	2.4664
② GBP,CHF		1.3548	1.5321	1.5519	1.7454	2.3362	2.4954

結果を見ると、全体的に I (二変量正規分布) は実際より楽観的な評価を与えてしまっているが、一方で、コピュラの理論を用いた II は現実のデータをよく捉えた比較的整合性の高い値となっているであろう。特に信頼水準が高いほどその傾向は顕著で、\*24 例えば I のモデルだと 99.9% VaR 以上の損失が何回も観測されてしまい、これはリスク管理をする上で問題があると言わざるをえない。収益率分布の fat tail 性が、二変量正規分布による VaR の計測に大きく影響してしまっていることがわかる。

II のモデルの方が、実際の分布に近い妥当な推定であると言えるだろう。実際、発生させた乱数を元に①について  $\log\left(\frac{S_t}{S_0}\right)$  の分布の左裾を描いてみれば (図6)、2つの違いはより明確である。

(図6) 経験分布関数と I (実線)、II (点線) の左裾



以上より、通常仮定されてしまう二変量正規分布ではなく、コピュラの理論とそのパラメトリックモデルを用いることで、より現実のデータに近いリスク管理ができることが示されたことになる。

\*24 95% VaR で各モデル間の差がほとんど無かったのは、図6を見れば分かるように (直感的に言えば) 下側5%ではまだ分布の裾に達していなかったためである。1つの VaR だけではリスクを適正に把握できない可能性があることを示唆している。

ただ今回の二変量正規分布との比較では、 $t$ コピュラ、混合ガンベル・コピュラだけでなく、ラプラス分布を仮定した周辺分布の違いによる影響も相当大きいと考えられる。そのため純粋な(裾の)依存構造のみの違いでは、この結果よりも相対的に小さな差異しか生まれず、コピュラを導入するメリットは少ないと思うかもしれない。しかし、このように周辺分布と依存構造を切り離して考えられたこと自体がコピュラ概念に基づいており、コピュラを用いることのメリットはⅠとⅡの比較結果に過不足なく集約されている、ということを経後に付け加えておく。

## 6 まとめと今後の課題

以上、コピュラと呼ばれる概念の性質を明らかにし、二変量正規分布に代わるモデリングで現実のデータへの応用を試みた。

[同時分布] = [周辺分布×コピュラ] という分解 (Sklarの定理) の結果、依存構造のみにスポットをあてた分析ができるようになり、モデリングの柔軟性も上昇する。これこそがコピュラ概念を用いることの最大のメリットであることが前半の議論で示された。そしてそのパラメトリック族としていくつかのコピュラの性質を明らかにすることもできた(この部分については2節~4節でも詳しくまとめている)。

実証分析では為替レートの収益率分布に焦点をあてた。そこで依存構造が多くの場合正規コピュラとは異なった形状をしており、相対的に $t$ コピュラの方が安定して優れていると言う結論を得るに至った。また $t$ コピュラも絶対ではないということが、混合ガンベル・コピュラによって示すことができた。そして少なくとも二変量正規分布については、データに対するフィットが最も悪いことが明らかとなった(ここも5.4でまとめている)。

しかし本論文の実証分析は多くの仮定・単純化のもとで成り立っており、厳密に検討すべき課題はまだ多い。

特に、時系列データをプールして独立同一分布の仮定をあてはめたことは直感的にも大きな問題点であり、他の時系列モデルと組み合わせるなど改善を図るべきである。

また、モデルが複雑であるために仮説検定を十分に行うことができなかったことも今回の課題である。モデルの選択基準としては簡単にAIC, BICを用いたが、この部分についてもより厳密な議論が必要であると考えられる。特にリスク管理を行う際には全体のフィットに加え、裾のフィットをより重視すべきであり、そのような重み付きのモデル選択を行うべきかもしれない。モデリングの柔軟性が増したということは、モデル選択が難しくなるということと裏表であり、コピュラを用いる上でモデル選択理論の検証は必要不可欠であると言える。

このように後半の理論的な内容に多くの課題を残したままであるが、最後にVaR等の計算を試すことができたように、二変量正規分布に代わるモデルとしてのコピュラの応用を少しは実践することができたと思う。

## 付録A 乱数の発生方法

各コピュラに従う乱数の生成について、特に本論文で用いた方法を述べる。より包括的な解説はCherubini et al. (2004), Embrechts et al. (2001)を見よ。

- 正規コピュラ  $C_\rho^N$

1. コレスキー分解を用い、相関係数  $\rho$  の2変量標準正規分布に従う乱数  $(X_1, X_2)$  を発生させる。
2.  $U_1 = \Phi(X_1), U_2 = \Phi(X_2)$  と変換すれば、 $(U_1, U_2)$  は  $C_\rho^N$  に従う乱数である。

- $t$ コピュラ  $C_{\rho, \nu}^t$

1. 2変量標準正規乱数と自由度  $\nu$  のカイ2乗乱数を用い、相関係数  $\rho$ 、自由度  $\nu$  の2変量標準  $t$  分布に従う乱数  $(X_1, X_2)$  を発生させる。
2.  $U_1 = t_\nu(X_1), U_2 = t_\nu(X_2)$  と変換すれば、 $(U_1, U_2)$  は  $C_{\rho, \nu}^t$  に従う乱数である。

- ガンベル・コピュラ  $C_\beta^{G_u}$ 、生存ガンベル・コピュラ  $\bar{C}_\beta^{G_u}$

1.  $U(0, 1)$  に従う独立な変数  $U_1, V$  を生成する。
2.  $U_2 = C_\beta^{-1}(V|U_1)$  とすれば、 $(U_1, U_2)$  はコピュラ  $C_\beta^{G_u}$  に従う乱数となる。  
(2. で  $(1 - U_1, 1 - U_2)$  とすれば  $\bar{C}_\beta^{G_u}$  に従う乱数となる)

ここで  $C(u_2|u_1)$  は、一般にコピュラ  $C$  の条件付分布関数<sup>\*25</sup>であり、

$$\begin{aligned} F(u_2|u_1) &= P(U_2 < u_2 | U_1 = u_1) = P(C^{-1}(V|u_1) < u_2 | U_1 = u_1) \\ &= P(V < C(u_2|u_1) | U_1 = u_1) \\ &= C(u_2|u_1) \end{aligned}$$

と変形し、 $c(u_2|u_1) = c(u_1, u_2)$ ,  $f(u_2|u_1) = f(u_1, u_2)$  に注意すれば、 $F(u_1, u_2) = C(u_1, u_2)$  を確認できる。

• 混合ガンベル・コピュラ  $C_{\alpha, \beta}^{MG}$

1.  $U(0, 1)$  に従う独立な変数  $U, V, W$  を生成する。
2.  $W \leq \alpha$  のとき  $U_2 = C_{\beta}^{-1}(V|U)$ ,  $U_1 = U$  とし、 $W > \alpha$  のとき  $U_2 = 1 - C_{\beta}^{-1}(V|U)$ ,  $U_1 = 1 - U$  とすれば、 $(U_1, U_2)$  はコピュラ  $C_{\alpha, \beta}^{MG}$  に従う乱数である。

(注) 本プログラミングも他と同様 R2.2.0 を用いて行った。

- 正規乱数、カイ 2 乗乱数の生成には `rnorm`, `rchisq` を、また標準正規分布関数、t 分布関数は `pnorm`, `pt` を用いた。
- 条件付ガンベル・コピュラの逆関数は解析的に求まらないため、`uniroot` を使ってそれぞれ数値的に計算した。

## 付録 B 各パラメトリック・コピュラの性質

### B.1 密度関数

最尤推定を行う際に用いた各コピュラの密度関数を載せておく。正規・t コピュラについては (2.3) 式より計算した。ガンベル・混合ガンベルコピュラについても導出は容易である。

$$c_{\rho}^N(u_1, u_2) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(\frac{2\rho xy - x^2 - y^2}{2(1-\rho^2)} + \frac{x^2 + y^2}{2}\right)$$

where  $x = \Phi^{-1}(u_1)$ ,  $y = \Phi^{-1}(u_2)$ .

$$c_{\rho, \nu}^t(u_1, u_2) = \frac{\nu \{\Gamma(\frac{\nu}{2})\}^2}{2\sqrt{1-\rho^2} \{\Gamma(\frac{\nu+1}{2})\}^2} \left\{ \left(1 + \frac{x^2}{\nu}\right) \left(1 + \frac{y^2}{\nu}\right) \right\}^{\frac{\nu+1}{2}} \left(1 + \frac{x^2 + y^2 - 2\rho xy}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}$$

where  $x = t_{\nu}^{-1}(u_1)$ ,  $y = t_{\nu}^{-1}(u_2)$ .

$$c_{\beta}^{Gu}(u_1, u_2) = \frac{1}{u_1 u_2} (-\log u_1)^{\beta-1} (-\log u_2)^{\beta-1} \exp\left(-\{(-\log u_1)^{\beta} + (-\log u_2)^{\beta}\}^{\frac{1}{\beta}}\right) \\ \times \left\{ (-\log u_1)^{\beta} + (-\log u_2)^{\beta} \right\}^{\frac{1}{\beta}-2} \left(\beta + 1 + \{(-\log u_1)^{\beta} + (-\log u_2)^{\beta}\}^{\frac{1}{\beta}}\right).$$

$$\bar{c}_{\beta}^{Gu}(u_1, u_2) = c_{\beta}^{Gu}(1-u_1, 1-u_2).$$

$$c_{\alpha, \beta}^{MG}(u_1, u_2) = \alpha c_{\beta}^{Gu}(u_1, u_2) + (1-\alpha) \bar{c}_{\beta}^{Gu}(1-u_1, 1-u_2).$$

<sup>\*25</sup>  $X_1 = x_1$  における  $X_2$  の条件付分布関数は、一般に  $F(x_2|x_1) = \lim_{\varepsilon \downarrow 0} P(X_2 \leq x_2 | x_1 - \varepsilon < X_1 \leq x_1 + \varepsilon)$  と定義されるが (高島 (1977) などを参照)、特にコピュラの場合  $U_1, U_2$  は一様分布するので

$$C(u_2|u_1) = \lim_{\varepsilon \downarrow 0} \frac{P(U_1 \leq u_1 + \varepsilon, U_2 \leq u_2) - P(U_1 \leq u_1 - \varepsilon, U_2 \leq u_2)}{2\varepsilon} = \frac{\partial C(u_1, u_2)}{\partial u_1}$$

である。

## B.2 極限分布

### B.2.1 正規コピュラ, tコピュラの $C^+, C^-$ への収束

$C^N, C^t$  は,  $\rho \rightarrow 1$  のとき  $C^+$  へ,  $\rho \rightarrow -1$  のとき  $C^-$  へ分布収束する.\*26直感的には, 相関行列が退化し 2次元正規分布 (2次元 t分布) が 1次元分布となることを, コピュラの立場から表現したものであるが, 証明を行っている文献が特に見当たらないため, ここで導出する. 以下では正規コピュラの場合のみ扱う (tコピュラの場合も同様).

#### (1) $C^+$ への分布収束

(証明)

$u_1 \leq u_2$  を仮定する. まず正規コピュラの  $U_1 = s$  における条件付分布は, B.1における結果から,

$$\begin{aligned} C_\rho^N(u_2|s) &= \int_0^{u_2} c_\rho^N(t|s) dt = \int_0^{u_2} c_\rho^N(s, t) dt \\ &= \int_0^{\Phi^{-1}(u_2)} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{(y-\rho\Phi^{-1}(s))^2}{2(1-\rho^2)}\right) dy \quad (t = \Phi(y) \text{ と置換して整理.)} \\ &= \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(s)}{\sqrt{1-\rho^2}}\right) \end{aligned} \quad (*)$$

と簡単に計算される. ここで  $\rho \rightarrow 1$  のとき, 分布関数  $C_\rho^N(u_2|s)$  は任意の  $s, u_2$  に対して関数

$$\delta_s(u_2) = \begin{cases} 0 & \text{if } u_2 < s \\ 1 & \text{if } u_2 \geq s \end{cases} \quad (**)$$

に各点収束する (すなわち  $U_2|U_1 = s$  が  $s$  に確率収束する)\*27. 特に (\*), (\*\*) を  $s$  の関数と見れば,  $\delta_s(u_2)$  は  $[0, u_1]$  で連続 ( $\because u_1 \leq u_2$  より常に 1), かつ  $|C_\rho^N(u_2|s)| \leq 1$  (有界) よりアルツェラの定理\*28から積分と極限の交換ができて

$$\begin{aligned} \lim_{\rho \rightarrow 1} C_\rho^N(u_1, u_2) &= \lim_{\rho \rightarrow 1} \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(s)}{\sqrt{1-\rho^2}}\right) ds \\ &= \int_0^{u_1} \delta_s(u_2) ds \\ &= \int_0^{u_1} 1 ds \quad (\because u_1 \leq u_2) \\ &= u_1. \end{aligned}$$

$u_2 \leq u_1$  のときも,  $U_1|U_2 = t$  の条件付分布について同様の議論を繰り返せば,  $\lim_{\rho \rightarrow 1} C_\rho(u_1, u_2) = u_2$ .

ゆえに  $\lim_{\rho \rightarrow 1} C_\rho(u_1, u_2) = \min(u_1, u_2)$ , すなわち  $C^+$  への分布収束が示された. ■

#### (2) $C^-$ への分布収束

(証明)

$u_1 \leq u_2$  を仮定する. (1) と同様, 条件付分布関数  $C_\rho^N(u_2|s)$  は  $\rho \rightarrow -1$  で

$$\delta_{1-s}(u_2) = \begin{cases} 0 & \text{if } u_2 < 1-s \\ 1 & \text{if } u_2 \geq 1-s \end{cases}$$

\*26 細かいことを言えば, それぞれ片側からの極限である.

\*27  $U_1 = s$  としたとき, (\*) から確率変数  $\Phi^{-1}(U_2)$  は平均  $\rho\Phi^{-1}(s)$ , 分散  $1-\rho^2$  の正規分布に従う. よって  $\rho \rightarrow 1$  のとき  $\Phi^{-1}(U_2)$  は  $\Phi^{-1}(s)$  に確率収束し,  $U_2$  も  $s$  に確率収束する.

\*28 小平 (1991) を参照.

に各点収束することが示される。ここで (1) と同様の議論のもと以下を計算する。

$$\begin{aligned}
\lim_{\rho \rightarrow -1} P_\rho(1 - u_1 < U_1 \leq 1, U_2 \leq u_2) &= \lim_{\rho \rightarrow -1} \int_{1-u_1}^1 \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(s)}{\sqrt{1-\rho^2}}\right) ds \\
&= \int_{1-u_1}^1 \delta_{1-s}(u_2) ds \\
&= \int_{1-u_1}^1 1 ds \quad (\because u_1 \leq u_2 \text{ より } 1 - (1 - u_1) \leq u_2) \\
&= u_1.
\end{aligned}$$

$u_2 \leq u_1$  のときもほぼ同じ様な計算ができて、 $\lim_{\rho \rightarrow -1} P_\rho(1 - u_1 < U_1 \leq 1, U_2 \leq u_2) = u_2$ . ゆえに

$$\begin{aligned}
\lim_{\rho \rightarrow -1} \{u_2 - C_\rho^N(1 - u_1, u_2)\} &= \min(u_1, u_2) \\
\Leftrightarrow u_2 - \lim_{\rho \rightarrow -1} C_\rho^N(u_1, u_2) &= \min(1 - u_1, u_2) \\
\Leftrightarrow \lim_{\rho \rightarrow -1} C_\rho^N(u_1, u_2) &= \max(u_1 + u_2 - 1, 0).
\end{aligned}$$

$C^-$  へ分布収束する. ■

### B.2.2 ガンベル・コピュラ, 混合ガンベル・コピュラの $C^+$ への収束

Nelsen (1999) 4.4 節にアルキメディアン・コピュラ族全体に対する  $C^+$  への収束条件が証明されているが、やや一般的で分かりづらい。しかしガンベル・コピュラの場合に限って計算したところ、以下のように非常に簡単な方法で  $C^+$  への収束が示された。

(証明)

任意の非負の実数  $a, b$  に対し  $a > b$  を仮定すると、 $(a^x + b^x)^{\frac{1}{x}} = a(1 + (\frac{b}{a})^x)^{\frac{1}{x}}$  は、

$$a \leq a(1 + (\frac{b}{a})^x)^{\frac{1}{x}} < a(1 + (\frac{a}{a})^x)^{\frac{1}{x}} = 2^{\frac{1}{x}} a$$

から、はさみうちの原理より  $x \rightarrow \infty$  のとき  $a$  に収束する。同様に  $a < b$  のとき  $(a^x + b^x)^{\frac{1}{x}} \rightarrow b$ . また  $a = b$  ならば  $(a^x + b^x)^{\frac{1}{x}} = 2^{\frac{1}{x}} a \rightarrow a$  となるから  $\lim_{x \rightarrow \infty} (a^x + b^x)^{\frac{1}{x}} = \max(a, b)$ .

この結果を用いれば、ガンベル・コピュラは任意の  $u_1, u_2 \in [0, 1]$  において

$$\begin{aligned}
\lim_{\beta \rightarrow \infty} C_\beta^{G^u}(u_1, u_2) &= \exp(-\max(-\log u_1, -\log u_2)) \\
&= \exp(\min(\log u_1, \log u_2)) = \min(u_1, u_2).
\end{aligned}$$

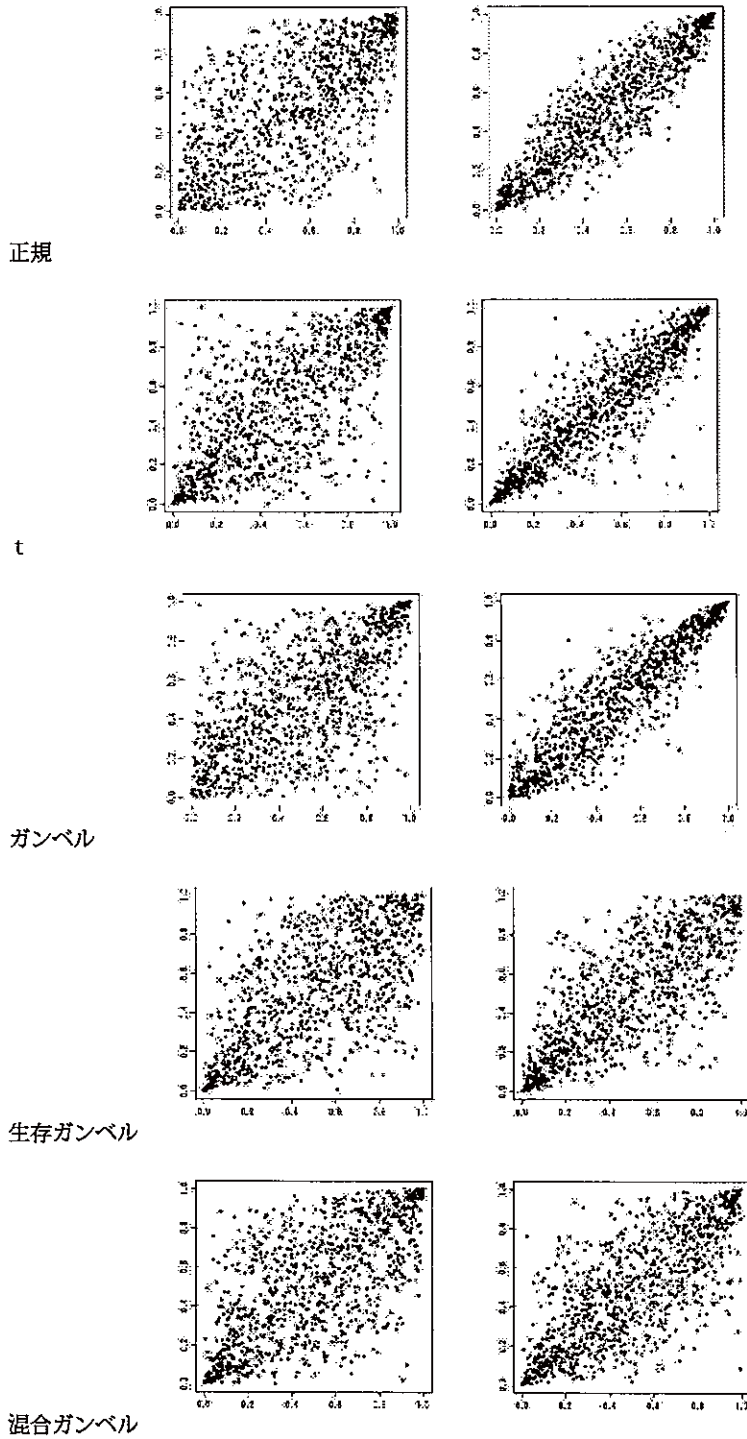
よって  $C^+$  へ分布収束する。ゆえに混合ガンベルも

$$\begin{aligned}
\lim_{\beta \rightarrow \infty} C_{\alpha, \beta}^{MG}(u_1, u_2) &= \alpha \lim_{\beta \rightarrow \infty} C_\beta^{G^u}(u_1, u_2) + (1 - \alpha) \lim_{\beta \rightarrow \infty} \bar{C}_\beta^{G^u}(u_1, u_2) \\
&= \alpha \min(u_1, u_2) + (1 - \alpha)(u + v - 1 + \min(1 - u_1, 1 - u_2)) \\
&= \min(u_1, u_2).
\end{aligned}$$

と  $C^+$  へ分布収束する。生存ガンベルについても同様である. ■

## 付録C 各コピュラの散布図

- ケンドール $\tau$ を0.5 (左), 0.75 (右) と固定して各コピュラの散布図をプロット.
- tコピュラは $\nu=3$ , 混合ガンベル・コピュラは $\alpha=0.3$ とした.



## 参考文献

- [1] Carmona, R. (2004). *Statistical Analysis of Financial Data in S-PLUS*, Springer Verlag, New York.
- [2] Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance*, John Wiley and Sons, London.
- [3] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer Verlag, London.
- [4] Demarta, S. and McNeil, A. J. (2004). *The t Copula and Related Copulas*, ETH Zurich.
- [5] Durrleman, V., Nikeghbali, A. and Roncalli, T. (2001). Which copula is the right one? *Credit Lyonnais*.
- [6] Embrechts, P., Lindskog, F. and McNeil, A. J. (2001). *Modelling Dependence with Copulas and Applications to Risk Management*, ETH Zurich.
- [7] Fantazzini, D. (2004). Copula's Conditional Dependence Measures for Portfolio Management and Value at Risk, *Microstructure of Financial Markets in Europe*.
- [8] Hoeffding, W. (1994). Scale-invariant correlation theory, *The collected works of Wassily Hoeffding*, Springer Verlag, New York.
- [9] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- [10] Liu, W. (2005). *Distributions of Currencies Portfolio Return: A Copula Methodology*, University of Toronto.
- [11] Mashal, R. and Zeevi, A. (2002). Beyond correlation: extreme co-movements between financial assets, Columbia University.
- [12] Nelsen, R. B. (1999). *An Introduction to Copulas*, Springer Verlag, New York.
- [13] Schmidt, R. (2004). Tail dependence, *Statistical Tools In Finance and Insurance*, Springer Verlag, New York.
- [14] 小平邦彦 (1991), 「解析入門」, 岩波書店.
- [15] 高島巳千雄 (1977), 「序説確率・統計」, 東京教学社.
- [16] 竹内啓 (1975), 「確率分布と統計解析」, 日本規格協会.
- [17] 養谷千風彦 (2003), 「統計分布ハンドブック」, 朝倉書店.
- [18] 山井康浩・吉羽要直 (2002), 「市場ストレス時におけるバリュアットリスクと期待ショートフォールの比較: 多変量極値分布のもとでの比較分析」, 『金融研究』第 21 巻別冊第 2 号, 日本銀行金融研究所.

## 第III部: 保険と統計学

8. 「保険リスクと破産理論 (I),(II)」室井芳史
9. 「変額年金保険の統計的リスク管理法：局面転換モデルの利用」（日本アクチュアリー会編「リスクと保険」2006年、）秋山豪太・国友直人
10. 「Swiss Solvency Test について」川崎能典



# 保険リスクと破産理論 (I) \*

室井 芳史

日本銀行金融研究所

e-mail:yoshifumi.muroi@boj.or.jp

**概要:** 本稿では、保険リスクと破産理論の入門部分について解説を行う。これらの理論は古くから研究が行われているが、近年数理ファイナンスとの関係等にも刺激され研究が深化している分野である。ここでは、余り難しい議論には立入らず、かつ、なるべく保険理論の概要が分かるように解説を行いたい。また、近年の数理ファイナンスの研究との関連が分かるような解説を行う。なお、本稿の流れは基本的に Dickson (2005) に従っている。なお、筆者が一見したところ Dickson (2005) はより包括的な数理保険学の教科書 Klugman, Panjer and Willmot (2004) の影響を多分に受けながら執筆されてたという印象を得ている。また、保険に関する包括的な議論は森本 (2000) を参照。

**キーワード:** 保険料算出原理、逐次計算法、破産理論

## 1 保険料率算出原理

本節では、保険料の満たすべき望ましい性質について議論を行い、どのように保険料率を決めればそれらの性質を満たすのかについて議論を行いたい。リスクを確率変数  $X$  で表現する。このリスクをカバーする保険のプレミアムは確率変数  $X$  を与えるとある実数へ対応させる (汎) 関数  $\phi(\cdot)$  を通じて計算される。すなわち、リスク  $X$  が存在するとき、リスクに見合った保険料をリスク (確率変数) から実数への写像と考える。保険料を  $\Pi_X$  と書くと  $\Pi_X = \phi(X)$  と書ける。さて、望ましい保険料の価格のつき方とはどのような性質を満たすべきであろうか。そこで、望ましいと思われる性質をいくつか挙げてみる。

- (i) 非負の運用実績:  $\Pi_X \geq E[X]$  が成り立つ。
- (ii) 加法性: リスク  $X_1$  と  $X_2$  は独立なリスクとする。このとき  $\Pi_{X_1+X_2} = \Pi_{X_1} + \Pi_{X_2}$  が成り立つ。
- (iii) スケール不変性: 正の実数  $a > 0$  があるとする。  $Z = aX$  とすると、  $\Pi_Z = a\Pi_X$  が成り立つ。
- (iv) 一致性:  $Y = X + c$  であれば、  $\Pi_Y = \Pi_X + c$  が成り立つ。

---

\*本稿は、室井個人の文責によるものであり日本銀行および金融研究所の公式見解を示すものではない。

(v) ずるがない (No rip-off): リスクの最大値を  $x_m$  とおくと  $\Pi_X \leq x_m$  が成り立つ。

これらの性質がなるべく多く成り立つ保険料の算出方法が良い方法であると思われる。これらの性質は、数理ファイナンスにおいてリスクをどのように捕らえるか論じた Artzner et. al. (1999) によるコヒーレント・リスク測度の研究と平行して研究されてきた<sup>1</sup>。このような議論がなされる背景には、保険リスクはファイナンスの世界と異なり非完備市場モデルを基礎にしていることが挙げられる。本稿では、代表的な保険料率算出原理であるゼロ効用原理、エッシャー原理、リスク調整プレミアム原理について主に言及を行う。なお、これ以外にもさまざまな保険料算出の方法が知られている。

1) 期待値原理: 任意の正数  $\theta > 0$  を用いて

$$\Pi_X = (1 + \theta)EX$$

で保険料を決める。一致性や no rip-off が成り立たない。

2) 分散原理: 任意の  $\alpha > 0$  について

$$\Pi_X = EX + \alpha Var[X]$$

で保険料を決める。加法性や no rip-off が成り立たない。これ以外にも、標準偏差原理や分位原理<sup>2</sup> (森本 (2000) を参照) 等の保険料算出原理が知られている。

## 1.1 ゼロ効用原理

いま、 $u'(x) > 0$  および  $u''(x) < 0$  を満たす効用関数  $u(x)$  を考える。効用関数の保険への適用は、例えば Gerber and Pafumi (1998) において保険料率の決定にどのように効用関数が利用されているか論じられている。ゼロ効用原理では保険を売った場合と売らなかった場合に、効用が同等になるように保険料を決定することを提案している。すなわち

$$u(W) = E[u(W + \Pi_X - X)]$$

で保険料率  $\Pi_X$  を決定する。特に、指数効用関数  $u(x) = -\exp\{-\beta x\}$  を考えれば、

$$\Pi_X = \beta^{-1} \log E[\exp(\beta X)]$$

<sup>1</sup> Artzner et. al. (1999) は、

Axiom T. リスク  $X$  は  $\rho(X + \alpha \cdot r) = \rho(X) - \alpha$  が成り立つこと。(Translation Invariance)

Axiom S. リスク  $X$  は  $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$  が成り立つこと。(劣加法性)

Axiom PH. リスク  $X$  と正数  $\lambda > 0$  は  $\rho(\lambda X) = \lambda \rho(X)$  が成り立つこと。(正值斉次性)

Axiom M. 任意の  $X \leq Y$  を満たす  $X$  と  $Y$  について  $\rho(Y) \leq \rho(X)$

Axiom R. 全ての  $X \leq 0$  を満たす  $X \neq 0$  について、 $\rho(X) > 0$  が成り立つ。

の各性質を満たすリスク測度の考察を行った。

<sup>2</sup> VaR のように確率分布の分位点をもとに保険料率を決める。

が成り立つ。この保険料の決定法を指数原理と呼ぶこともある。ゼロ効用原理のもとでは、保険料は非負の運用実績 (i) であることが示される。Jensen の不等式より

$$u(W) = E[u(W + \Pi_X - X)] \leq u(W + \Pi_X - E[X])$$

であるが、 $u'(x) > 0$  が成り立つので  $\Pi_X \geq E[X]$  であることが示される。また、同様の方法で No rip-off であることも簡単に示すことができる。その一方、ゼロ効用原理のもとでは加法性が成り立たない。また、保険料率が初期時点での資本に依存する等の好ましくない点が指摘されている。ところが、指数原理を用いて保険料率を算出した場合は加法性や初期時点での資本の非依存性が成り立つ。その一方で、スケール不変性は成り立たない。

ゼロ効用原理の考え方は、近年、数理ファイナンスにおいても重要視されている。非完備市場において派生商品の価格をどのように算出するべきか考えよう。例えば、株式価格過程が確率ボラティリティ・モデル

$$\begin{aligned} dS_t &= S_t(\mu dt + f(Y_t)dW_t) \\ dY_t &= a(Y_t)dt + b(Y_t)dZ_t \end{aligned}$$

と書けるとする。二つの確率過程  $W_t$  と  $Z_t$  は相関  $cov(dW_t, dZ_t) = \rho dt$  を持つ標準ブラウン運動である。いま、預金金利は  $r (> 0)$  で一定とする。このモデルでオプションの評価を行なうと非完備市場モデルを考えないといけぬ。このとき、株式と預金による複製戦略を考え、預金の持ち高を  $\pi^0(t)$ 、株式の持ち高を  $\pi(t)$  とすると、ポートフォリオの価値過程  $\{X_t\}$  は

$$dX_t = rX_t dt + (\mu - r)\pi(t)dt + \pi(t)\sigma(Y_t)dW_t$$

と書ける。満期における株式の価格  $S(T)$  でペイオフ  $C(T) = C(S(T))$  が決まるオプションの評価を行ないたい。そこで、次の2対の確率制御問題

$$\begin{aligned} V(x, y, t) &= \sup_{\pi_1} E[u(X(T)) | X(t) = x, Y(t) = y] \\ U(s, x, y, t) &= \sup_{\pi_2} E[u(X(T) - C(T)) | S(t) = s, X(t) = x, Y(t) = y] \end{aligned}$$

を考える。このとき、

$$V(x, y, t) = U(s, x, y + h^w, t)$$

を満たす、 $h^w = h^w(t, s, y)$  を売り手の "indifference price"、

$$V(x, y - h^b, t) = U(x, y, t)$$

を満たす、 $h^b = h^b(t, s, y)$  を買い手の "indifference price" と呼ぶ。指数原理を考えると、初期時点での資本の非依存性や売り手と買い手の "indifference price" の一致等が示される。具体的なヘッジ方法や効用の最大値、オプションの価格を計算するためには HJB 方程式 (Hamilton=Jacobi=Bellman 方程式) と呼ばれる、確率制御理論でしばしば現れる非線形偏微分方程式を計算する必要がある。このような話題の数理ファイナンスでの応用は Zariphopoulou (1999) 等を保険での応用は Young and Zariphopoulou (2002) 等を参照せよ。

## 1.2 エッシャー原理

エッシャー原理は

$$\Pi_X = \frac{E[Xe^{hX}]}{E[e^{hX}]}$$

で保険料を決めることである。この原理はスケール不変性を除く多くの望ましい保険料の性質を満たす。 $(0, \infty)$  で定義される確率変数  $X$  の確率密度関数を  $f$  とおき、新しい関数  $g$  を

$$g(x) = \frac{e^{hx} f(x)}{\int_0^\infty e^{hx} f(x) dx}$$

とおく。 $g$  は確率密度関数で、対応する分布関数は

$$G(x) = \frac{\int_0^x e^{hy} f(y) dy}{M_X(h)}$$

である。ただし、 $M_X(h)$  は確率変数  $X$  の積率母関数。すると、 $G$  を確率分布関数とする確率変数の積率母関数  $M_{\tilde{X}}(t)$  とすると

$$M_{\tilde{X}}(t) = \frac{M_X(t+h)}{M_X(h)}$$

簡単な計算により、

$$E[\tilde{X}^r] = \frac{M_X^{(r)}(h)}{M_X(h)}$$

が成り立つ。よって、

$$\frac{d}{dh} \Pi_X = \frac{d}{dh} E[\tilde{X}] = \frac{d}{dh} \frac{M_X'(h)}{M_X(h)} = E[\tilde{X}^2] - E[\tilde{X}]^2 \geq 0$$

により  $\Pi_X$  は  $h$  の増加関数であることが分かる。よって、 $\Pi_X \geq E[X]$  が成り立つ。よって、エッシャー原理において非負の運用実績 (i) が成り立つ。なお、加法性、一致性、No rip-off の3つの性質は簡単な計算により示すことができる。エッシャー原理はスケール不変性を満たさない。

一見、この式で保険料を決定する意味は分かりにくいだが、実は Bühlmann (1980) に見られるようにリスクを交換する交換経済で一般均衡を考えることでこの価格式が導出される。また、Gerber and Pafumi (1998) にも詳細な解説がある。以下の説明は、森本 (2000) にも簡単な解説があるが、重複を恐れずにここでも解説を行なう。いま、 $n$  人の経済主体を考え、事象  $\omega$  が起これば経済主体  $i$  に損失  $X_i(\omega)$  が発生するものとする。経済主体はリスクに対して、保険取引ができるとする。事象  $\omega$  の発生で経済主体  $i$  の得られる保険金額を  $Y_i(\omega)$  であるとする。保険料がある確率変数  $\phi$  を用いて

$$\int_{\Omega} Y_i(\omega) \phi(\omega) dP(\omega) = E[Y_i \phi]$$

と書けるものとする。また、純粋交換経済を考え  $\sum_{i=1}^n Y_i(\omega) = 0$  が成り立つものとする。

この設定のもと経済主体  $i$  の効用関数が  $u_i(x)$  であるならば、各個人が自己の効用を最大化する問題

$$\max \int_{\Omega} u_i(-X_i(\omega) + Y_i(\omega) - \int_{\Omega} Y(\omega') \phi(\omega') dP(\omega')) dP(\omega)$$

を考える。そこで、上の問題を最大化する  $(\phi, Y)$ 、(パレート) 均衡点を見つける。

各効用関数がリスク回避度のパラメータ  $\delta$  を持つ指数効用を考えれば、

$$\phi(\omega) = \frac{e^{\delta Z}}{E[e^{\delta Z}]}$$

となる。ただし、 $Z = \sum_{i=1}^n X_i$  である。よって、リスク  $X$  に対する保険料は

$$\frac{E[X e^{\delta Z}]}{E[e^{\delta Z}]}$$

であるが、 $X$  と  $Z - X$  が独立な場合は

$$\frac{E[X e^{\delta Z}]}{E[e^{\delta Z}]} = \frac{E[X e^{\delta X}]}{E[e^{\delta X}]}$$

が成り立つ。近年、Esscher 変換はファイナンスへの応用も考えられている。これについては Gerber and Shiu (1994,1996) を参照。

### 1.3 リスク調整プレミアム原理

$X$  を非負の値を取る確率変数として分布関数を  $F$  とする。リスク調整済みプレミアム原理とは

$$\Pi_X = \int_0^{\infty} Pr[X > x]^{1/\rho} dx = \int_0^{\infty} (1 - F(x))^{1/\rho} dx$$

で保険料を決定する方法である。 $\rho \geq 1$  をリスク指数と呼ぶ。

$$1 - H(x) = (1 - F(x))^{1/\rho}$$

を満たす関数  $H(x)$  も確率分布関数であり、Esscher 原理の時と同様にもとの分布に重みを付け直して平均を取っている。このような、保険料算出法は Wang (1995) により研究されたものである。なお、上記の積分で  $\rho = 1$  のとき、

$$\Pi_X = \int_0^{\infty} Pr[X > x] dx = \int_0^{\infty} (1 - F(x)) dx (= E[X])$$

を満たすことに注意しておく。リスク調整プレミアム原理では加法性が成立しない。しかし、劣加法性を示すことができる。

## 2 リスク集計モデル

本節では、決められた時間内 (例えば1年間) のリスクを集計する手法について考察を行う。いま、確率変数  $S$  を1年間に発生する保険の請求の集計額とする。確率変数  $N$  で請求件数を表現し、 $X_i$  で第  $i$  請求の請求額を表現する。すると、

$$S = \sum_{i=1}^N X_i$$

と書くことができる。ここで、 $\{X_i\}$  は独立同一分布に従うものとする。このとき  $G(x) = Pr[S \leq x]$  を求めるのは数理保険学において重要な問題である。

$$G(x) = Pr[S \leq x] = \sum_{n=0}^{\infty} Pr[S \leq x, N = n] = \sum_{n=0}^{\infty} Pr[S \leq x | N = n] Pr[N = n]$$

ところが、

$$Pr[S \leq x | N = n] = Pr\left[\sum_{i=1}^n X_i \leq x\right] = F^{n*}(x)$$

が成り立つので、確率分布関数は

$$G(x) = \sum_{n=0}^{\infty} p_n F^{n*}(x)$$

で与えられる。また、確率密度関数  $g_x$  は

$$g_x = \sum_{n=1}^{\infty} p_n f_x^{n*}$$

ただし、

$$f_x^{n*} = Pr\left[\sum_{i=1}^n X_i = x\right]$$

である。ただし、一般的に分布関数や密度関数の畳み込みを計算するのは難しい問題である。それでは、どのように計算を行えば効率的に  $g_x$  を計算することができるだろうか？そのための手法の一つが Panjer (1981) による逐次計算法である。なお、逐次計算法は数理保険学の代表的な研究の一つであり今日でも研究が行なわれている。逐次計算法は、Klugman, Panjer and Willmot (2004) でも詳しく解説が行なわれているようである。

### 2.1 (a,b,0) クラス分布

確率分布が  $(a, b, 0)$  分布に従っているとは、請求件数が  $n$  である確率  $p_n$  が

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad p_0 > 0$$

を満たすことを言う。このモデルでは、退化しない分布ではポアソン分布、負の2項分布、2項分布を表現できることが Sundt and Jewell (1981) によって示されている。確率変数  $X$  の確率母関数を  $P_X(r)$  と書くとする。すると、

$$P_N(r) = p_0 + \sum_{n=1}^{\infty} r^n p_n$$

とすれば、

$$\begin{aligned} P'_N(r) &= \sum_{n=1}^{\infty} n r^{n-1} p_n = \sum_{n=1}^{\infty} n r^{n-1} \left(a + \frac{b}{n}\right) p_{n-1} \\ &= a \sum_{n=1}^{\infty} n r^{n-1} p_{n-1} + b P_N(r) \\ &= a \sum_{n=1}^{\infty} (n-1) r^{n-1} p_{n-1} + a \sum_{n=1}^{\infty} r^{n-1} p_{n-1} + b P_N(r) \\ &= a r P'_N(r) + (a+b) P_N(r) \end{aligned} \tag{2.1}$$

## 2.2 Panjer 逐次計算法

集計されたりスクが0の確率、すなわち  $S=0$  の確率は『 $N=0$  であるか、 $N=n$  であり  $X_i=0$  ( $i=1, \dots, n$ ) であるか』の確率と等しい。確率変数  $X_i$  の独立性より

$$Pr\left[\sum_{i=1}^n X_i = 0\right] = f_0^n$$

が成り立つ。確率変数  $S$  が0である確率は

$$g_0 = p_0 + \sum_{n=1}^{\infty} p_n f_0^n = P_N(f_0)$$

である。確率変数  $S$  の確率母関数  $P_S(r) = P_N(P_X(r))$  は与えられ、(2.1) より

$$P'_S(r) = P'_N(P_X(r)) P'_X(r) = \{a P_X(r) P'_N(P_X(r)) + (a+b) P_N(P_X(r))\} P'_X(r)$$

が成り立つ。よって、

$$P'_S(r) = a P_X(r) P'_S(r) + (a+b) P_S(r) P'_X(r)$$

である。ここで、

$$P_S(r) = \sum_{j=0}^{\infty} r^j g_j, \quad P_X(r) = \sum_{k=0}^{\infty} r^k f_k$$

を代入する。すると、

$$\sum_{j=0}^{\infty} j r^{j-1} g_j = a \left(\sum_{k=0}^{\infty} r^k f_k\right) \left(\sum_{j=0}^{\infty} j r^{j-1} g_j\right) + (a+b) \left(\sum_{j=0}^{\infty} r^j g_j\right) \left(\sum_{k=0}^{\infty} k r^{k-1} f_k\right)$$

が成り立つので両辺に  $r$  を掛けて比較すると、

$$\begin{aligned} xg_x &= a \sum_{k=0}^x f_k(x-k)g_{x-k} + (a+b) \sum_{k=0}^x kf_k g_{x-k} \\ &= af_0 xg_x + a \sum_{k=1}^x f_k(x-k)g_{x-k} + (a+b) \sum_{k=1}^x kf_k g_{x-k} \end{aligned}$$

式を再整理して、

$$g_x = \frac{1}{1-af_0} \sum_{k=1}^x \left(a + \frac{bk}{x}\right) f_k g_{x-k}$$

が得られる。

### 2.3 Panjer 逐次計算法の拡張

Panjer による  $(a, b, 0)$  クラス分布の拡張としては、Sundt and Jewell (1981) による  $(a, b, 1)$  クラスや Schröter (1991) による拡張が挙げられる。 $(a, b, 1)$  クラスとは、 $(a, b, 0)$  が  $q_0$  を与えた上で保険請求数の逐次計算法を提案しているのに対して、 $(a, b, 1)$  クラスでは  $q_0$  と  $q_1$  を与えた上で保険請求数の分布を求める手法である。また、Schröter クラスとは

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2}$$

を満たすクラスである。Panjer による逐次公式とほぼ同様の計算方法で  $g_x$  を計算できるが、請求数の密度は

$$g_x = \frac{1}{1-af_0} \sum_{j=1}^x \left\{ \left(a + \frac{bj}{x}\right) f_j + \frac{cj}{2x} f_j^{2*} \right\} g_{x-j}$$

となり  $f_x$  の畳み込みがあらかじめ計算されていなければ、 $g_x$  は逐次的にもとまらない。近年では、高次元での Panjer 逐次公式等の研究も行われているようである。

## 3 個別リスク集計モデル

本節では、個別リスクを集計したモデルを考える場合、どのようにモデリングするか考察を行いたい。

### 3.1 De Pril による逐次計算法

De Pril (1986) により、次のような次のような問題の数値計算法が考案された。次の1年間の死亡率が  $q_j$  ( $j = 1, \dots, J$ ) の人がある生命保険に  $i$  ( $i = 1, \dots, I$ ) 口入っている人を  $n_{ij}$  と書くことにする。このとき、1年間での保険請求数の分布を求めたい。すなわち、保険請求数を確率変数  $S$  で表すとすれば  $g_x = P[S = x]$  を計算したい。保険に  $i$  口入っていて、死亡率が  $q_j$  で表される個人による保険請求数の確率母関数は

$$P_{ij}(r) = 1 - q_j + q_j r^i$$



と書ける。このクラスに入っている人が  $n_{ij}$  人いたとすれば、各個人の保険請求の有無が独立ならば総保険請求数の確率母関数は

$$P_{ij}(r) = \prod_{i=1}^I \prod_{j=1}^J (1 - q_j + q_j r^i)^{n_{ij}} = \sum_{x=0}^{\infty} r^x g_x \quad (3.2)$$

と書ける。ここで、

$$\frac{d}{dr} \log P_S(r) = \frac{P'_S(r)}{P_S(r)} = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \frac{q_j i r^{i-1}}{1 - q_j + q_j r^i}$$

よって

$$\begin{aligned} P'_S(r) &= P_S(r) \sum_{i=1}^I \sum_{j=1}^J n_{ij} i \frac{q_j r^i}{1 - q_j + q_j r^i} \\ &= P_S(r) \sum_{i=1}^I \sum_{j=1}^J n_{ij} i \frac{q_j r^i}{1 - q_j} \left(1 + \frac{q_j r^i}{1 - q_j}\right) \\ &= P_S(r) \sum_{i=1}^I \sum_{j=1}^J n_{ij} i \sum_{k=1}^{\infty} (-1)^{k-1} \left(\frac{q_j}{1 - q_j}\right)^k r^{ik} \end{aligned}$$

が成り立つ。ここで、最後の計算は全ての  $i$  と  $j$  において

$$\left| \frac{q_j r^i}{1 - q_j} \right| < 1$$

がいえる場合にのみ成り立つ。通常は  $q_j$  はとても小さいので、この条件は成り立つ。ここで、

$$\begin{aligned} h(i, k) &= i(-1)^{K-1} \sum_{j=1}^J n_{ij} \left(\frac{q_j}{1 - q_j}\right)^k \quad (i = 1, \dots, I) \\ &= 0 \quad (\text{otherwise}) \end{aligned}$$

とおけば、

$$P'_S(r) = P_S(r) \sum_{i=1}^I \sum_{k=1}^{\infty} r^{ik} h(i, k)$$

であるので、(3.2) 式の右辺を代入して各項を比較すると

$$g_x = \frac{1}{x} \sum_{i=1}^x \sum_{k=1}^{\lfloor x/i \rfloor} g_{x-ik} h(i, k)$$

として計算ができる。ただしこの逐次式を開始するためには

$$g_0 = \prod_{i=1}^I \prod_{j=1}^J (1 - q_j)^{n_{ij}}$$

を用いる。通常は、 $g_0^K = g_0$  かつ

$$g_x^K = \frac{1}{x} \sum_{i=1}^{\min(x, I)} \sum_{k=1}^{\min(K, \lfloor x/i \rfloor)} g_{x-ik} h(i, k)$$

として  $g_x^K$  を計算する。  $K = 4$  程度で十分に精度を取ることが可能である。この際の計算誤差については De Pril (1988) を参照。

### 3.2 Kornyaの方法

De prilの方法は  $S$  の正確な分布を計算する方法であるのに対して、Kornyaの方法は  $S$  の分布を近似する方法である。この手法は Kornya (1983) により導入された。確率  $p_j = 1 - q_j$  を導入すると、確率変数  $S$  の母関数は

$$P_S(r) = \prod_{i=1}^I \prod_{j=1}^J (p_j + q_j r^i)^{n_{ij}}$$

と書くことができる。すると、

$$P_S(r) = \prod_{i=1}^I \prod_{j=1}^J \left(1 + \frac{q_j r^i}{p_j}\right)^{n_{ij}} \left(1 + \frac{q_j}{p_j}\right)^{-n_{ij}}$$

である。いま、条件  $q_j < 1/2$  と  $|q_j r^i / p_j| < 1$  を仮定する。

$$\begin{aligned} \log P_S(r) &= \sum_{i=1}^I \sum_{j=1}^J n_{ij} [\log(1 + \frac{q_j r^i}{p_j}) - \log(1 + \frac{q_j}{p_j})] \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \sum_{i=1}^I \sum_{j=1}^J n_{ij} [(\frac{q_j r^i}{p_j})^k - (\frac{q_j}{p_j})^k] \end{aligned}$$

そこで、

$$S_k(r) = \sum_{i=1}^I \sum_{j=1}^J n_{ij} [(\frac{q_j r^i}{p_j})^k - (\frac{q_j}{p_j})^k]$$

を導入して、

$$Q_S(r) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} S_k(r), \quad Q_K(r) = \sum_{k=1}^K \frac{(-1)^{k+1}}{k} S_k(r) \quad (3.3)$$

を定義する。 $P_S(r)$  と  $P_S(r)$

$$P_S(r) = \exp\{Q_S(r)\} = \sum_{x=0}^{\infty} r^x g_x, \quad P_K(r) = \exp\{Q_K(r)\} = \sum_{x=0}^{\infty} r^x g_x^{(K)} \quad (3.4)$$

として、 $g_x^{(K)}$  を求めたい。これにより、 $\sum_{x=0}^y g_x$  の近似として  $\sum_{x=0}^y |g_x^{(K)}|$  を計算したい。まず、 $Q_K(r) = \sum_{x=0}^{\infty} r^x b_x^{(K)}$  とすると、

$$Q_K(r) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \sum_{i=1}^I \sum_{j=1}^J n_{ij} [(\frac{q_j r^i}{p_j})^k - (\frac{q_j}{p_j})^k]$$

と書けることから、係数の比較により

$$\begin{aligned} b_0^{(K)} &= \sum_{k=1}^K \frac{(-1)^k}{k} \sum_{i=1}^I \sum_{j=1}^J n_{ij} (\frac{q_j}{p_j})^k \\ b_x^{(K)} &= \sum_{k=\lfloor x/I \rfloor, k \leq x}^{\min\{K, x\}} \frac{(-1)^{k+1}}{k} \sum_{i=1}^I \sum_{j=1}^J n_{x/k, j} (\frac{q_j}{p_j})^k \end{aligned}$$

ここで  $k|x$  は  $k$  は  $x$  の約数であることを意味している。また、 $[x/I]$  は  $x/I$  より大きい最小の整数であるとする。ところで、 $g_x^{(K)}$  は

$$g_x^{(K)} = \frac{1}{x} \sum_{j=1}^x j b_j^{(K)} g_{x-j}^{(K)} \quad (3.5)$$

という関係があるので<sup>3</sup>、

$$P_K(0) = g_0^{(K)} = \exp\{Q_K(0)\} = \exp\{h_0^{(K)}\}$$

からスタートして式 (3.5) を用いて計算していくことが可能である。すると、

$$g_x^{(K)} = \frac{1}{x} \sum_{j=1}^{\min\{x, IK\}} j b_j^{(K)} g_{x-j}^{(K)}$$

が成り立つ。なお、通常  $K = 4$  程度で十分な精度で  $g_x$  を近似できる。なお、この逐次手法を用いた場合の近似誤差の上限は Kornya (1983) で計算されている。

また、個別リスクモデルの他の計算法としては De Pril and Dhaene (1992) による、2重2項分布 (compund binomial distribution) を2重ポアソン分布 (compund Poisson distribution) で近似する手法等が挙げられる。

## 4 破産理論入門

### 4.1 無限時間での破産確率

時刻  $n = 1, 2, 3, \dots$  における保険家の留保額過程 ('surplus process') を

$$U_d(n) = u + n - \sum_{i=1}^n Z_i$$

と書くことにする。ここで、 $u = U_d(0)$  は初期時点での保険家の内部留保の総額、 $Z_i$  で時刻  $i$  における保険家への支払い請求額とする。ここで、 $Z_i$  ( $i = 1, \dots$ ) は正の整数の値を取る独立同一分布に従う確率変数であり、 $E[Z_i] < 1$  を満たすものとする。このモデルでは簡単化のために保険金による収入は毎期1であるとしてモデルを作成している。保険家の破産時刻を

$$T_{d,u} = \min\{n \geq 1 : U_d(n) \leq 0\}$$

で定義する。ただし、全ての時刻  $n$  において  $U_d(n) > 0$  が成り立つとき  $T_{d,u} = \infty$  と書くことにする。さて、将来時点で破産が起こる確率

$$\psi_d(u) = Pr[T_{d,u} < \infty] = Pr[u + n - \sum_{i=1}^n Z_i \leq 0, \text{ for some } n = 1, 2, 3, \dots]$$

<sup>3</sup>  $P'_K(r) = Q'_K(r)P_K(r)$  に (3.3) と (3.4) を代入して求まる。

を求める方法を知りたい。もし、時刻1において留保額過程が  $U_d(1) > 0$  であれば、この位置での破産確率は

$$Pr[U_d(1) + n - 1 - \sum_{i=2}^n Z_i, \text{ for some } n = 2, 3, 4, \dots]$$

であり  $\psi(U_d(1))$  とかける。よって、

$$\psi_d(u) = \sum_{j=0}^u h_j \psi_d(u+1-j) + 1 - H(u) = \sum_{r=1}^{u+1} h_{u+1-r} \psi_d(r) + 1 - H(u)$$

これより

$$\begin{aligned} \sum_{u=0}^w \psi_d(u) &= \sum_{u=0}^w \sum_{r=1}^{u+1} h_{u+1-r} \psi_d(r) + \sum_{u=0}^w [1 - H(u)] \\ &= \sum_{r=1}^{w+1} \psi_d(r) \sum_{u=r-1}^w h_{u+1-r} + \sum_{u=0}^w [1 - H(u)] \\ &= \sum_{r=1}^w \psi_d(r) H(w+1-r) + \psi_d(w+1) h_0 + \sum_{u=0}^w [1 - H(u)] \end{aligned}$$

である。これから、

$$\psi_d(w+1) h_0 = \psi_d(w) - \sum_{r=1}^w \psi_d(r) [1 - H(w+1-r)] - \sum_{u=0}^w [1 - H(u)]$$

が言える。特に、 $\psi_d(0) = E[Z_1]$  である。これを示す。初期留保額が0であるとする。このとき、破産のモデルを考える。破産したときの赤字額が  $y$  である確率を  $g_d(y)$  とする。いま、初期の留保額を  $u$  とし、初めて  $u$  を下回った時点での留保額が  $u-y$  である確率を考えることで

$$\psi_d(u) = \sum_{y=0}^{u-1} g_d(y) \psi_d(u-y) + \sum_{y=u}^{\infty} g_d(y)$$

が求まる。また、

$$\psi_d(0) = \sum_{y=0}^{\infty} g_d(y)$$

が成り立つ。よって、

$$\begin{aligned} \psi_d(u) &= \sum_{y=0}^{u-1} g_d(y) \psi_d(u-y) + \psi_d(0) - \sum_{y=0}^{u-1} g_d(y) \\ &= \psi_d(0) + \sum_{y=1}^u g_d(u-y) \psi_d(y) - \sum_{y=0}^{u-1} g_d(y) \end{aligned}$$

これと(?)を比較して、

$$g_d(y) = 1 - H(y)$$

を得る。これより

$$\psi_d(0) = \sum_{y=0}^{\infty} (1 - H(y)) = E[Z_1]$$

が成り立つ。よって示された。また、

$$\psi_d(u) = \sum_{y=0}^{u-1} [1 - H(y)] \psi_d(u - y) + \sum_{y=u}^{\infty} [1 - H(u)]$$

も成り立つ。有限時間での破産確率の評価法は De Vylder and Goovaerts (1988) を参照。

## 4.2 Lundberg の不等式

いま、定数  $R_d$  を

$$E[\exp\{r(Z_1 - 1)\}] = 1$$

の解であるとする。すなわち  $E[\exp\{R_d(Z_1 - 1)\}] = 1$  である。

$$g(r) = E[\exp\{r(Z_1 - 1)\}]$$

とおくと、

$$g'(0) = E[Z_1] - 1 < 0$$

である。また、 $g''(r) > 0$  かつ  $\lim_{r \rightarrow \infty} g(r) = \infty$  が成り立つ<sup>4</sup>。このことから、関数  $g(r)$  は  $r = 0$  から減少し、0 と  $R_d$  の間の点で減少が止まり、増加に転じる。

Lundberg の不等式は

$$\psi_d(u) \leq e^{-R_d u}$$

と書ける。それを示すには、時刻  $t$  までに破産が起こる確率を  $\psi_d(u, t)$  と書くことにすると、

$$\psi_d(u, t) < e^{-R_d u}$$

が全ての  $t$  で成り立つことを示せばよい。まず、

$$\psi_d(u, 1) = \sum_{k=u+1}^{\infty} h_k$$

が成り立つ。

$$\begin{aligned} \psi_d(u, 1) &= \sum_{k=u+1}^{\infty} e^{-R_d(u+1-k)} h_k \\ &\leq \sum_{k=0}^{\infty} e^{-R_d(u+1-k)} h_k \\ &\leq e^{-R_d u} \sum_{k=u+1}^{\infty} e^{R_d(k-1)} h_k = e^{-R_d u} \end{aligned}$$

<sup>4</sup> ここは

$$g(r) = \sum_{k=0}^{\infty} e^{r(k-1)} h_k > \sum_{k=2}^{\infty} e^{r(k-1)} h_k > e^r \sum_{k=2}^{\infty} h_k = e^r (1 - H(1)) \rightarrow \infty \quad (r \rightarrow \infty)$$

から示される。

ここで、最後の等式では  $R_d$  の性質

$$E[\exp\{R_d(Z_1 - 1)\}] = 1$$

を用いていることに注意する。数学的帰納法を用いて Lundberg の不等式を示す。ここで、 $t$  において  $\psi_d(u, t+1) \leq e^{-R_d u}$  が成り立つと仮定する。よって、数学的帰納法の仮定から

$$\begin{aligned} \psi_d(u, t+1) &= \psi_d(u, 1) + \sum_{k=0}^u h_k \psi_d(u+1-k, t) \\ &\leq \sum_{k=u+1}^{\infty} h_k + \sum_{k=0}^u h_k e^{-R_d(u+1-k)} \\ &\leq \sum_{k=0}^{\infty} h_k e^{-R_d(u+1-k)} = e^{-R_d u} \end{aligned}$$

が成り立ち、Lundberg の不等式が成り立つ。

## Reference

- [1] Bühlmann, H., An economic premium principle, *ASTIN Bulletin*, **11**, 52-60, 1980.
- [2] De Pril, N., On exact computation of aggregate claims distribution of a life insurance portfolio, *ASTIN Bulletin*, **16**, 109-112, 1986.
- [3] De Pril, N., Improved approximations for the aggregate claims distribution of a life insurance portfolio, *Scandinavian Actuarial Journal*, 61-68, 1988.
- [4] De Pril, N. and Dhaene, J. Error bounds for compound Poisson approximations of individual risk model, *ASTIN Bulletin*, **22**, 135-148, 1986.
- [5] De Vylder, F. and Goovaerts, M. J. Recursive calculation of finite time survival probabilities, *Insurance Mathematics & Economics*, **7**, 1-8, 1988.
- [6] Dickson, D. C. M., *Insurance Risk and Ruin*, Cambridge, 2005.
- [7] Gerber, H. U. and Pafumi, G., Utility functions: from risk theory to finance, *North American Actuarial Journal*, **2**, 74-100, 1998.
- [8] Klugman, S.A., Panjer, H. H. Panjer, H. H. and Wilmot, G.E. *Loss Models - from data to decisions (2 nd. ed.)*, John Wiley, 2004.
- [9] Kornya, P.S., Distribution of aggregate claims in the individual risk theory model, *Transactions of Society of Actuaries*, **2**, 74-100, 1998.
- [10] Panjer, H. H., Recursive evaluation of family of compound distributions, *ASTIN Bulletin*, **12**, 21-26, 1981.

- [11] Schröter, K. J., On a family of counting distributions and recursions for related compound distributions, *Scandinavian Actuarial Journal*, 161-175, 1991.
- [12] Sundt, B. and Jewell, W.S., Further results on recursive evaluation of compound distributions, *ASTIN Bulletin*, 12, 27-39, 1981.
- [13] Wang S., Insurance pricing & increased limits ratemaking by proportional hazard transform, *Insurance: Mathematics & economics*, 17, 43-54, 1995.
- [14] Young, V. and Zariphoulou, T., Pricing Dynamic Insurance Risks Using the Principle of Equivalent Utility, *Scandinavian Actuarial Journal*, 246-279, 2002.
- [15] Zariphoulou, T., A solution approach to valuation with unhedgeable risks *Finance and Stochastics*, , -, 1999.

# 保険リスクと破産理論 (II)

室井 芳史

大阪大学金融・保険教育研究センター  
e-mail:muroi@sigmath.es.osaka-u.ac.jp

**概要:** 本稿では、80年代および90年代に研究された破産理論の簡単な解説を行う。この時代の保険理論、特に破産理論は Gerber (1979) の内容を発展させながら研究してきたような印象を抱く。本稿でも、保険リスクと破産理論 (I) と同様に Dickson (2005) に従い、破産理論について言及を行う。現在、破産理論で研究されてきた内容は多岐にわたっているが、本稿では破産確率の計算や破産時の赤字額、破産までの時間について記述を行う。破産理論のほかの理論や、破産理論にまつわる数値計算法、再保険などについては他の機会に譲りたい。

**キーワード:** 破産理論、破産時の赤字、破産までの時間

## 1 古典的破産理論

### 1.1 モデルの設定

本節では、保険会社の剰余金が古典的リスク過程に従う仮定の下で保険会社が破産する確率を計算する。この話題は、数理保険学では中心的な話題であり、古くから議論が行われている。破産確率の上限である Lundberg 不等式の証明や、保険金支払い額が指数分布に従うと仮定したときの保険会社の倒産確率などについて議論を行う。

古典的破産理論において、保険会社の剰余金は、初期時刻  $t = 0$  における剰余金の額、時刻  $t$  までに受け取った収入、時刻  $t$  までに支払った請求額の3つの要素で決まる。このモデルでは、この3つの中で確率的に決まるのは時刻  $t$  までに支払った請求額であり、これを  $S(t)$  と書くことにする。 $N(t)$  で時刻  $[0, t]$  において請求が行われた回数を表すとする。古典モデルでは  $N(t)$  は強度が  $\lambda$  のポアソン過程であるとしている。各請求は一回目の請求から  $X_1, X_2, \dots$  と書け独立同一分布に従うとする。確率変数  $X$  の確率密度関数は  $f(x)$  であらわされるものとする。すると  $S(t)$  は

$$S(t) = \sum_{i=1}^{N(t)} X_i$$

と書け、複合ポアソン過程となっている。ただし、 $N(t) = 0$  の場合は  $S(t) = 0$  と定義することとする。初期時刻  $t = 0$  における剰余金の額を  $u$ 、単位時間当たりの保険料収入を



$c$ とすると、剰余金の総額は

$$U(t) = u + ct - S(t)$$

と書くことができる。

いま、 $X_1$ の積率母関数  $M_X(r)$  が存在するとして、この積率母関数はある  $0 < \gamma \leq \infty$  について、任意の  $r < \gamma$  で有限の値をとり、

$$\lim_{r \rightarrow \gamma^-} M_X(r) = \infty$$

を満たすものとする。これらの条件の下で無限時間における破産確率

$$\phi(u) = Pr[U(t) < 0 \text{ for some } t > 0]$$

を計算したい。すなわち剰余金が負になる場合を破産とみなしその確率を計算する。本稿では特に断らない限り、 $m_1$  を確率変数  $X_1$  の平均値として

$$c > \lambda m_1$$

を満たすとする。また、単位時間当たりの保険料収入は  $c = (1 + \theta)\lambda m_1$  と書くことができるものとする。

## 1.2 破産確率の計算

古典的破産理論において方程式

$$\lambda M_X(r) - \lambda - cr = 0 \tag{1.1}$$

の解を調整係数 (adjustment coefficient) と呼び  $R$  と記する。すなわち

$$\lambda + cR = \lambda M_X(R)$$

を満たす。 $c = (1 + \theta)\lambda m_1$  とかけるので、 $R$  は  $\lambda$  と独立な表現を持っている。調整係数  $R$  は方程式 (1.1) の唯一の解であり、これは以下のように証明することが可能である。関数  $g(r)$  を

$$g(r) = \lambda M_X(r) - \lambda - cr$$

とおく。 $g(r)$  は以下のような性質を持つ。

$$\begin{aligned} \frac{d}{dr}g(r)|_{r=0} &= \lambda \frac{d}{dr}M_X(r)|_{r=0} - c \\ &= \lambda m_1 - c \quad (< 0) \end{aligned}$$

より原点付近で減少関数となる。さらに

$$\frac{d^2}{dr^2}g(r) = \lambda \frac{d^2}{dr^2}M_X(r) = \lambda \int_0^\infty x^2 e^{rx} f(x) dx > 0$$

より  $g'(r)$  は単調増加関数である。もし、 $\gamma < \infty$  ならば容易に  $\lim_{r \rightarrow \gamma^-} g(r) = \infty$  であることが分かる。このことから  $g(r) = 0$  は唯一解を持つことが分かる。一方、 $\gamma = \infty$  であれば、ある正数  $\epsilon$  と  $p$  が存在し、

$$P[X_1 > \epsilon] = p > 0$$

とすることができる。よって、

$$M_X(r) = \int_0^{\infty} e^{rx} f(x) dx \geq \int_{\epsilon}^{\infty} e^{rx} f(x) dx \geq e^{r\epsilon} p$$

が成り立つ。これから、

$$\lim_{r \rightarrow \infty} g(r) \geq \lim_{r \rightarrow \infty} (\lambda e^{r\epsilon} p - \lambda - cr) = \infty$$

よって、どちらの場合でも

$$\lim_{r \rightarrow \gamma^-} g(r) = \infty$$

が成り立ち、方程式  $g(r) = 0$  は唯一解を持つことが分かる。

### 1.3 Lundberg の不等式

離散モデルで得られたのと同じ形の不等式

$$\psi(r) \leq \exp\{-Ru\} \tag{1.2}$$

が連続モデルでも成り立つ。これを示したい。 $n$  回以下の保険金請求で倒産が起こる確率を  $\psi_n(u)$  とおくと、任意の整数  $n$  で

$$\psi_n(u) \leq e^{-Ru}$$

が成り立つことを示せば、極限  $n \rightarrow \infty$  を取ることで不等式 (1.2) が示される。この不等式を数学的帰納法により示したい。さて、 $n$  まではすでに示されているとしよう。ここで、 $n+1$  番目までの保険金請求によって倒産が起こる場合は

- (i) 初めての保険金請求で  $x > u + ct$  を満たすような請求額  $x$  を求められてすぐに倒産。
- (ii) 初めての保険金請求では請求額  $x$  が発生するものの、倒産せずに保険会社の剰余金は  $u + ct - x (> 0)$  となる。この剰余金の額を初期時点での剰余金と見直して、もう一度、古典リスク過程を考える。その後  $n$  回以下の保険金請求により倒産が起こる。

のどちらかである。保険金請求の起こる回数は強度  $\lambda$  のポアソン過程に従っているので、請求が起こるまでの間隔はパラメータ  $\lambda$  の指数分布に従う。よって

$$\psi_{n+1}(u) = \int_0^{\infty} \lambda e^{-\lambda t} \int_{u+ct}^{\infty} f(x) dx dt + \int_0^{\infty} \lambda e^{-\lambda t} \int_0^{u+ct} f(x) \psi_n(u + ct - x) dx dt$$

よって、帰納法の仮定より

$$\psi_{n+1}(u) \leq \int_0^\infty \lambda e^{-\lambda t} \int_{u+ct}^\infty f(x) dx dt + \int_0^\infty \lambda e^{-\lambda t} \int_0^{u+ct} f(x) e^{-R(u+ct-x)} dx dt$$

が成り立つが、 $x \geq u + ct$  であれば

$$\exp\{-R(u + ct - x)\} \geq 1$$

が言えるので

$$\int_{u+ct}^\infty f(x) dx \leq \int_{u+ct}^\infty f(x) e^{-R(u+ct-x)} dx$$

が成り立つ。よって、

$$\begin{aligned} \psi_{n+1}(u) &\leq \int_0^\infty \lambda e^{-\lambda t} \int_0^\infty f(x) e^{-R(u+ct-x)} dx dt \\ &\leq e^{-Ru} \int_0^\infty \lambda e^{-(\lambda+cR)t} \int_0^\infty e^{Rx} f(x) dx dt \\ &\leq e^{-Ru} \int_0^\infty \lambda e^{-(\lambda+cR)t} M_X(R) dt \end{aligned}$$

ここで、方程式 (1.1) より  $\lambda + cR = \lambda M_X(R)$  が成り立つので

$$\psi_{n+1}(u) \leq \exp(-Ru)$$

が言えた。よって、 $n+1$  のときも成立することが分かった。あとは、 $n=1$  の場合を示すばよい。

$$\begin{aligned} \psi_1(u) &= \int_0^\infty \lambda e^{-\lambda t} \int_{u+ct}^\infty f(x) dx dt \\ &\leq \int_0^\infty \lambda e^{-\lambda t} \int_{u+ct}^\infty f(x) e^{-R(u+ct-x)} dx dt \\ &\leq \int_0^\infty \lambda e^{-\lambda t} \int_0^\infty f(x) e^{-R(u+ct-x)} dx dt = e^{-Ru} \end{aligned}$$

よって題意が示された。

## 1.4 生存確率

次に、初期時点での剰余金を  $u$  としたときの保険会社の生存確率  $\phi(u) = 1 - \psi(u)$  の計算法について考える。生存確率  $\phi(u)$  は積分方程式

$$\begin{aligned} \phi(u) &= \int_0^\infty \lambda e^{-\lambda t} \int_0^{u+ct} f(x) \phi(u + ct - x) dx dt \\ &= \frac{1}{c} \int_u^\infty \lambda e^{-\lambda(s-u)/c} \int_0^s f(x) \phi(s - x) dx ds \\ &= \frac{\lambda}{c} e^{\lambda u/c} \int_u^\infty e^{-\lambda s/c} \int_0^s f(x) \phi(s - x) dx ds \end{aligned}$$

を満たす。両辺を  $u$  で微分して

$$\begin{aligned}\frac{d}{du}\phi(u) &= \frac{\lambda^2}{c^2}e^{\lambda u/c} \int_u^\infty e^{-\lambda s/c} \int_0^s f(x)\phi(s-x)dx ds - \frac{\lambda}{c} \int_0^u f(x)\phi(u-x)dx \\ &= \frac{\lambda}{c}\phi(u) - \frac{\lambda}{c} \int_0^u f(x)\phi(u-x)dx\end{aligned}\quad (1.3)$$

ここで、簡単化のため保険金支払い額が母数  $\alpha$  の指数分布に従う場合を考える。すなわち、確率分布関数が  $F(x) = 1 - e^{-\alpha x}$  である場合について考える。

$$\begin{aligned}\frac{d}{du}\phi(u) &= \frac{\lambda}{c}\phi(u) - \frac{\lambda}{c} \int_0^u \alpha e^{-\alpha x} \phi(u-x)dx \\ &= \frac{\lambda}{c}\phi(u) - \frac{\lambda}{c} e^{-\alpha u} \int_0^u \alpha e^{\alpha x} \phi(x)dx\end{aligned}\quad (1.4)$$

再度微分して、

$$\frac{d^2}{du^2}\phi(u) = \frac{\lambda}{c} \frac{d}{du}\phi(u) + \frac{\alpha^2 \lambda}{c} e^{-\alpha u} \int_0^u e^{-\alpha x} \phi(x)dx - \frac{\alpha \lambda}{c} \phi(u)\quad (1.5)$$

式(1.4)と(1.5)により、両辺から積分を消して常微分方程式

$$\frac{d^2}{du^2}\phi(u) + \left(\alpha - \frac{\lambda}{c}\right) \frac{d}{du}\phi(u) = 0$$

を得る。よって、解は

$$\phi(u) = a_0 + a_1 e^{-(\alpha - \frac{\lambda}{c})u}$$

とかける。ただし、 $a_0$ と $a_1$ はある定数であるが、 $\lim_{u \rightarrow \infty} \phi(u) = 1$ および $\phi(0) = 1 - \psi(0)$ より $a_0 = 1$ かつ $a_1 = -\psi(0)$ である。すなわち

$$\phi(u) = 1 - \psi(0)e^{-(\alpha - \lambda/c)u}$$

である。よって $\psi(0)$ が求まれば、生存確率の計算ができたことになる。(1.3)の両辺を積分すると

$$-\psi(0) = \frac{\lambda}{c} \int_0^\infty \psi(u)du - \frac{\lambda}{c} \int_0^\infty \int_0^u f(x)\psi(u-x)dx du - \frac{\lambda}{c} \int_0^\infty (1 - F(u))du$$

ここで、第2項目の積分は

$$\begin{aligned}\int_0^\infty \int_0^u f(x)\psi(u-x)dx du &= \int_0^\infty \int_x^\infty \psi(u-x)du f(x)dx \\ &= \int_0^\infty \int_0^\infty \psi(y)dy f(x)dx \\ &= \int_0^\infty \psi(y)dy\end{aligned}$$

であるので、第1項目の積分と打ち消しあう。よって、

$$\psi(0) = \frac{\lambda}{c} \int_0^\infty (1 - F(u))du = \frac{\lambda m_1}{c}\quad (1.6)$$

この結果は、 $F$ の分布によらないことに注意しよう。よって、請求額の分布が指数分布に従うときには

$$\phi(u) = 1 - \frac{\lambda}{\alpha c} \exp\{-(\alpha - \lambda/c)u\}$$

を得る。いま、請求額の分布が母数 $\alpha$ の指数分布に従うときに、調整係数は $R = \alpha - \lambda/c$ なので

$$\psi(u) = \frac{\lambda}{\alpha c} e^{-Ru} = \psi(0)e^{-Ru} (\leq e^{-Ru})$$

が成り立つ。請求額の分布が指数分布に従う場合以外の、一般の場合はPanjer(1981)の逐次公式アルゴリズムを使うことで、破産確率を計算することができる。これらの話題については、Panjer (1986) や Dufresne and Gerber (1989) および Dickson and Waters (1991) を参照のこと。

## 2 破産時の赤字額

この章では、保険会社の剰余金が古典的リスク過程に従うとしたときに、保険会社が倒産したらどの程度の赤字額が出るかについて議論を行いたい。これらの話題はGerber, Goovaerts and Kaas (1987) や Dufresne and Gerber (1988) をはじめとして、数理保険学では多くの研究がなされた話題である。

$u$ を起点とする古典的リスク過程が上側の境界 $b (> u)$ に到達することなく、破産が起こる確率を $\xi(u, b)$ と書くことにする。古典的リスク過程が上側の境界に到達すると、そこで吸収されてしまうときに、破産が起こる前に吸収が起こる確率(すなわち、破産するより先に古典的リスク過程が上側の境界に到達する確率)を $\chi(u, b)$ と書くことにする。 $\phi(u) = \chi(u, b)\phi(b)$ が成り立つので

$$\chi(u, b) = \frac{1 - \psi(u)}{1 - \psi(b)}$$

が成り立つ。同じく $\psi(u) = \xi(u, b) + \chi(u, b)\psi(b)$ がいえるので

$$\xi(u, b) = \psi(u) - \frac{1 - \psi(u)}{1 - \psi(b)}\psi(b) = \frac{\psi(u) - \psi(b)}{1 - \psi(b)}$$

が言える。

新たに確率過程 $\{L(t)\}$ を $L(t) = S(t) - ct$ によって定義する。この確率過程を総損出過程と呼ぶ。総損失過程の最大値を確率変数 $L$ と書くとすると $L$ と $\phi$ は

$$\phi(u) = Pr[U(t) \geq 0 \text{ for all } t > 0] = Pr[L(t) \leq u \text{ for all } t > 0] = Pr[L \leq u]$$

の関係を持っている。いま、積分微分方程式(1.3)にラプラス変換を行うと

$$s\phi^*(s) - \phi(0) = \frac{\lambda}{c}\phi^*(s) - \frac{\lambda}{c}f^*(s)\phi^*(s)$$

を得る<sup>1</sup>。よって、

$$\phi^*(s) = \frac{c\phi(0)}{cs - \lambda(1 - f^*(s))}$$

である。このことから、請求額の分布が一般の確率分布のときはこの式を反転すれば破産確率を計算できる。さて、 $L$ の積率母関数を計算すると

$$\begin{aligned} L^*(s) &= E[e^{-sL}] = \int_0^\infty e^{-su} d\phi(u) \\ &= \phi(0) + \int_0^\infty e^{-su} \frac{d}{du} \phi(u) du \\ &= s\phi^*(s) \\ &= \frac{cs\phi(0)}{cs - \lambda(1 - f^*(s))} \end{aligned} \quad (2.7)$$

となる。

起点が  $u$  の古典的リスク過程が  $u$  を下回る確率は  $\psi(0)$  である。初めて  $u$  を下回ったときに、 $u$  から下回った大きさ  $L_1$  は確率変数である。このときの古典的リスク過程は、例えば、水準  $u - l_1$  にあるとする。次に水準  $u - l_1$  を下回る確率はまたしても  $\psi(0)$  である。このとき、 $u - l_1$  から下回った大きさ  $L_2$  も  $L_1$  と独立同一分布の確率変数である。このようにして、 $L_1, L_2, \dots$  を決めていき、水準の下回り（記録の更新）が  $n$  回で終わる確率は  $\psi(0)^n \phi(0)$  である。これは、幾何分布である。記録の更新が起こる回数も確率変数であり  $N$  と書くことにする。すると

$$L = \sum_{i=1}^N L_i$$

と書くことができる。すなわち、 $L$  は複合幾何分布に従う。 $L_1$  の確率分布関数を  $K(x) = Pr[L_1 \leq x]$  と書き、密度関数を  $k(x)$  と書くことにする。よって  $k^*(s) = E[e^{-sL_1}]$  とおくと

$$E[e^{-sL}] = \frac{\phi(0)}{1 - \psi(0)k^*(s)}$$

すると、式 (2.7) から

$$\frac{\phi(0)}{cs - \lambda(1 - f^*(s))} = \frac{\phi(0)}{1 - \psi(0)k^*(s)}$$

がいえる。 $\psi(0) = \lambda m_1 / c$  なので直前の式より

$$k^*(s) = \frac{1}{m_1 s} (1 - f^*(s))$$

がわかる。反転させて

$$k(x) = \frac{1}{m_1} (1 - F(x))$$

<sup>1</sup> 関数  $h(u)$  のラプラス変換を  $h^*(s)$  と書く。すなわち

$$h^*(s) = \int_0^\infty e^{-su} h(u) du$$

と書ける。

が言えた。ここで、停止時刻  $T_u$  を破産時刻とする。すなわち

$$T_u = \inf\{t : U(t) < 0\}$$

とおく。ただし、もし、 $U(t) \geq 0$  ならば  $T_u = \infty$  とする。よって、 $\psi(u) = Pr[T_u < \infty]$  である。破産時の赤字額が  $y$  以下である確率を  $G(u, y)$  とおく。つまり、

$$G(u, y) = Pr[T_u < \infty \text{ and } U(T_u) \geq -y]$$

とおくと

$$\lim_{y \rightarrow \infty} G(u, y) = \psi(u)$$

なので、

$$\frac{G(u, y)}{\psi(u)} = Pr[|U(T_u)| \leq y | T_u < \infty]$$

は確率分布関数になっている。ここで、

$$g(u, y) = \frac{\partial}{\partial y} G(u, y)$$

を導入する。初めての記録が起る確率は  $\psi(0)$  であったので

$$k(y) = \frac{d}{dy} \frac{G(0, y)}{\psi(0)} = \frac{g(0, y)}{\psi(0)}$$

と書くことができる。このことと (1.6) より

$$g(0, y) = \psi(0)k(y) = (1 - \phi(0)) \frac{1}{m_1} (1 - F(y)) = \frac{\lambda m_1}{c} \frac{1}{m_1} (1 - F(y)) = \frac{\lambda}{c} (1 - F(y))$$

が成り立つ。よって、初期時点での保険会社の剰余金が 0 であったときの破産時の赤字額が求まった。

次に、一般の場合の破産時の赤字額を計算しよう。破産時の赤字額が高々  $y$  であるとは

(1) 古典的リスク過程が初めて  $u$  を下回ったときに  $u - x$  まで下落し、その後、破産時の赤字額は高々  $y$  である。

(2) 古典的リスク過程が初めて  $u$  を下回ったときに破産が発生し、破産時の赤字額は高々  $y$  である。

のどちらかである。よって、

$$\eta(u, y) = \int_u^{u+y} k(x) dx = K(u+y) - K(u)$$

を導入すると

$$\begin{aligned} G(u, y) &= \int_0^u g(0, x) G(u-x, y) dx + \int_u^{u+y} g(0, x) dx \\ &= \psi(0) \int_0^u k(x) G(u-x, y) dx + \psi(0) \eta(u, y) \end{aligned}$$

が成り立つ。両辺をラプラス変換してやり

$$G^*(s, y) = \int_0^\infty e^{-su} G(u, y) du, \quad \text{and} \quad \eta^*(s, y) = \int_0^\infty e^{-sy} \eta(u, y) du$$

とおけば

$$G^*(s, y) = \frac{\psi(0)\eta^*(s, y)}{1 - \psi(0)k^*(s)} \quad (2.8)$$

が言える。請求額の確率分布が指数分布に従うときはラプラス変換を簡単に反転することができる。すなわち  $f(x) = \alpha e^{-\alpha x}$  とおくと、 $k(x) = \alpha e^{-\alpha x}$  なので

$$\eta(u, y) = e^{-\alpha u} - e^{-\alpha(u+y)}$$

が成り立つので

$$\eta^*(u, y) = \frac{1 - e^{-\alpha y}}{s + \alpha}$$

である。このことから

$$G^*(s, y) = \frac{\psi(0)(1 - e^{-\alpha y})}{s + \alpha - \psi(0)\alpha} = \frac{\lambda (1 - e^{-\alpha y})}{\alpha c s + \alpha - \lambda/\alpha}$$

が成り立つ。反転して

$$G(u, y) = \frac{\lambda}{\alpha c} e^{-(\alpha - \lambda/c)u} (1 - e^{-\alpha y}) = \psi(u)(1 - e^{-\alpha y})$$

が成り立つ。

一方、保険金支払いが一般の確率分布に従う場合においては (1.8) より、

$$G^*(s, y) = \frac{\psi(0)}{\phi(0)} \eta^*(s, y) L^*(s)$$

が成り立つ。これを反転させて

$$\begin{aligned} G(u, y) &= \frac{\psi(0)}{\phi(0)} \int_0^u \eta(u-x, y) d\phi(x) \\ &= \frac{\psi(0)}{\phi(0)} \int_0^u \eta(K(u-x+y) - K(u-x)) d\phi(x) \\ &= \psi(0)(K(u+y) - K(u)) + \frac{\psi(0)}{\phi(0)} \int_0^u (K(u-x+y) - K(u-x)) \phi'(x) dx \end{aligned}$$

が成り立つ。ところで、

$$\phi(u) = \phi(0) + \sum_{n=1}^{\infty} \psi(0)^n \phi(0) K^{n*}(u)$$

から、 $u > 0$  において

$$\phi'(u) = \sum_{n=1}^{\infty} \psi(0)^n \phi(0) k^{n*}(u)$$



であることと、

$$K(u-x+y) - K(u-x) = \int_{u-x}^{u-x+y} k(z) dz$$

より

$$G(u, y) = \psi(0)(K(u+y) - K(u)) + \sum_{n=1}^{\infty} \psi(0)^{n+1} \int_0^u k^{n*}(x) \int_{u-x}^{u-x+y} k(z) dz dx$$

となる。

### 3 破産までの時間

本節では、保険会社の剰余金が古典的リスク過程に従う場合に保険会社の破産時刻  $T_u$  の確率分布およびモーメントを求める。これらの話題は、Gerber and Shiu (1998) や Dreik and Wilmot (2003) で議論されてきた話題であり、本節でも保険会社の剰余金が古典的リスク過程に従うと仮定した場合を考察し、破産時刻の確率密度関数などをラプラス変換を用いて解析する手法について解説を行う。

まず、破産時刻  $T_u$  のラプラス変換

$$\varphi(u, \delta) = E[e^{-\delta T_u} 1_{T_u < \infty}]$$

を計算する。ここで、 $\delta$  は正の値をとるパラメータである。前節までと同様の議論により  $\varphi$  は積分方程式

$$\varphi(u, \delta) = \int_0^{\infty} \lambda e^{-\lambda t} e^{-\delta t} \int_0^{u+ct} f(x) \varphi(u+ct-x, \delta) dx dt + \int_0^{\infty} \lambda e^{-\lambda t} e^{-\delta t} \int_{u+ct}^{\infty} f(x) dx dt$$

を満たす。 $s = u + ct$  を代入して

$$\varphi(u, \delta) = \frac{\lambda}{c} \int_u^{\infty} e^{-(\lambda+\delta)(s-u)/c} \int_0^s f(x) \varphi(s-x, \delta) dx ds + \frac{\lambda}{c} \int_u^{\infty} e^{-(\lambda+\delta)(s-u)/c} \int_s^{\infty} f(x) dx ds$$

が成り立つ。両辺を  $u$  で微分して

$$\frac{\partial}{\partial u} \varphi(u, \delta) = \frac{\lambda + \delta}{c} \varphi(u, \delta) - \frac{\lambda}{c} \int_0^u f(u-x) \varphi(x, \delta) dx - \frac{\lambda}{c} (1 - F(u))$$

ここでは、簡単化のために特に保険請求額の分布が指数分布に従う場合について考える。すなわち  $f(x) = \alpha e^{-\alpha x}$  として

$$\frac{\partial}{\partial u} \varphi(u, \delta) = \frac{\lambda + \delta}{c} \varphi(u, \delta) - \frac{\lambda}{c} e^{-\alpha u} \int_0^u \alpha e^{-\alpha x} \varphi(x, \delta) dx - \frac{\lambda}{c} e^{-\alpha u}$$

である。さらに微分して

$$\frac{\partial^2}{\partial u^2} \varphi(u, \delta) = \frac{\lambda + \delta}{c} \frac{\partial}{\partial u} \varphi(u, \delta) + \frac{\alpha \lambda}{c} e^{-\alpha u} \int_0^u \alpha e^{-\alpha x} \varphi(x, \delta) dx - \frac{\alpha \lambda}{c} \varphi(x, \delta) + \frac{\alpha \lambda}{c} e^{-\alpha u}$$

積分を消すと常微分方程式

$$\frac{d^2}{du^2}\varphi(u, \delta) + \left(\alpha - \frac{\lambda + \delta}{c}\right)\frac{d}{du}\varphi(u, \delta) - \frac{\alpha\delta}{c}\varphi(u, \delta) = 0$$

が得られる。この方程式を解くと

$$\varphi(u, \delta) = \kappa_1 e^{\rho_\delta u} + \kappa_2 e^{-R_\delta u}$$

を得る。ただし、 $\rho_\delta > 0$  と  $-R_\delta < 0$  は特性方程式

$$s^2 + \left(\alpha - \frac{\lambda + \delta}{c}\right)s - \frac{\alpha\delta}{c} = 0 \quad (3.9)$$

であり、定数  $\kappa_1$  と  $\kappa_2$  は  $\delta$  に依存している。この2つの定数  $\kappa_1$  と  $\kappa_2$  を求めよう。まず、 $\varphi(u, \delta) \leq \psi(u)$  より

$$\lim_{u \rightarrow \infty} \varphi(u, \delta) = 0$$

が成り立つ。よって  $\kappa_1 = 0$  かつ  $\kappa_2 = \varphi(0, \delta)$  と書くことができる。そこで、 $\varphi(u, \delta)$  を計算する。この式を (3.9) に代入して整理すると

$$0 = \varphi(0, \delta) e^{-R_\delta u} \left(R_\delta + \frac{\lambda + \delta}{c} - \frac{\lambda\alpha}{c} \frac{1}{\alpha - R_\delta}\right) + e^{-\alpha u} \left(\frac{\lambda\alpha}{c} \frac{\varphi(0, \delta)}{\alpha - R_\delta} - \frac{\lambda}{c}\right)$$

が成り立つので

$$\varphi(0, \delta) = 1 - R_\delta/\alpha$$

を得る<sup>2</sup>。

$$\varphi(u, \delta) = (1 - R_\delta/\alpha) e^{-R_\delta u}$$

が成り立つ。これにより破産時刻を表す確率変数  $T_u$  の積率母関数が計算できた。このことから、 $T_u$  のモーメントも計算することができる。積率母関数を  $k$  回微分すると

$$(-1)^k \frac{\partial^k}{\partial \delta^k} \varphi(u, \delta) \Big|_{\delta=0} = E[T_u^k 1_{T_u < \infty}]$$

である。このことから  $\varphi$  を順々に  $\delta$  で微分すれば  $T_u$  のモーメントを得ることができる。例えば、

$$\frac{\partial}{\partial \delta} \varphi(u, \delta) = -\frac{R'_\delta}{\alpha} e^{-R_\delta u}$$

であり、(3.10) から

$$R'_\delta = \frac{1}{2c} \{-1 + ((c\alpha - \delta - \lambda)^2 + 4c\delta\alpha)^{-1/2} (\delta + \lambda + c\alpha)\}$$

<sup>2</sup> 次の等式

$$R_\delta + \frac{\lambda + \delta}{c} - \frac{\lambda\alpha}{c} \frac{1}{\alpha - R_\delta} = \frac{-1}{\alpha - R_\delta} (R_\delta^2 - (\alpha - \frac{\lambda + \delta}{c})R_\delta - \frac{\alpha\delta}{c}) = 0$$

を用いた。

を求めることにより

$$R'_0 = \frac{\lambda}{c(c\alpha - \lambda)}$$

を与える。 $R_0$  は調整計数であり、 $R_0 = \alpha - \lambda/c$  である。よって、

$$E[T_u 1_{T_u < \infty}] = \frac{R'_0}{\alpha} e^{-R_0 u} + (1 - R_0/\alpha) R'_0 u e^{-R_0 u}$$

である。生存確率  $\psi(u) = (1 - R_0/\alpha) e^{-R_0 u}$  で辺々を割ることで、

$$E[T_{u,c}] = \frac{R'_0}{\alpha - R_0} + R'_0 u = \frac{c + \lambda u}{c(c\alpha - \lambda)}$$

がいえ。ただし  $T_{u,c} = T_u |_{T_u < \infty}$  である。高次モーメントについても (面倒ではあるが) 同様に計算することができる。ここで、 $\varphi(u, \delta)$  が関数  $\omega$  を用いて

$$\varphi(u, \delta) = \int_0^{\infty} e^{-\delta t} \omega(u, t) dt$$

とかけるとする。このとき、関数  $\omega$  を定めたい<sup>3</sup>。さて、

$$\zeta_\delta = 1 - R_\delta/\alpha$$

とおくと、

$$\begin{aligned} \varphi(u, \delta) &= \zeta_\delta \exp\{-\alpha(1 - \zeta_\delta)u\} \\ &= \exp\{-\alpha u\} \zeta_\delta \sum_{j=0}^{\infty} \frac{(\alpha \zeta_\delta u)^j}{j!} \\ &= \exp\{-\alpha u\} \sum_{j=0}^{\infty} \frac{(\alpha u)^j}{j!} \left(1 - \frac{R_\delta}{\alpha}\right)^{j+1} \end{aligned}$$

さらに

$$\begin{aligned} 1 - \frac{R_\delta}{\alpha} &= \frac{1}{2c\alpha} (c\alpha + \lambda + \delta - \sqrt{(c\alpha - \lambda - \delta)^2 + 4c\delta\alpha}) \\ &= \frac{1}{2c\alpha} (c\alpha + \lambda + \delta - \sqrt{(c\alpha + \lambda + \delta)^2 - 4c\delta\alpha}) \end{aligned}$$

より

$$\varphi(u, \delta) = \exp\{-\alpha u\} \sum_{j=0}^{\infty} \frac{(\alpha u)^j}{j!} \left( \frac{c\alpha + \lambda + \delta - \sqrt{(c\alpha + \lambda + \delta)^2 - 4c\delta\alpha}}{2c\alpha} \right)^{j+1}$$

ここで、 $s = c\alpha + \lambda + \delta$  および  $a = 2\sqrt{c\alpha\lambda}$  とおくと、

$$\varphi(u, \delta) = \frac{\exp\{-\alpha u\}}{2c\alpha} \sum_{j=0}^{\infty} \left(\frac{u}{2c}\right)^j \frac{(s - \sqrt{s^2 - a^2})^{j+1}}{j!}$$

<sup>3</sup> この式の意味は、 $\varphi(u, \delta) = E[e^{-\delta T_u} 1_{T_u < \infty}]$  に注意すると分かる。

ここで、ラプラス変換  $\beta^*(\delta)$  が

$$\beta^*(\delta) = \int_0^\infty e^{-\delta t} \beta(t) dt = (\delta - \sqrt{\delta^2 - a^2})^v$$

であるならば  $\beta(t)$  は

$$\beta(t) = \frac{va^v}{t} I_v(at)$$

である。ただし、 $I_v(t)$  は変形ベッセル関数

$$I_v(t) = \sum_{n=0}^{\infty} \frac{(t/2)^{2n+v}}{n!(n+v)!}$$

である。この結果を用いると

$$\omega(u, t) = \frac{\exp\{-\alpha u - (\lambda + c\alpha)t\}}{2c\alpha t} \sum_{j=0}^{\infty} \left(\frac{u}{2c}\right)^j \frac{(j+1)(2\sqrt{c\alpha\lambda})^{j+1}}{j!} I_{j+1}(2t\sqrt{c\alpha\lambda}) \quad (3.10)$$

がいえる。

ここで、

$$E[e^{-\delta T_u}] = E[e^{-\delta T_u} | T_u < \infty] Pr[T_u < \infty] + E[e^{-\delta T_u} | T_u = \infty] Pr[T_u = \infty] = E[e^{-\delta T_{u,c}}] \psi(u)$$

である。さらに、

$$e^{-\delta T_u} = e^{-\delta T_u} 1_{T_u < \infty}$$

なので

$$E[e^{-\delta T_u} 1_{T_u < \infty}] = E[e^{-\delta T_u}] = E[e^{-\delta T_{u,c}}] \psi(u)$$

である。このことから

$$E[e^{-\delta T_{u,c}}] = \frac{\varphi(u, \delta)}{\psi(u)}$$

が成り立つが、これは  $T_{u,c}$  のラプラス変換である。 $T_{u,c}$  は密度関数

$$\frac{1}{\psi(u)} \frac{\partial}{\partial t} \psi(u, t)$$

を持つ。 $\omega(u, t) = \frac{\partial}{\partial t} \psi(u, t)$  が成り立つ。

このことから、この式を  $\omega_c(u, t)$  は  $\psi(u)$  で割ることで

$$\omega_c(u, t) = \frac{\exp\{-(\lambda + c\alpha)t - \lambda u/c\}}{2\lambda t} \sum_{j=0}^{\infty} \left(\frac{u}{2c}\right)^j \frac{(j+1)(2\sqrt{c\alpha\lambda})^{j+1}}{j!} I_{j+1}(2t\sqrt{c\alpha\lambda})$$

を得る。

## Reference

- [1] Dickson, D. C. M., *Insurance Risk and Ruin*, Cambridge, 2005.
- [2] Dickson, D. and Waters, H. R. Recursive Calculation of Survival Probabilities, *ASTIN Bulletin*, **21**, 199-221, 1981.
- [3] Drekcic, S. and Wilmot, G. E. On the density and moments of the time to ruin, *ASTIN Bulletin*, **33**, 11-21, 2003.
- [4] Dufresne, F. and Gerber, H. Three methods to calculate the probability of ruin, *ASTIN Bulletin*, **19**, 71-90, 1981.
- [5] Dufresne, F. and Gerber, H. The surpluses immediately before and at ruin, and the amount of the claim causing ruin *Insurance: Mathematics and Economics*, **7**, 193-199, 1988.
- [6] Gerber, H. *An Introduction to Mathematical Risk Theory*, Philadelphia, 1979.
- [7] Gerber, H., Goovaerts, M.J. and Kaas, R. On the probability and severity of ruin, *ASTIN Bulletin*, **17**, 151-163, 1987.
- [8] Gerber, H. and Shiu, E.S. On the time value of ruin, *North American Actuarial Journal*, **2**, 48-78, 1998.
- [9] Panjer, H. H., Recursive evaluation of family of compound distributions, *ASTIN Bulletin*, **12**, 21-26, 1981.
- [10] Panjer, H. H., Direct Calculation of Ruin Probabilities, *Journal of Risk and Insurance*, **53**, 521-529, 1986.

—研究論文—

変額年金保険の統計的リスク管理法  
：局面転換モデルの利用\*

秋山豪太<sup>†</sup>

国友直人<sup>‡</sup>

2005年9月30日投稿

2006年3月6日受理

概要

生命保険業界では変額年金保険が近年における一つの大きな話題として注目されている。変額年金保険では保険契約者は何らかの最低保証を受けられることが一般的であるが、生命保険会社にとってはそのリスク管理が問題となる。本稿では変額年金保険に関する局面転換対数正規 (RSLN) モデルと呼ばれている、一種の隠れマルコフ (Hidden Markov) モデルを利用したリスク管理法の理論と実際の問題を議論する。特にこの間に日本が経験した米国・カナダなどとかなり異なる様相のマクロ経済の動向に依存して、既存のリスク管理法を応用するときに生じる問題点を指摘し、改善可能性を議論する。

キーワード

変額年金保険、最低保証、局面転換 (RS) モデル、隠れマルコフ (Hidden Markov) モデル、VaR、条件付裾期待値 (CTE)、責任準備金

---

\*この論文は秋山豪太・国友直人 (2005) 「変額年金保険の理論と実際」 (Discussion Paper CIRJE J-141, Graduate School of Economics, University of Tokyo) の改訂稿であり、原論文は 2005 年 9 月の日本統計学会・統計関連学会連合大会 (広島) 及び 2005 年 10 月の JARIP 大会 (東京) において報告された。本誌のレフェリー及び川崎能典氏 (統計数理研究所) からの有益なコメントに感謝する。なお、この論文の内容は三井アセット信託銀行の見解を示すものではない。

<sup>†</sup>三井アセット信託銀行

<sup>‡</sup>東京大学大学院経済学研究科

## 1 はじめに

近年になり日本の生命保険業においても変額年金保険を巡る議論がかなり大きな話題になっている。日本の生命保険市場では従来から伝統的に存在しているタイプの生命保険の契約が低迷している中で、変額年金と呼ばれている年金保険の金融契約が急速に拡大を遂げている。変額年金保険は、その特徴として契約保有者が何らかの最低保証を受けられると同時に、支払った保険料の運用益も享受できる仕組みになっているのが一般的という意味では従来の生命保険とは異なるといえる。しかし、他方で生命保険会社は株式市場など金融市場で資金を運用していく中で、もしも運用結果が支払保証を下回った場合には、将来に損失が発生する可能性があるので、変額年金保険の販売は保険会社にとっては大きなリスクの源泉となりうる。特に生命保険のリスクを社会的に制御する仕組みとして機能しているアクチュアリー（保険数理人）にとっては重要な責任準備金の積立額の評価法も重要な課題となる。いずれの評価法を採用するにしても、生命保険会社は何らかの統計学的方法でリスクを計測し、変額年金保険の販売から将来に生じうるリスクを制御することが必要である。

変額年金保険、より一般的にはエクイティリンク型保険と呼ばれる年金保険はこの間にアメリカやカナダなど北米諸国を中心として既にかなりの販売実績がある。日本の生命保険業界においても、近年では様々な形で販売されており、成長が著しい年金保険の新しい動きの原動力となっている。こうした年金保険を巡る実務界の動向を反映して、北米のアクチュアリーなどの関係者を中心に、変額年金保険の統計的リスクの測定に関しての理論が新たな展開を見せている。そこでの議論は従来から伝統的に議論されてい

る保険数理の枠組みとはかなり異なる形で進んでおり、変額年金保険のリスク管理論を巡る新たな問題が注目されている。特にRSLN(regime switching log-normal)モデル、すなわち局面転換対数正規モデルと呼ばれている一種の隠れマルコフ（Hidden Markov）タイプの統計的時系列モデルによる変額年金保険のリスク測定が北米では盛んに行われるようになってきている。ここで局面転換モデルと呼ばれている統計的モデルとは、データがしがたがっている想定される状態変数（state variable）がある時刻に属する局面（レジューム）や分布が時間とともに転換（あるいは変化）していく時系列モデルであるのがその最大の特長である。こうした隠れマルコフ型時系列モデルやボラティリティ（volatility）変動の時系列モデルによる統計的リスク管理法について、日本の生命保険関係者も注目するようになってきている。（例えば田中・松山（2004）を参照<sup>1</sup>。）

本稿ではこうした変額年金保険を巡る議論で重要と考えられる、局面転換（RS）モデルによる変額年金保険のリスク評価法に焦点をあてる。特に統計的モデルとしての局面転換（RS）モデルの統計的性質を明らかにする。さらに、より重要な問題として、カナダや米国で議論されている統計的リスク管理法が日本のデータにおいて実証的に支持されうるか、という実務的に重要な論点について考察する。そして局面転換（RS）モデルを用いた日本における変額年金保険のリスクについての分析結果を報告する。我々の分析によれば、この間に経験した日本のマクロ経済変動を踏まえると、カナダや米国のデータ分析にもとづく変額年金のリスク評価をそのまま日本に適用すると、リスクを巡る幾つかの基本的な問題が生じるので、慎重に様々な統計的

<sup>1</sup>同問題に関する日本アクチュアリー会の説明（<http://www.actuaries.jp/info/hennen.html>）も参考となろう。

方法を利用し、当面の問題を解決すべきであるとの結論を得た。

あらかじめ本稿の構成を説明しておく、2節では隠れマルコフ型の統計的時系列モデルとしての局面転換 (RS) モデルの統計的モデルとしての性質を調べる。特に定常分布、自己共分散関数、同時密度関数などを考察する。さらに、変額年金保険のリスク評価に対して局面転換 (RS) モデルを利用した日本とカナダのデータ解析の結果を示す。次に3節では推定したパラメータを利用したシミュレーションを利用して、責任準備金の分析を行う。その際、VaR(バリュー・アット・リスク)やCTE(条件付裾期待値)などのリスク評価基準についても議論する。最後に4節では、本稿で得られた結論をまとめる。

## 2 局面転換 (RS) モデルの性質と応用

株価の収益率を説明する伝統的な確率過程モデルでは、ファイナンス分野で著名なブラック・ショールズ・アプローチも含めて、株価が連続時間の幾何ブラウン運動にしたがうと仮定することが少なくない。こうした確率過程モデルが示唆するところは、いかなる離散的な時間の間隔に対して、対数差分で求めた株価の収益率は正規分布に従い、重なり合わない間隔における収益率は互いに独立である、ということになる。すなわち、 $S_t$  を時点  $t$  における株価とすれば

$$\log \frac{S_t}{S_r} \sim N(\mu(t-r), \sigma^2(t-r)), \quad t > r \quad (2.1)$$

であり、未知母数としては一定値をとる  $\mu$  とボラティリティ  $\sigma$  を用いる。このように増分が互いに独立に対数正規分布にしたがうという独立(増分)対数正規 ILN (independent lognormal

model) モデルは単純で扱いやすく、比較的短い時間の間隔に関してはかなり良い近似を与えることがファイナンス分野では知られている。しかしながら、長期の問題を考える場合には適切でないことが多いと考えられる。例えばこのモデルでは現実の金融市場において時々観察されている極端な価格変動やボラティリティ母数の変動をとらえることが出来ない<sup>2</sup>。

ボラティリティ変動をとらえる離散時間の統計的時系列モデルについては近年では様々な研究があるが、ここではボラティリティ水準が  $K$  個 ( $K > 1$ ) の離散的な値の間を変動し、離散的な値の間を確率的に転換するタイプの統計的時系列モデルを考えよう。こうしたタイプの統計モデルは一般的には ILN モデルの単純な面を維持しつつ、ILN モデルよりも現実に観察される収益率変動をより適切にとらえることが出来ることが期待できよう。特に  $K = 2$  としたときに得られる離散時系列モデルの解釈としては、市場が時間の経過とともに安定的でボラティリティの低い状態と、不安定でボラティリティの高い状態とを推移するという、直観的ではあるが市場関係者の間に根強く存在する見方と整合的である。ここで高いボラティリティの状態は、例えば突発的に生じる大きな市場をとりまく不確実性に対応していると見ることができよう。

異なる経済変動の局面をある種の確率的メカニズムで推移する、というタイプの統計的時系列モデルはこれまで様々な分野で応用されている。特に局面転換が観察不能な状態変数 (state variable) に基づきマルコフ的に生じるモデルは、統計的時系列解析 (statistical time series analysis) の分野ではマルコフ転換時系列モデル<sup>3</sup>と呼ばれている。計量経済学においてこの

<sup>2</sup>関連する数理ファイナンス分野における諸問題については、例えば国友・高橋 (2003) が参考となろう。

<sup>3</sup>例えばこの種の統計的非線形時系列モデルについては Kitagawa (1987) を参照されたい。



種の局面転換モデルを導入したのは Hamilton (1989) であり、主としてマクロ経済分析における景気循環の複雑な時系列計量経済モデルとして利用されている。さらに、この種のマルコフ転換時系列モデルの中でも局面転換対数正規 (RSLN) モデルによる変額年金のリスク評価を提唱したのが Hardy (2001, 2003) であり、特に北米のアクチュアリー (保険数理人) の間で話題となっている。

ここでは各局面での収益率がしたがう分布として、特に対数正規分布に限ることなく、より一般の局面転換 (RS) モデルの性質を考察しておく。局面転換 (RS) モデルを株価収益率に適用する為に、各期において株価の収益率の過程が  $K$  個の状態の局面 (レジューム) の一つにあると考えよう。ここで  $\rho_t$  で間隔  $[t, t+1)$  における局面 (レジューム) を表し ( $\rho_t = 1, 2, \dots, K$ )、 $S_t$  は  $t$  におけるポートフォリオの総価値、

$$\log \frac{S_{t+1}}{S_t} \Big| \rho_t \sim F(\mu_{\rho_t}, \sigma_{\rho_t}^2) \quad (2.2)$$

とする。ただし  $\mu_{\rho_t}$  と  $\sigma_{\rho_t}^2$  は各局面での位置 (location) とスケール (scale) をそれぞれ表すものとする。ここで、局面転換の推移確率行列  $\mathbf{P} = (p_{ij})$  を

$$p_{ij}(t+1) = \mathbf{P}(\rho_{t+1} = j | \rho_t = i) \quad (2.3)$$

$$i = 1, 2, \dots, K, j = 1, 2, \dots, K$$

として、収益率  $R_{t+1} = \log \frac{S_{t+1}}{S_t}$  の定常分布を考えよう。ここでマルコフ型推移確率が時間に依存せず定常的であり、条件付確率  $p_{ij}(t+1) = p_{ij}$  が方程式

$$\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi} \quad (2.4)$$

を満たす定常分布を  $\boldsymbol{\pi}$  としよう。ただし  $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ ,  $\sum_{i=1}^k \pi_i = 1$  ( $\pi_k \geq 0$ ) であ

る。ここで各局面における収益率は互いに独立にそれぞれ分布関数  $F_i$  ( $i = 1, \dots, k$ ) にしたがうと仮定すれば、収益率の無条件分布は分布  $F_i$  ( $i = 1, \dots, k$ ) の混合分布 (mixture distribution) となる。これを  $R_t \sim \sum_{i=1}^k \pi_i F_i(\theta)$ ,  $F_i(x|\theta) = F_i(\frac{x - \mu_i}{\sigma_i})$  で表そう。このとき、収益率の条件付分布は

$$\begin{aligned} & \mathbf{P}(R_t \leq r_0 | \rho_{t-1} = i_1) \quad (2.5) \\ &= \sum_{i_0=1}^K \mathbf{P}(R_t \leq r_0, \rho_t = i_0 | \rho_{t-1} = i_1) \\ &= \sum_{i_0=1}^K \mathbf{P}(R_t \leq r_0 | \rho_t = i_0, \rho_{t-1} = i_1) \\ & \quad \times \mathbf{P}(\rho_t = i_0 | \rho_{t-1} = i_1) \\ &= \sum_{i_0=1}^K F_{i_0}(\frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}) \pi_{i_1 i_0} \end{aligned}$$

となる。したがって、収益率の分布関数は

$$\begin{aligned} & \mathbf{P}(R_t \leq r_0) \quad (2.6) \\ &= \sum_{i_1=1}^K \mathbf{P}(R_t \leq r_0 | \rho_{t-1} = i_1) \mathbf{P}(\rho_{t-1} = i_1) \\ &= \sum_{i_0=1}^K F_{i_0}(\frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}) \pi_{i_0} \end{aligned}$$

で与えられる。特に分布関数がルベグ測度に関して絶対連続であって、各局面での密度関数が  $f_i(\cdot)$  で与えられるとすると、 $R_t$  の期待値は

$$\begin{aligned} & \mathbf{E}[R_t] \quad (2.7) \\ &= \frac{1}{\sigma_{i_0}} \sum_{i_0=1}^K \int_{-\infty}^{\infty} r_0 f_{i_0}(\frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}) dr_0 \times \pi_{i_0} \\ &= \sum_{i_0=1}^K \pi_{i_0} \{ \mu_{i_0} + \sigma_{i_0} \mathbf{E}[x_{i_0}] \} \end{aligned}$$

と表現できる。ただし、 $x_{i_0}$  は基準化した収益率  $x_{i_0} = \frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}$  であるが、 $\mathbf{E}(x_{i_0}) = 0$  にとれば表現は簡単化される。次に収益率分布の同時分布の性質を調べてみよう。まず  $u \geq 1$  に対

して条件付確率の表現

$$\begin{aligned}
 & \mathbf{P}(R_t \leq r_0, R_{t-u} \leq r_u | \rho_{t-u-1} = i_{u+1}) \quad (2.8) \\
 &= \sum_{i_u=1}^K \mathbf{P}(R_t \leq r_0, R_{t-u} \leq r_u | \rho_{t-u} = i_u, \\
 & \quad \rho_{t-u-1} = i_{u+1}) \times \mathbf{P}(\rho_{t-u} = i_u | \rho_{t-u-1} = i_{u+1}) \\
 &= \sum_{i_u=1}^K \sum_{i_{u-1}=1}^K \cdots \sum_{i_0=1}^K \mathbf{P}\left(\frac{R_t - \mu_{i_0}}{\sigma_{i_0}} \leq \frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}, \right. \\
 & \quad \left. \frac{R_{t-u} - \mu_{i_u}}{\sigma_{i_u}} \leq \frac{r_u - \mu_{i_u}}{\sigma_{i_u}} \mid \right. \\
 & \quad \left. \rho_t = i_0, \dots, \rho_{t-u} = i_u, \rho_{t-u-1} = i_{u+1}\right) \\
 & \quad \times p_{i_{u+1}i_u} p_{i_u i_{u-1}} \cdots p_{i_1 i_0}
 \end{aligned}$$

を利用しよう。これより、二時点の収益率の同時分布関数は

$$\begin{aligned}
 & \mathbf{P}(R_t \leq r_0, R_{t-u} \leq r_u) \quad (2.9) \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K F_{i_0}\left(\frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}\right) F_{i_u}\left(\frac{r_u - \mu_{i_u}}{\sigma_{i_u}}\right) \pi_u p_{i_u i_0}
 \end{aligned}$$

と表現 ( $u > 0$ ) することができる。ここで  $p_{i_u i_0} = \mathbf{P}(\rho_u = i_0 | \rho_{t-u} = i_u)$  であることに注意しておく。同時分布関数より同時密度関数は

$$\begin{aligned}
 & f(R_t = r_0, R_{t-u} = r_u) \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K f_{i_0}\left(\frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}\right) f_{i_u}\left(\frac{r_u - \mu_{i_u}}{\sigma_{i_u}}\right) \\
 & \quad \times \frac{1}{\sigma_{i_u}} \frac{1}{\sigma_{i_0}} \pi_u p_{i_u i_0}
 \end{aligned}$$

で与えられることがわかる。次に収益率分布の自己共分散 (autocovariance) 構造について調べておこう。ここで積率 ( $u > 0$ ) についての評価

$$\begin{aligned}
 & \mathbf{E}[R_t R_{t-u}] \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K \int_{-\infty}^{\infty} r_0 r_u f_{i_0}\left(\frac{r_0 - \mu_{i_0}}{\sigma_{i_0}}\right) f_{i_u}\left(\frac{r_u - \mu_{i_u}}{\sigma_{i_u}}\right) \\
 & \quad \times \frac{1}{\sigma_{i_u}} \frac{1}{\sigma_{i_0}} \pi_u p_{i_u i_0} dr_0 dr_u \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K [\mu_{i_0} + \sigma_{i_0} \mathbf{E}(x_{i_0})][\mu_{i_u} + \sigma_{i_u} \mathbf{E}(x_{i_u})] \\
 & \quad \pi_u p_{i_u i_0}
 \end{aligned}$$

を用いる。自己共分散関数 (autocovariance func-

tion) は

$$\begin{aligned}
 & \mathbf{Cov}(R_t, R_{t-u}) \quad (2.10) \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K [\mu_{i_0} + \sigma_{i_0} \mathbf{E}(x_{i_0})][\mu_{i_u} + \sigma_{i_u} \mathbf{E}(x_{i_u})] \\
 & \quad \times \pi_{i_u} p_{i_u i_0} \\
 & \quad - \sum_{i_u=1}^K [\mu_{i_0} + \sigma_{i_0} \mathbf{E}(x_{i_0})] \pi_{i_0} \\
 & \quad \times \sum_{i_0=1}^K [\mu_{i_u} + \sigma_{i_u} \mathbf{E}(x_{i_u})] \pi_{i_u} \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K [\mu_{i_0} + \sigma_{i_0} \mathbf{E}(x_{i_0})][\mu_{i_u} + \sigma_{i_u} \mathbf{E}(x_{i_u})] \\
 & \quad \times \pi_{i_u} (p_{i_u i_0} - \pi_{i_0})
 \end{aligned}$$

で与えられる。したがって、もし  $p_{i_u i_0} = \mathbf{P}(\rho_u = i_0 | \rho_{t-u} = i_u) = \pi_{i_0}$  となるならば、共分散はゼロ ( $\mathbf{Cov}(R_u, R_{t-u}) = 0$  ( $u \geq 1$ )) であることが分かる。また  $\mathbf{E}[x_{i_0}] = 0$  にとれば収益率の分散は、

$$\mathbf{Var}(R_t) = \sum_{i_0=1}^K [\mu_{i_0}^2 + \sigma_{i_0}^2 \mathbf{E}(x_{i_0}^2)] \pi_{i_0} - \left( \sum_{i_0=1}^K \mu_{i_0} \pi_{i_0} \right)^2$$

となる。全く同様に、 $n \geq 1$  に対する積率は

$$\mathbf{E}[R_t^n] = \sum_{i_0=1}^K \pi_{i_0} \mathbf{E}[(\mu_{i_0} + \sigma_{i_0} x_{i_0})^n] \quad (2.11)$$

と表現される。さらに、 $n_1, n_2 \geq 1$  に対する同時積率を求めると

$$\mathbf{E}[R_t^{n_1} R_t^{n_2}] = \sum_{i_u=1}^K \sum_{i_0=1}^K \mathbf{E}[(\mu_{i_0} + \sigma_{i_0} x_{i_0})^{n_1} (\mu_{i_u} + \sigma_{i_u} x_{i_u})^{n_2}] \pi_u p_{i_u i_0}$$

で与えられる。したがって、特に二乗収益率の  $R_t^2$  の自己相関関数 ( $u > 0$ ) をとりあげると

$$\begin{aligned}
 & \mathbf{Cov}(R_t^2, R_{t-u}^2) \quad (2.12) \\
 &= \sum_{i_u=1}^K \sum_{i_0=1}^K \mathbf{E}[(\mu_{i_0} + \sigma_{i_0} x_{i_0})^2 (\mu_{i_u} + \sigma_{i_u} x_{i_u})^2] \\
 & \quad \times \pi_{i_u} (p_{i_u i_0} - \pi_{i_0})
 \end{aligned}$$

で与えられることがわかる。

なお、収益率の二乗変動はボラティリティの変動に対応していると考えられる。株価に幾何ブラウン運動を想定すると、異なる期間での収益率

は独立で正規分布にしたがい、ボラティリティは一定値をとる。それに対して、局面転換モデルでは自己共分散関数はゼロではなく、収益率のボラティリティも一定とは限らずに変動するので重要な意味を持っている。ここで、さらに収益率がしたがう各局面において分散の存在を仮定しよう。すなわち、ある正定数  $M_1$  と  $M_2$  が存在して  $\mathbf{E}[(\mu_{i_0} + \sigma_{i_0} x_{i_0})^4] \leq M_1$ ,  $\mathbf{E}[(\mu_{i_u} + \sigma_{i_u} x_{i_u})^2] \leq M_2$  とする。このとき、もし状態変数の (有限) マルコフ連鎖で表現された局面転換過程がエルゴード性 (ergodic) を持てば、それぞれある正定数  $\eta (< 1)$  が存在して、

$$\begin{aligned} & |\mathbf{Cov}(R_t^2, R_{t-u}^2)| \quad (2.13) \\ & \leq M_1 \left| \sum_{i_u=1}^K \sum_{i_0=1}^K [P(\rho_t = i_0, \rho_{t-u} = i_u) \right. \\ & \quad \left. - P(\rho_t = i_0)P(\rho_{t-u} = i_u)] \right| \\ & = O(\eta^u) \end{aligned}$$

および

$$\begin{aligned} & |\mathbf{Cov}(R_t, R_{t-u})| \quad (2.14) \\ & \leq M_2 \left| \sum_{i_u=1}^K \sum_{i_0=1}^K [P(\rho_t = i_0, \rho_{t-u} = i_u) \right. \\ & \quad \left. - P(\rho_t = i_0)P(\rho_{t-u} = i_u)] \right| \\ & = O(\eta^u) \end{aligned}$$

を満足することがわかる。以上の考察より得られた局面転換 (RS) モデルの一般的性質は次のようにまとめることができる。

**定理 1** : 局面転換 (RS) モデルにおいて局面転換の推移確率 (transition probability) が時間に依存せず非既約 (irreducible) かつ非循環 (aperiodic) のマルコフ連鎖 (Markov chain) とする。このとき収益率  $R_t$  と  $R_{t-s}$  の同時分布は (2.9) で与えられる。各局面における収益率の分散が有限ならば自己共分散関数は (2.10) で与えられる。またこれらの仮定の下で、収益率と二乗収益率は弱従属 (weakly dependent) の弱定常

過程 (weakly stationary process) となる。

次に各局面において収益率がしたがう分布を仮定して母数を推定する最尤推定法 (maximum likelihood method) を考察しよう。ただし、ここでは議論を単純化するために  $K = 2$  のケースを考えることにする。一般には  $K > 2$  の場合も同様に扱うことができるが、データへのフィットや推定した結果の解釈はそれほど自然なものとはならないことが多いことをここでは注意しておく。特に収益率が各局面でしたがう分布が互いに独立な正規分布にしたがうと仮定すると、推定する母数ベクトルは  $\theta = \{\mu_1, \mu_2, \sigma_1, \sigma_2, p_{12}, p_{21}\}$  の 6 個となる。ここで  $\pi$  は定常性の条件、 $\pi_1 p_{11} + \pi_2 p_{21} = \pi_1$ ,  $\pi_1 p_{12} + \pi_2 p_{22} = \pi_2$  より

$$\pi_1 = \frac{p_{21}}{p_{12} + p_{21}}, \pi_2 = 1 - \pi_1 = \frac{p_{12}}{p_{12} + p_{21}}$$

で与えればよい。

ここで観測期間において  $n$  個のデータが観察されたとすると、観察量  $\mathbf{R} = (R_1, R_2, \dots, R_n)$  の尤度関数は、

$$\begin{aligned} & L_n(\theta) \quad (2.15) \\ & = f(R_1|\theta)f(R_2|R_1, \theta) \cdots f(R_n|R_{n-1}, \dots, R_1, \theta) \end{aligned}$$

と表現される。ここで  $f$  は収益率  $R$  の密度関数、 $\theta$  は未知母数ベクトル、さらに  $t-1$  時点までの情報を所与とした  $t$  番目の観測量がしたがう対数条件付密度関数は

$$\log f(R_t|R_{t-1}, R_{t-2}, \dots, R_1, \theta)$$

である。尤度関数を構成する要素、例えば時刻  $t$  における条件付分布はそれぞれ  $t-1$  までの観測値を所与として逐次的に計算することができる。実際に与えられたデータに対して尤度関数を最大化する解は明示的には可能ではない。そこで最適化アルゴリズムとしてよく知られている EM (Expectation-Maximization) アルゴリズムを利用することで実行することができる。こ

の計算アルゴリズムは、母数ベクトルについてまず初期値を与え、その初期値得られた観察値を元に各時点における局面変化の確率を推定することをを行う。次に推定された確率を利用して時点  $t = 1, 2, \dots, n$  のデータに対する尤度関数を求め、それを4個の母数について最適化することにより母数を推定する。さらに、推定された母数を二回目の初期値として再び計算を実行し、この繰り返し計算を収束するまで行うというものである<sup>4</sup>。

ここで、こうして得られた最尤推定量の漸近的性質については、独立標本に関する議論を直接に適用することはできないことに注意しておく。定理1にまとめておいたように、収益率データは弱定常過程となるので、統計的時系列モデルのパラメトリック推定問題となっている。局面転換モデルは非線形時系列モデルであるので教科書的な線形モデルに関する議論を適用することもまた可能ではない。ところが、幸いなことに隠れマルコフ (Hidden Markov) モデルについてはこれまで統計的時系列解析においては幾つかの研究があり、その理論的結果を適用することで最尤推定量の漸近的結果が得られる。特に観測されない離散的状態変数の取りうる値が有限個であって観測データが連続変量、である隠れマルコフ型の統計的時系列モデルにおける最尤推定量の一致性については Leroux (1982) の結果、漸近正規性については Bickel et. al. (1998) の結果が適用できる。各局面での収益率の分布が一般の場合において最尤推定量のよい性質を導くための正則条件についてはこれら二つの論文に詳しく説明されている。例えば各局面における分布が正規分布であれば、その正則条件はかなり単純化される。ここで母数空間  $\Theta = \{\mu_1, \mu_2, 0 <$

<sup>4</sup>例えば Hamilton (1990) は計算アルゴリズムを詳しく説明しているが、本稿でのデータ分析はその EM 計算アルゴリズムを利用した。

$\sigma_1, 0 < \sigma_2, 0 < p_{12} < 1, 0 < p_{21} < 1\}$  とすると、真の母数値が母数空間の境界上にあると、通常の漸近的結果が必ずしも成立しないことに注意しておこう。統計家の間ではよく知られている重要な例としては、各局面の分布が正規分布にしたがうとき、分散の母数  $\sigma_2 \rightarrow 0$  とすると尤度関数をいくらかでも大きくすることができ、尤度関数が発散する可能性がある。こうした問題を回避するには、母数空間を十分に大きい正実数  $M_i$  ( $i = 3, 4, 5$ ) により

$$\Theta = \{|\mu_1| \leq M_3, |\mu_2| \leq M_3, 1/M_4 \leq \sigma_1, \sigma_2 \leq M_4, 1/M_5 \leq p_{12}, p_{21} \leq 1 - 1/M_5\}$$

と制約することが考えられる。ここで新たな結果ではないが、収束計算や推定結果の解釈など応用上も重要であるので念のためにここでの議論を次のようにまとめておこう。

**定理2:** 局面転換 (RS) モデルにおいて局面転換の推移確率が時間に依存せず、非既約 (irreducible) かつ非循環 (a-periodic) のマルコフ連鎖 (Markov chain) とする。収益率  $R_t$  は各局面において密度関数を持ち、分散が存在し、さらに同時分布が母数について識別可能、母数空間がコンパクトで真の母数とその内点となる正則条件を満足するものとする。このとき最尤推定量は一致性と漸近正規性を持つ。すなわち、 $n \rightarrow \infty$  のとき

$$\hat{\theta} \xrightarrow{p} \theta_0, \quad (2.16)$$

かつ

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}) \quad (2.17)$$

となる。ただし  $\hat{\theta}$  は最尤推定量、 $\theta_0$  は真の母数ベクトル、 $I(\theta_0)$  は尤度関数より求められる (正定符号行列である) フィッシャー情報行列

$$I(\theta_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E} \left[ - \frac{\partial^2 \log L_n(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta_0} \right]$$

とした。

次に各局面での収益率の条件付分布を正規分布と仮定した上で、カナダのデータと日本のデータを用いた推定結果を示しておく。ここでカナダのデータとしては主要な年金保険の分離ファンド契約において共通の指数になっているトロント証券取引所 300 種 (TSE300 : Toronto Securities Exchange300) を用いる。ここで配当込みの TSE300 の月次の対数収益率とその 12カ月の移動標準偏差 (移動ボラティリティ) を図 1 に示しておくが、データの期間は TSE が導入された 1956 年 1 月から 1999 年 12 月までの 527 個である。日本に関しては、株式指標として標準的な TOPIX の月次の対数収益率 (配当無) を用いた。月次対数収益率とその 12カ月の移動標準偏差を図 2 に示しておくが、データの期間は 1956 年 1 月から 1999 年 12 月まで、データ数は 527 個である。

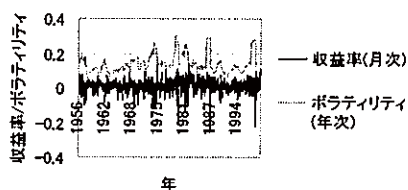


図 1: 月次収益率とボラティリティ(TSE300)

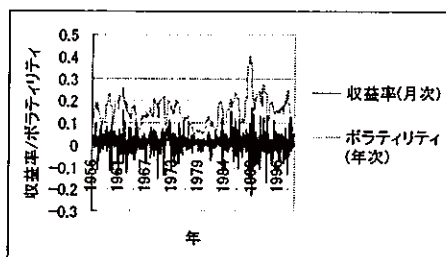


図 2: 月次収益率とボラティリティ(TOPIX)

	$\hat{\mu}$	$\hat{\sigma}$
TSE 300 1956-1999	0.008	0.156
TOPIX 1956-1999	0.007	0.175
TOPIX 1956-1979	0.009	0.160
TOPIX 1980-1989	0.014	0.144
TOPIX 1990-1999	-0.004	0.225

表 1: 月次収益率の平均と年次のボラティリティ

全期間の対数収益率を用いたデータの平均とボラティリティの推定値をまとめると、表 1 のようになる。1956 年から 1999 年までのデータで見ると、TSE300 の年次のボラティリティの推定値は 0.156、TOPIX の年次のボラティリティの推定値は 0.175 となっている。TSE300 の月次の対数収益率の平均は 0.008、TOPIX の月次の対数収益率の平均は 0.007 となっている。なお、TOPIX に関しては期間を三つに分けて、その平均、ボラティリティの推定値を示しておく。これは 4 節で局面転換 (RS) モデルを利用して責任準備金の問題を考えるが、その際には楽観的なシナリオと悲観的なシナリオとして利用し、責任準備金の額がどう変化するかを調べるのである。

図 1 より、TSE のデータでは 1980 年初頭に、TOPIX のデータでは 1990 年初頭に、特にボラティリティの高い状態が数ヶ月続いている。こうした観察事実を独立 (増分) 対数正規 (ILN) モデルの実現値としてとらえることが難しいので、局面転換 (RS) モデルでどのようにとらえられるのか、推定結果を表 2 に示しておく。表 2 より、TSE300 と TOPIX はともに、1956 年から 1999 年のデータに関して予想された結果が得られることがわかる。すなわち、両ケースとも高ボラティリティ局面の平均が低ボラティリティ局面の平均を下回っており、年次のボラティリティは 20% 前半になっている。しかしながら、相違

点も少なくない。TSE300 と TOPIX では、局面間の平均の差が 0.029, 0.012 とかなり異なっている。局面が転換する確率も、特に  $p_{21}$  に関しては、TSE300 が 0.191, TOPIX が 0.045 と異なっている。このことは、TSE300 のデータのほうが TOPIX のデータより高ボラティリティの状態が続かないということを示唆している。ここで局面  $i$  が  $t$  月続く確率は  $p_{ii}^{t-1} p_{ij}$  であるから局面  $i$  が継続する平均期間は  $\frac{1}{p_{ij}}$  になる。したがって、TSE300 のデータの高ボラティリティ局面の状態にある平均期間は、約 5.2 ヶ月であり、TOPIX では約 22 ヶ月と推定される。

ここで TOPIX のデータの期間別の推定結果について見ると、予想していた結果とは異なるところがある。特に、1980-1989 においては、高ボラティリティ局面の方が平均が高く出ている。これはバブル期を含む当時のマクロ経済の動向を反映してのものと解釈できよう。1990-1999 においては、バブル崩壊をうけて高ボラティリティ局面と低ボラティリティ局面の平均がともに負になっている。また、それぞれの局面（レジューム）間の平均の差が異なっていること、 $p_{21}$  に関しては、結果に大きな違いが見られることが読み取れる。すなわち、日本のデータ分析においては局面間の推移確率の推定が不安定となったことが一因であろうと考えられる。

ところで上で求めた推定結果を用いて独立(増分)対数正規(ILN)モデルとの比較を行おう。まず ILN モデル

$$R_t = \mu + \sigma \epsilon_t, \quad \epsilon_t \sim N(0, 1) \quad (2.18)$$

のパラメーターの最尤推定法による推定結果を表 3 に示しておく。収益率が互いに独立に対数正規分布に従うと仮定すると、標本平均と標本分散が最尤推定値になるが、ここでは比較のため、全 5 種類の局面転換対数正規(RSLN)モデルと独立(増分)対数正規(ILN)モデルの尤度関

	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{p}_{12}$
	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{p}_{21}$
TSE 300 1956-1999	0.012 (0.002)	0.039 (0.001)	0.031 (0.008)
	-0.017 (0.014)	0.068 (0.010)	0.191 (0.059)
TOPIX 1956-1999	0.014 (0.002)	0.033 (0.001)	0.055 (0.015)
	0.002 (0.004)	0.061 (0.003)	0.045 (0.012)
TOPIX 1956-1979	0.013 (0.002)	0.046 (0.002)	0.053 (0.013)
	-0.047 (0.005)	0.065 (0.003)	0.723 (0.169)
TOPIX 1980-1989	0.012 (0.003)	0.029 (0.002)	0.033 (0.019)
	0.022 (0.011)	0.059 (0.007)	0.070 (0.042)
TOPIX 1990-1999	-0.001 (0.005)	0.058 (0.004)	0.008 (0.008)
	-0.035 (0.056)	0.105 (0.040)	0.116 (0.131)

表 2: 最尤推定法による RSLN モデルの母数推定値

(注: カッコ (.) は標準偏差の推定値を示している。)

	$\hat{\mu}$	$\hat{\sigma}$
TSE 300 1956-1999	0.008 (0.002)	0.046 (0.001)
TOPIX 1956-1999	0.007 (0.002)	0.051 (0.002)
TOPIX 1956-1979	0.009 (0.003)	0.046 (0.002)
TOPIX 1980-1989	0.014 (0.004)	0.042 (0.003)
TOPIX 1990-1999	-0.004 (0.006)	0.065 (0.004)

表 3: 最尤推定法による ILN モデルの母数推定値

(注: カッコ ( ) は標準偏差の推定値を示している。)

	RSLN	ILN
TSE 300 1956-1999	917.38	888.19
TOPIX 1956-1999	852.48	825.14
TOPIX 1956-1979	477.14	472.16
TOPIX 1980-1989	223.07	206.66
TOPIX 1990-1999	161.40	145.69

表 4: 2つのモデルの尤度関数値

数の値を表 4 に示しておく。

ここでエクイティリンク型保険では長期にわたる収益率分布の裾をどの様にとらえるかが問題であることに注意しよう。そこで局面転換モデルの局面に滞在回数分布関数を求め、そこから、局面転換モデルによる長期における収益率の分布を求め、通常の独立(増分)対数正規モデルによる結果と比較することを考えよう。

確率変数  $M$  を局面 1 に滞在した総回数とすると、 $M \in \{0, 1, 2, \dots, n\}$  の可能性がある。また確率  $P(M = m) = p(m)$  で表現し、さらに  $M_t$  を期間  $[t, n]$  において局面 1 に滞在した総回数としよう。そして、 $P(M_t = m | \rho_{t-1})$  ( $m = 0, 1, \dots, n-t; t = 1, \dots, n-1$ ) を考察しよ

う。まず明らかに、 $m > n-t$ ,  $m < 0$  に対しては  $P(M_t = m | \rho_{t-1}) = 0$  である。また、例えば  $P(M_{n-1} = 0 | \rho_{t-1} = 1)$  は前の時点  $t-1$  では局面 1 に滞在し、その後は局面 1 に滞在することがない確率を表している。ここで  $t \in [n-2, n-1]$  に対して  $P(M_{n-1} = 1 | \rho_{t-1} = j) = p_{j1}$  ( $j = 1, 2$ ),  $P(M_{n-1} = 0 | \rho_{t-1} = j) = p_{j2}$  であり、初期条件として  $P(\rho_t = j) = \pi_j$  ( $j = 1, 2$ ) を用いることが考えられよう。したがって、定常性の仮定の下では関係

$$\begin{aligned} & P(M_t = m | \rho_{t-1} = j) \\ &= p_{j1} P(M_{t+1} = m-1 | \rho_t = 1) \\ & \quad + p_{j2} P(M_{t+1} = m | \rho_t = 2) \quad (j = 1, 2) \end{aligned}$$

を使えば、逐次的に確率  $P(M = M_0)$  を求めることができる。

次に  $M$  の分布関数を利用して  $n$  時点における長期の指数収益率の分布を表現することを考えよう。 $S_n$  を  $n$  時点での指数水準、初期条件を  $S_0 = 1$  と仮定しよう。確率変数  $\{M = m\}$  及び状態変数  $\{\rho_1 = i_1, \dots, \rho_n = i_n\}$  を条件とすると、指数水準の対数変換値  $\sum_{t=1}^n R_t$  は  $m$  個の互いに独立な状態 1 の確率変数と  $n-m$  個の互いに独立な確率変数の和として表現される。状態変数を取りうる値  $\{M = m\}$  で条件づけた条件付確率を事象  $\{M = m\}$  の範囲でとれば 1 であるので、各局面での分布を正規分布に特定化すれば

$$S_n | M \sim \text{lognormal}(\mu_*(M), \sigma_*^2(M)),$$

$$\mu_*(M) = M\mu_1 + (n-M)\mu_2,$$

$$\sigma_*^2(M) = M\sigma_1^2 + (n-M)\sigma_2^2$$

となる。(ここで対数正規分布  $\text{lognormal}(\mu, \sigma^2)$  とは対数をとると正規分布  $N(\mu, \sigma^2)$  にしたがう分布と定義する。) すなわち各局面での分布を正規分布に特定化すれば、 $M$  の分布関数  $p(m)$

を用いて、

$$\begin{aligned}
 F_{S_n}(x) &= P(S_n \leq x) & (2.19) \\
 &= \sum_{m=0}^n P(S_n \leq x | M = m) p(m) \\
 &= \sum_{m=0}^n \Phi\left(\frac{\log x - \mu_*(m)}{\sigma_*(m)}\right) p(m)
 \end{aligned}$$

となる。ここで  $\Phi$  は標準正規分布の累積分布関数を表している。同様に、 $S_n$  の確率密度関数は、

$$f_{S_n}(x) = \sum_{m=0}^n \phi\left(\frac{\log x - \mu_*(m)}{\sigma_*(m)}\right) \frac{1}{x} \frac{1}{\sigma_*(m)} p(m) \quad (2.20)$$

と書けるが、 $\phi$  は標準正規分布の確率密度関数である。これらの累積密度関数、確率密度関数を使って局面転換対数正規 (RSLN) モデルと独立 (増分) 対数正規 (ILM) モデルにしたがう収益率の分布関数や確率密度関数などを推定することも出来る。長期間の収益率分布については独立増分対数正規 (ILN) モデルに比べて、局面転換対数正規 (RSLN) モデルの左の裾が厚いことが分かる。こうした裾の厚さは、エクイティリンク型保険において長期の契約におけるリスクをモデリングする際に有用であろう。しかしながら、このコメントは TSE300 の全データと TOPIX の全データ、並びに、TOPIX の 1956-1979 のデータに関して言えるものである。TOPIX の 1980-1989 のデータと TOPIX の 1990-1999 のデータに関して求めた結果は、少なくとも高ボラティリティ期間でもネガティブの平均を持つという結果になっている。

ここでは表 2 で推定したパラメーターを使い、さらに初期値  $S_0 = 100$  とした時の収益率の確率密度関数を図 3 から図 7 に与えておく。長期間の収益率分布については独立対数正規 (ILN) モデルに比べて、局面転換対数正規 (RSLN) モデルの左の裾が厚いことがみてとれる。こうし

た裾の厚さが、エクイティリンク型保険において長期の契約におけるリスクをモデリングする際に有用になってくる。

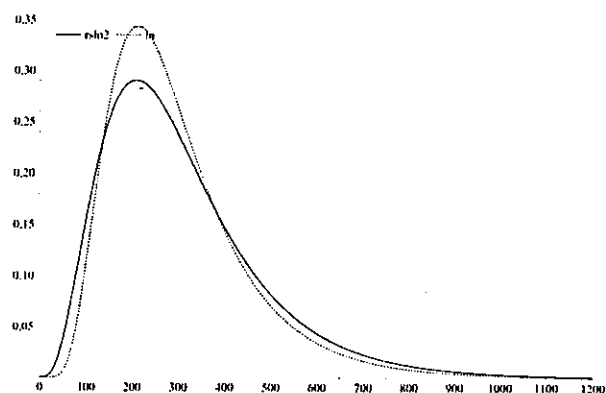


図 3: TSE300 の確率密度関数

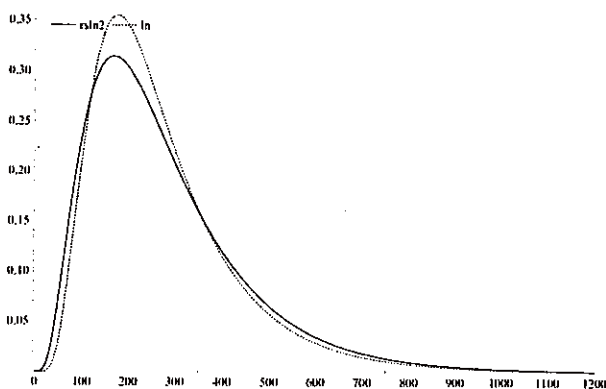


図 4: TOPIX1956-1999 の確率密度関数

次に滞在総回数の分布関数を用いれば、任意の時点における株価の積率を求めることができることに注意する。条件付対数正規分布の性質



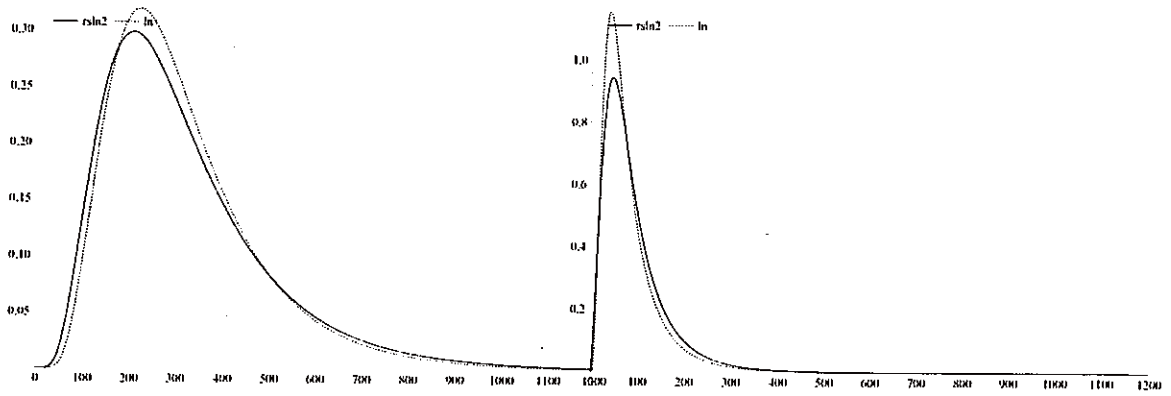


図 5: TOPIX1956-1979 の確率密度関数

図 7: TOPIX1990-1999 の確率密度関数

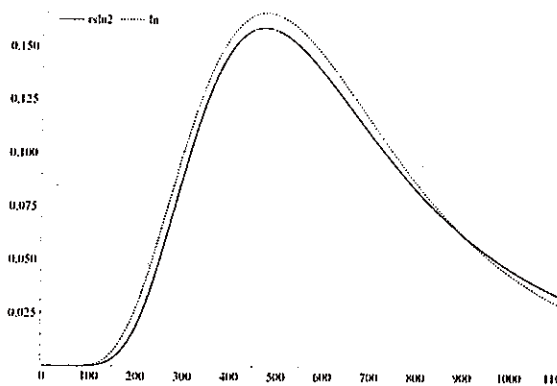


図 6: TOPIX1980-1989 の確率密度関数

おく。また比較の為に収益率データから計算した結果を表 6 にまとめておく。なお TOPIX は全期のデータを、TOPIX1,2,3 はそれぞれ 1956 年～1979 年、1980 年～1989 年、1990 年～1999 年の期間のデータをそれぞれ意味している。

### 3. 変額年金保険のリスク測定

を利用すると

$$\begin{aligned}
 & \mathbf{E}[(S_n)^k] \\
 = & \mathbf{E}\left[\exp\left(k(M\mu_1 + (n-M)\mu_2) + \frac{k^2}{2}(M\sigma_1^2 + (n-M)\sigma_2^2)\right)\right] \\
 = & \mathbf{E}\left[\exp\left(M(k(\mu_1 - \mu_2) + \frac{k^2}{2}(\sigma_1^2 - \sigma_2^2))\right) \times \exp\left(kn\mu_2 + \frac{k^2}{2}n\sigma_2^2\right)\right] \\
 = & \exp\left(kn\mu_2 + \frac{k^2}{2}n\sigma_2^2\right) \\
 & \times \sum_{m=0}^n \exp\left(m\left(k(\mu_1 - \mu_2) + \frac{k^2}{2}(\sigma_1^2 - \sigma_2^2)\right)\right) p(m)
 \end{aligned} \tag{2.21}$$

となることがわかる。この (2.21) 式を用いて、局面転換 (RSLN) モデルに関して上記の 5 ケースについて表 5 に、平均、2 次、3 次、4 次のモーメント、歪度、尖度を求めた結果を与えて

局面転換 (RS) モデルを用いたリスク評価法を考えよう。特に具体的な例として局面転換 (RS) モデルを用いたエクイティリンク型保険のリスク評価の例として、分離ファンドのリスク評価について考察する。分離ファンド契約の最も典型的な形式として、10 年間保険料が投資信託に投資されるという状況考え、運用などの諸費用は月々差し引かれるとする。保険の満期には契約保有者は、受け取りが支払った保険料を下回らないという保証がついた形で投資の成果を受け取る契約を想定しよう。ここでは、仮想的<sup>5</sup>に保険会社の満期における責任準備金 (以下では

<sup>5</sup>なお責任準備金については実際には、オプション・プレミアムである最低保証料収入など中途の収入と支出、死亡率の動向、金利の変動と解約率や引き出し率の動向等、日本アクチュアリー会別冊 (2004) が詳しく議論しているように、具体的な商品設計に応じた様々な要因を含む分析が必要となる。こうした様々な実際の論点について指摘されたレフェリーに感謝する。

簡略準備金とよぶ) は、 $G$  を満期において保証している額、 $F$  を満期におけるファンドとする、 $\max[G - F, 0]$  で与えられると考える。したがって、この契約はヨーロッパ型プット・オプション (European put-options) の一種と解釈することが出来る。こうした金融契約を販売する時、その潜在的なリスクを分析するためには、特に左側の裾が厚いことを許す長期の株式収益率モデルが必要となる。すなわち、保険会社はプット・オプション契約の販売に伴い、株式運用から生じうる将来における不利な経済状況に陥る可能性をも考慮する必要がある。

1つのリスクを測定する基準としては、簡略準備金の分布のパーセント点を見ることが考えられる。これがバリュエーション・アット・リスク (VaR: value-at-risk) アプローチである。ここでは、株価の収益率を説明するモデルとして独立増分対数正規モデルを用いたときと局面転換 (RS) モデルを用いたときのパーセント点を比較してみよう。

ここで単純化の為に保険契約の中途解約については考えないことにする。また  $S_1 = 100$ ,  $G = 100$  を最低保証額、運用コストとして、複利で月  $h$  パーセントがファンドから差し引かれ、 $S_n$  を  $n$  時点における当該資産の価格とする。ここで議論の単純化の為に、 $h$  が小さく満期におけるファンドの価値を  $S_n e^{-nh}$  で近似して、満期時における簡略準備金を

$$X = \max(G - S_n e^{-nh}, 0) \quad (3.1)$$

と見なそう。ここで  $\zeta = P(S_n e^{-nh} > G)$  とすると、もし  $\alpha \leq \zeta$  であれば、簡略準備金の分布の  $100\alpha$  パーセント点は  $V_\alpha = 0$  である。さらに  $\alpha > \zeta$  となる場合には  $V_\alpha$  は

$$F_{S_n}((G - V_\alpha)e^{nh}) = (1 - \alpha) \quad (3.2)$$

より

$$V_\alpha = G - e^{-nh} F_{S_n}^{(-1)}(1 - \alpha) \quad (3.3)$$

で与えられる。ここで  $F_{S_n}$  は当該資産の  $n$  時点における分布関数、 $F_{S_n}^{(-1)}(t) = \inf\{x | F(x) \geq t\}$  で定める。特に独立増分対数正規 (ILN) モデルの場合には標準正規分布の累積分布関数は逆関数が存在するので、 $z_\alpha = \Phi^{-1}(\alpha)$  と置けば、

$$V_\alpha = G - S_0 \exp[-z_\alpha \sqrt{n}\sigma + n\mu - nh] \quad (3.4)$$

である。

ここで VaR (バリュエーション・アット・リスク) はパーセント点であり、その点よりも左側の分布の形状は考慮しない基準になっていることに注意しよう。それに対して、近年、北米のアクチュアリー間でも有力となっているリスク尺度 (risk measure) として、条件付裾期待値 (CTE, conditional tail expectation) がある。この統計量は、簡略準備金の値が分布の  $(1 - \alpha)$  の裾に落ちることを所与としたときの簡略準備金の期待値に対応する。CTE はパーセント点を越えた分布の裾をすべて利用しているからバリュエーション・アット・リスクよりもさらに悪化する状況をもより現実的に考慮していることになる。したがって、VaR に比べるとより保守的な基準を意味している。

ここで、連続かつ単調増加の簡略準備金の分布に対して  $0 \leq \alpha < 1$  の範囲にある任意の母数  $\alpha$  に対する CTE は

$$CTE(\alpha) = E[X | X > V_\alpha] \quad (3.5)$$

で与えられる。ここで  $V_\alpha$  はバリュエーション・アット・リスク値である。こうした CTE の定義によれば、例えば  $\alpha < \zeta$  に対しては  $V_\alpha = 0$  となる。このようなケースが実用的に望ましくないとす

れば、例えば  $\beta' = \max \beta : V_\alpha = V_\beta$  という  $\beta$  に基づき

$$CTE(\alpha) = \frac{(1 - \beta')E[X|X > V_\alpha] + (\beta' - \alpha)V_\alpha}{1 - \alpha} \quad (3.6)$$

と定義することを、例えば Hardy (2001) は提案しているが、この CTE の定義は一種の実用的な簡便法と考えられよう。

ここで  $\alpha \geq \zeta$  を仮定すると

$$\begin{aligned} CTE(\alpha) &= E[X|X > V_\alpha] \\ &= E[G - S_n e^{-nh} | S_n < (G - V_\alpha)e^{nh}] \\ &= \frac{1}{1 - \alpha} \int_0^{(G - V_\alpha)e^{nh}} (G - ye^{-nh}) f_{S_n}(y) dy \\ &= \frac{1}{1 - \alpha} \left[ G \times F_{S_n}((G - V_\alpha)e^{nh}) - \int_0^{(G - V_\alpha)e^{nh}} ye^{-nh} f_{S_n}(y) dy \right] \\ &= G - \frac{e^{-nh}}{1 - \alpha} \int_0^{(G - V_\alpha)e^{nh}} y f_{S_n}(y) dy \end{aligned} \quad (3.7)$$

と表現されことに注意しよう。ここで初期値  $S_0 = 1$  かつ  $S_n \sim \text{lognormal}(n\mu, n\sigma^2)$  であるならば、 $\alpha \geq \zeta$  に対して

$$\begin{aligned} CTE(\alpha) &= G - \frac{\exp(n\mu - nh + n\sigma^2/2)}{1 - \alpha} \\ &\quad \times \Phi\left(\frac{\log(G - V_\alpha) - n\mu + nh - n\sigma^2}{\sqrt{n}\sigma}\right) \end{aligned} \quad (3.8)$$

となる。なお、この表現は初期値  $S_0 = 1$  を仮定しているの、より一般の初期値  $S_0$  に対しては、同様の議論により

$$\begin{aligned} CTE(\alpha) &= G - \frac{\exp(n\mu - nh + \log(S_0) + n\sigma^2/2)}{1 - \alpha} \\ &\quad \times \Phi\left(\frac{\log(G - V_\alpha) - \log(S_0) - n\mu + nh - n\sigma^2}{\sqrt{n}\sigma}\right) \end{aligned} \quad (3.9)$$

で与えられる。

また、 $S_n$  が局面転換対数正規 (RSLN) モデルにしたがうときには、

$$\begin{aligned} S_n | M &\sim \text{lognormal}(\mu_*(M), \sigma_*^2(M)), \\ \mu_*(M) &= M\mu_1 + (n - M)\mu_2, \end{aligned}$$

$\sigma_*^2(M) = M\sigma_1^2 + (n - M)\sigma_2^2$  となることに注意すると、 $\alpha \geq \zeta$  ならば

$$\begin{aligned} CTE(\alpha) &= G - \frac{e^{-nh}}{1 - \alpha} \sum_{m=0}^n p(m) \\ &\quad \times (\exp(\mu_*(m) + \log(S_0) + \sigma_*^2(m)/2)) \\ &\quad \times \Phi\left(\frac{\log(G - V_\alpha) - \log(S_0) - \mu_*(m) + nh - \sigma_*^2(m)}{\sigma_*(m)}\right) \end{aligned}$$

となる。

特に  $\alpha < \zeta$  の場合には、 $\beta' = \zeta$  かつ  $V_\alpha = V_\zeta = 0$  であるから  $CTE(\zeta) = E[X|X > 0]$  となるので (3.6) の定義を採用すれば、

$$CTE(\alpha) = \frac{1 - \zeta}{1 - \alpha} CTE(\zeta)$$

で与えられる。

次に既に推定された 5 種類の母数を用いて局面転換対数正規 (RSLN) モデルによりシミュレーションを行い、バリュエーション・アット・リスクと CTE のリスク評価を表 7 にまとめておく。比較のために独立対数正規 (ILN) モデルについても表 8 にまとめておく。これらの表より、1980 年-89 年のような 2 つの局面の平均が正であるような場合においては、ドリフト (あるいはトレンド) 部分の寄与が大きく、当然ながら事後的にはここで定義したような意味での簡略準備金の額がほとんど不要であったことが分かる。他方、1990 年-1999 年のような 2 つの局面での平均が負であるような場合においては、プット・オプションの販売から生じる簡略準備金の額が非常に大きくなることもまた分かる。

次に時間的経過に伴い、将来における簡略準備金の評価をどうすべきかについて考えてみよう。ある  $j$  期 ( $0 \leq j < 120$ ) の簡略準備金の額は確率密度関数

$$\begin{aligned} f_{S_{120-j}}(x) &= \sum_{m=0}^{120-j} \phi\left(\frac{\log x - \mu^*(m) - \log S_j}{\sigma^*(m)}\right) \frac{1}{x} \frac{1}{\sigma^*(m)} p(m) \end{aligned} \quad (3.10)$$

にしたがうと考えられる。ただし、 $S_{120-j}$  は残り 120 -  $j$  月の当該資産を表すが、 $j$  期では  $S_j$

が確定することになる。ここでは、1年後から8年後まで1年ごとに簡略準備金はどのように調整すべきであるかをシミュレーションにより考察してみる。系列  $S_{12}, S_{24}, \dots, S_{96}$  までを100回乱数発生させ、簡略準備金の評価を100回行いその平均をとってみた。それらのシミュレーションを利用し、表9はTSEの計算結果、表10にはTOPIXの計算結果をそれぞれまとめておいた。ここでyearは残存年数であり、yearが減ると簡略準備金の額も減る傾向にある事が観察される。収益率が独立(増分)対数正規(ILN)モデルにしたがう場合には対数正規分布の性質を利用して簡単に様々な統計量が計算できるが、局面転換モデルの場合には残存期間がある程度あると、その効果は複雑となるのでシミュレーションが必要となろう。

## 4 結語

本稿では生命保険業において近年に注目されている変額年金保険とそれに関わる新しいリスク管理について、新たに分析が必要となる統計的側面を中心に検討した。ここで得られた主要な結論は次のようにまとめることができる。

第一に変額年金保険のリスク管理法には伝統的に保険数理(アクチュアリアル法)が依拠してきた古典的な大数の法則や中心極限定理が十分な役割を果たし得ないという意味で、生命保険業においてかなり新たな問題を提起していることは明らかである。

第二に北米の保険関係者を中心にこの間議論されている局面転換(RS)モデルによるリスク管理法は、ファイナンス分野で開発されたブラック・ショウルズ・モデルに対応する離散時系列モデルである互いに独立な増分が対数正規分布にしたがうという、独立(増分)対数正規(ILN)モ

デルを適用する時に生じる様々な実証的問題点の幾つかをとりあえず解決してくれる、という側面があるという意味で興味深い方法である。

局面転換モデルにおいては必ずしも各局面での分布を正規分布に限る必要はないが、各局面での収益率分布を正規分布とすると、収益率分布の周辺分布は混合正規分布となり、ボラティリティ変動を許容し、局面転換に関する一般的な条件の下で、弱相関構造を有する直観的に分かりやすい統計的時系列モデルと見なすことができる。

しかしながら、より大きな第三の論点として変額年金保険に関するリスク評価における局面転換モデルの利用を考えると幾つかの問題点も同時に浮上してきた。変額年金保険に関連するリスク全体に目を向けると、伝統的な生命保険で問題としている死亡率リスクというよりも、大きな割合で将来の株価水準といったマクロ経済変数の動向により直接的に依存している。ところが、例えばこの間の北米のマクロ経済動向と日本のマクロ経済動向には共通点はあるものの、異なる大きな変動も少なくない。変額年金の場合には、場合によっては数十年の単位での長期多期間におよぶ保険リスク評価をある程度まで行う必要がある。したがって、当然ではあるが長期の収益率に対し適当に期間を設定して局面転換(RS)モデルを適用してリスク評価を行うことから問題が発生する可能性があることが確認された。

ここでは局面転換(RS)モデルは統計的時系列分析(statistical time series analysis)においては隠れマルコフ(Hidden Markov)モデルと呼ばれているタイプの時系列モデルであることを指摘しておく。この種の離散時系列モデルは時間的従属性を持つ同時確率分布が非正規性や非線型性をも表現しうる便利な統計学的手段を提供

している。ただし、非正規性と非線型性を満足する離散時系列モデルは既にこれまで数多くの研究がなされている。例えば北米のアクチュアリー<sup>6</sup>の議論では確率的ボラティリティ(stochastic volatility)変動の時系列モデルや(意味はかなり異なるものの)安定分布(stable distribution)モデル<sup>6</sup>なども言及されている。あるいは、経済的考察から開発された非線形時系列モデルとして Kunitomo and Sato(1999) は同時転換自己回帰モデル(SSAR)を提案している。また、よりノン・パラメトリック統計学的方法に基づく多期間リスク管理法については例えば国友・一場(2004)が検討している。どのような統計の時系列モデルを用いたリスク管理が適切なのか、実務面での応用も含めて今後もさらに検討していく必要があるだろう。

最後に本稿では変額年金のリスク管理問題の中でも特に統計学的にも議論が必要となった幾つかの問題に絞った限定的な分析を行ったことに注意しておく。当然ではあるが、多くの残された重要な検討課題がある。例えば変額年金の運用資産間の依存性を無視することができないが、近年では様々な新しい統計学的分析方法<sup>7</sup>が提案されている。また、現実的方法として採用されることがあるボラティリティ母数を適当に調整するという、リスク調整によるリスク管理法(アクチュアリー会別冊(2004)を参照)の妥当性についてもさらなる検討が必要であろう。また、変額年金の実際的なリスク管理では死亡率の動向、金利の変動と解約率や引き出し率の動向など契約者の選択動向の分析も重要である。こうした変額年金問題を巡る様々な側面につい

<sup>6</sup>安定分布やより一般的な無限分解可能分布(infinitely divisible distributions)について詳しいことは、佐藤(1990)が参考となろう。

<sup>7</sup>例えば資産運用でよく利用されている相関係数は正規分布の仮定をはずすと従属性に関するもっともらしい指標とは限らない。確率変数間の従属性をより一般に表現するコピュラ関数については Nelson(1999)を参照されたい。

ても、今後、さらに検討することが望まれよう。

## 引用文献

- [1] Bickel, P.J, Ritov, Y, and Ryden, T. (1998), "Asymptotic Normality of the Maximum Likelihood for General Hidden Markov Models," *Annals of Statistics*, 26-4, 1614-1635.
- [2] Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Non-stationary Time Series," *Econometrica*, 57, 357-84.
- [3] Hamilton, J.D. (1990), "Analysis of Time Series subject to Changes in Regime," *Journal of Econometrics*, 45, 39-70.
- [4] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.
- [5] Hardy, M.R. (2001), "A Regime-Switching Model of Long-Term Stock Returns," *North American Actuarial Journal*, 5-2, 42-53.
- [6] Hardy, M.R. (2003), *Investment Guarantees*, John Wiley & Sons Inc.
- [7] Kitagawa, G. (1987), "Non-Gaussian State Space Modelling of Non-stationary Time Series," *Journal of the American Statistical Association*, 82 (with Discussions), 1032-1063.
- [8] 国友直人・高橋明彦(2003), 「数理ファイナンスの基礎: マリアバン解析と漸近展開の応用」, 東洋経済新報社。
- [9] 国友直人・一場知之(2004), 「多期間リスク管理法と変額年金保険」, Discussion Paper CIRJE-J-113, Graduate School of Economics, University of Tokyo (<http://www.e.u-tokyo.ac.jp/cirje/>よりダウンロード可能), 日本統計学会(和文)誌(近刊)。
- [10] Kunitomo, N. and Sato, S. (1999), "Stationary and Non-stationary Simultaneous Switching Autoregressive Models with an Application

to Financial Time Series," *Japanese Economic Review*, 50-2, 161-190, Blackwell.

[11] Leroux, B. (1982), "Maximum-likelihood Estimation for Hidden Markov Models," *Stochastic Processes and their Applications*, 40, 127-143.

[12] 田中周二・松山直樹 (2004), 「統計学とアクチュアリーの現代的課題」, 日本統計学会 (和文) 誌, Vol.34-1, 41-55.

[13] 日本アクチュアリー会別冊 (2004), 「変額年金等の最低保証リスクに係る責任準備金の積立等について」.

[14] Nelson, R. (1999), *An Introduction to Copulas*, Springer.

[15] 佐藤健一 (1990), 「加法過程」, 紀伊國屋書店.

	TSE300	TOPIX	TOPIX1	TOPIX2	TOPIX3
平均	3.1056	2.6331	3.2296	7.1764	0.90306
2次モーメント	12.853	8.3228	14.181	69.183	1.3333
3次モーメント	68.875	30.420	84.092	955.40	3.0986
4次モーメント	468.95	128.66	669.59	20376.	11.160
歪度	1.5727	0.72517	1.9362	2.7587	2.5754
尖度	7.5859	5.3237	10.262	20.377	16.770

表 5: RSLN モデルの積率・歪度・尖度

	TSE300	TOPIX	TOPIX1	TOPIX2	TOPIX3
平均	2.9993	2.7612	3.1991	6.4926	0.80748
2次モーメント	11.479	10.363	13.228	51.858	1.0826
3次モーメント	56.059	52.869	70.704	509.54	2.4099
4次モーメント	349.34	366.60	488.48	6159.1	8.9074
歪度	1.7212	2.0134	1.7810	1.5498	2.9745
尖度	8.6916	10.979	9.1229	7.5538	22.025

表 6: 収益率より推定された積率・歪度・尖度

	TSE300	TOPIX	TOPIX1	TOPIX2	TOPIX3
$\zeta$	0.8724	0.8302	0.9135	0.99968	0.1805
$V_{0.90}$	8.8053	19.473	0	0	81.035
$V_{0.95}$	28.215	37.030	15.028	0	86.443
$V_{0.975}$	42.216	49.254	29.252	0	90.073
$CTE(0.90)$	31.558	39.669	18.784	0.0366	88.524
$CTE(0.95)$	44.837	51.547	32.174	0.0731	91.258
$CTE(0.975)$	55.008	60.114	42.562	0.1463	93.322

表 7: RSLN モデルによるリスク評価

	TSE300	TOPIX	TOPIX1	TOPIX2	TOPIX3
$\zeta$	0.9046	0.8327	0.9390	0.99864	0.1364
$V_{0.90}$	0	16.213	0	0	81.702
$V_{0.90}$	15.621	31.622	4.8049	0	85.850
$V_{0.90}$	27.992	42.661	18.806	0	88.740
$CTE(0.90)$	18.835	34.317	11.256	0.1650	86.620
$CTE(0.95)$	30.811	45.023	21.754	0.3301	89.290
$CTE(0.975)$	39.869	53.026	32.178	0.6601	91.397

表 8: ILN モデルによるリスク評価

表 9: TSE の動学的リスク評価

year	VaR90	VaR95	VaR97.5	CTE90	CTE95	CTE97.5
10	8.8053	28.215	42.216	31.558	44.837	55.008
9	8.0621	22.197	36.101	29.422	39.673	49.742
8	8.3658	18.841	31.260	28.835	37.775	45.937
7	7.7355	16.155	26.694	27.763	35.888	41.939
6	8.1875	14.937	23.709	27.951	35.890	40.669
5	8.0847	13.944	20.849	27.059	33.814	39.786
4	7.7826	12.688	18.839	26.978	32.046	35.930
3	7.6623	12.149	16.871	23.647	25.464	31.862
2	6.9010	9.6155	12.873	20.322	22.788	27.921

表 10: TOPIX の動学的リスク評価

year	VaR90	VaR95	VaR97.5	CTE90	CTE95	CTE97.5
10	19.473	37.030	49.254	39.669	51.547	60.114
9	14.487	30.632	43.208	35.407	46.047	54.988
8	13.289	25.519	37.810	33.358	42.507	50.434
7	11.378	21.602	32.523	32.586	39.269	46.073
6	10.712	19.244	28.370	32.734	38.389	43.498
5	9.0013	15.990	23.811	29.286	32.484	39.869
4	10.066	15.339	21.231	27.600	32.949	36.413
3	9.2178	13.307	17.371	25.421	29.886	34.586
2	8.6963	11.173	13.855	19.466	25.232	26.820



**Risk Management Methods of Equity-Linked Insurance  
and Practical Problems**

**Gota Akiyama**

**Mitsui Asset Trust and Banking Company, Limited  
3-23-1, Shiba, Minato-ku, Tokyo 105-8574**

**Naoto Kunitomo**

**Graduate School of Economics, University of Tokyo  
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033**

**Abstract**

Recently the various types of the equity-linked insurance have been introduced and actively traded in Japanese financial markets. We investigate the basic problems of the actuarial risk management methods for those products based on the Markovian regime-switching time series models, which was originally proposed by Hamilton (1989). We argue that they should be carefully used in Japan mainly because the macro-economic performance of Japan in the past decades have been quite different from the macro-economies of Canada and the U.S..

# Swiss Solvency Test について\*

川崎 能典†  
統計数理研究所

## 1 はじめに

本稿の目的は、スイス連邦民間保険局 (Swiss Federal Office of Private Insurance) による “White Paper of the Swiss Solvency Test” (2004) に基づき、SST を概観することである。FOPI はスイスにおける保険の規制当局、監督機関である。以下本稿ではこの文献を「SST 白書」と記す。

保険会社の経営の健全性を表す指標として従来から用いられているのがソルベンシー・マージン比率である。これは、大規模災害など不測の事態で保険金支払額が急増した場合でも、保険会社が契約通りに保険金を払う原資を持っているかどうかを示す指標であり、保険会社の支払い余力に関する目安である。一般にはこの値が 200% を超えているか否かが判断の分かれ目になる。

$$\frac{\text{ソルベンシー・マージン総額}}{\text{リスクの合計額}/2} \times 100(\%) \geq 200(\%)$$

しかしながら近年本邦でも、200% を上回っていた保険会社が相次いで破綻した経緯から、基準値の一層厳しい見直しが迫られている。(なお、SST 白書によれば、スイスではまだ保険会社の破綻事例はない。)

ソルベンシー・マージンの主な構成要素は、基金（資本金）、価格変動準備金、危険準備金、貸倒引当金、株式含み益、劣後ローンなどである。リスクの合計額は、

- 保険リスク：大災害の発生などにより、保険金支払いが急増するリスク
- 予定利率リスク：運用環境の悪化により、資産運用利回りが予定利率を下回るリスク
- 資産運用リスク：株価暴落・為替相場の激変などにより資産価値が大幅に下落するリスク、および貸付先企業の倒産などにより貸倒れが急増するリスク

\*Preliminary につき、著者に断りなく引用しないで下さい。

†情報・システム研究機構 統計数理研究所モデリング研究系/リスク解析戦略研究センター：  
〒106-8569 東京都港区南麻布 4-6-7

- 経営管理リスク：業務の運営上通常の予想を超えて発生し得るリスク

の合計として与えられる。SSTの狙いは、各保険会社が実際に負っているリスクに基づき支払い余力の基準を与えることである。そのためには、各社が自己のリスクを把握し、それを目標資本(target capital)の算出に組み入れる責任を負うことが求められる。義務ばかりではなく、そうするインセンティブも当然与えられなければならないわけで、透明性と競争力が増し、リスク管理がきちんとできている会社ほど報われる仕組みになっている。

SSTを構成する基本コンセプトを、SST白書の序章に做って列挙しておこう。ひとつは透明性である。これは、リスク・ベースで支払い余力を算出することによって実現されるとする、という基本コンセプトを指している。第2は資産・負債の整合的価格付けである。これは、時価評価を基本とし、市場と整合的に資産・負債評価を行うということである。第3は、リスク管理への誘因である。これは、標準(ひな形)モデルを与えると同時に内部モデルの開発を奨励するものである。その目的は、より精度の高い内部モデルを考えることでリスクの過大な見積もりを避けることができるような仕組みを考え、金融機関にリスク管理へのインセンティブをあたえることにある。第4は、破綻はゼロではないと前提することである。破綻に備えて、経営環境悪化時の是正措置が高じられなければならない。

第5の論点は、Solvency IIとの整合性である。Solvency IIとは、支払い余力に関する国際的新基準で、銀行業界におけるBasel IIと類似した位置づけを持つものである。その構成も、以下に見るようにBasel IIと類似している。

- Pillar 1: 最小資本要請
  - － 法定準備(statutory requirements)とSolvency I
- Pillar 2: 監督官庁による検査
  - － 各種金融・保険リスクの計算による目標資本(target capital)の決定[SSTの枠組みによる]
- Pillar 3: 情報開示

第6の論点は、最小支払能力と対象資産である。最小支払能力(minimum solvency)とは、貸借対照表上法的に規定される額であって、算出は容易だが、個々の保険会社の資産構成に伴うリスク・エクスポージャーを反映していない。これに対して目標資本(target capital)は、リスク生成資産(Risk-Baring Capital, 以下RBCと略記)と対応している。RBCは市場整合的(時価評価)であり、リスク・ベースの評価である。対象資産に関する「基準」が達成されなければ、監督官庁からは是正措置が取られることになる。

実際にこうしたリスク・ベースの規制・監督を導入している国は多数あり、フィンランド、カナダ、米国、オーストラリア、英国、シンガポール、オラン

ダなどがそうである。また、これらの国における制度は、多かれ少なかれ国際アクチュアリー会 (IAA) による勧告を受け入れて設計されている。SSTはその多くを取り入れているが、特徴としては以下のような点が挙げられる。

- リスク測度としての期待ショートフォールの採用
- トータル・バランスシート・アプローチ
- 計測期間は1年単位
- 明示的なリスク・マージンの算出

## 2 基本要素と分析の流れ

ここでは、SSTを構成する基本要素を逐次概観する。

### 2.1 分析の流れ

まず、分析の流れはSST白書のFigure 2(本稿で図1として引用)に示されている。それによると、SSTとは、資産、負債、信用リスク等に対する1組の標準モデルから成る。信用リスクモデルを除いては、標準モデルの結果は確率分布で与えられる。ここでの確率分布とは、モデル化されたリスク要因に起因するリスク生成資本(RBC)の変動の分布である。アクチュアリーは、その会社特有のリスクをうまく捉えるようなシナリオを選択肢から選び、その影響を加味する。標準モデルの結果とシナリオ分析の結果は、それらの加重平均により統合される。(統合方法の詳細はSST白書第8節。)

### 2.2 標準モデル

SSTが含む標準モデルがカバーするリスクは、以下の通りである。ただし、各モデルの具体像は白書のみからは読み取れない。

- 市場リスク
- 生保リスク
- 非生保(損保)リスク
- 医療保険リスク
- 信用リスク

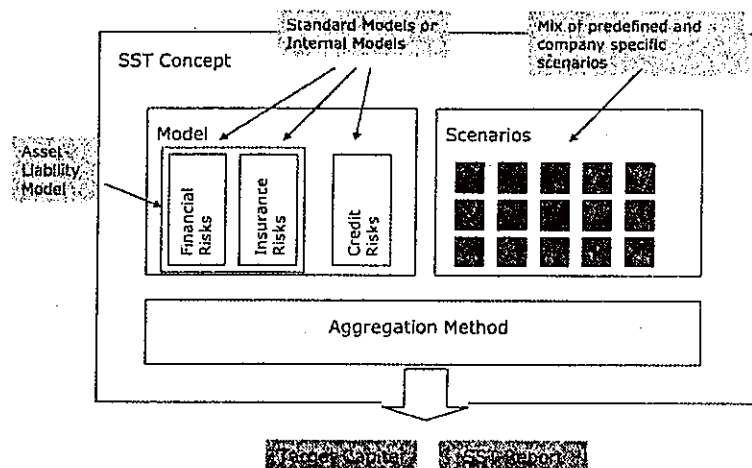


図 1: SST の概念図

信用リスクに関しては、Basel II で採用されている標準的方法に従うとされている。

標準モデルのパラメータには、幾つかのクラスが設定されている。このうち、タイプ1のパラメータとは、規制当局が定めるもので勝手な変更が許されないものを指す。具体的には、無リスク金利等マクロ経済環境に関するもの、シナリオ分析の対象事象に関するパラメータなどである。これに対して、各生保会社が決めて良いパラメータをタイプ2パラメータと呼んでいる。これは例えば、各社の資産構成に応じたボラティリティの推定値等が該当する。タイプ3パラメータとは、規制当局が定めるが、各保険会社が変更してよいものを指す。実際上は、殆どのパラメータはこのクラスに属する。推定法のガイドラインは規制当局が与えるが、各社は実際の推定方法を規制当局に報告する義務がある。

### 2.3 市場整合的評価

可能な場合は観測される市場価格をもとに資産評価を行う。非流動的資産の場合など、直接価格を観測することが難しい場合は、比較可能な資産の価格等から代理変数を構成する。(当然モデルに依存する。) SST では、自社株を除く全ての負債は、BS に計上されていなくても考慮に入れる。評価にあたっては、商品に組み入れられたオプションや保証の影響も明示的に考慮される。

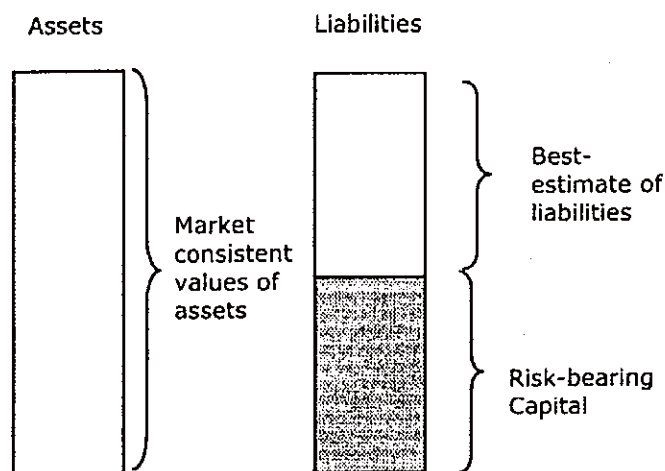


Figure 3. Definition of risk-bearing capital

図 2: リスク生成資本の定義

## 2.4 目標資本

まず資産の時価評価を行う。次に負債の最良推定値 (best estimates) を算出する。その差がリスク生成資本 (RBC) である。白書 Figure 3(本稿では図 2) 参照。目標資本とは、リスク生成資本 (RBC) を元に、保険会社が負っているリスクと、資本準備として要する額とを関連づけたものである。SST の計測は 1 年単位である。目標資本とは、1 年の終わりに、全負債をカバーするのに、ある信頼水準で必要とされる資本額のことを指す。目標資本は、

$$\text{目標資本} = \text{RBC の期待ショートフォール (ES)} + \text{リスク} \cdot \text{マージン}$$

として定義される。Figure 4(本稿では図 3) 参照。

目標資本とは、基本的にはリスク資産の変動性をカバーするだけの十分な資本額であるから、一年以内に生じるリスクに必要な資本は ES で計測する。SST 白書 Figure 7(本稿では図 4) も参照。一方リスク・マージンとは、仮に全負債を返済するとその際必要となる規制資本に対する (仮想的) コストのことである。例えば、ある保険会社の資産と負債を別の保険会社が引き継ぐとなったとき、規制資本の積み上げコストはポートフォリオ内で賄われなければならない。(そうでなければ引き受ける第 3 者はいないだろう。)

では、負債の最良推定とは何であろうか。市場整合的に負債を評価する方法に決め手はないが、いずれにせよ保険契約に組み込まれたオプションは全て評価しなければならない。標準的には、発生が予想されるキャッシュフローを無

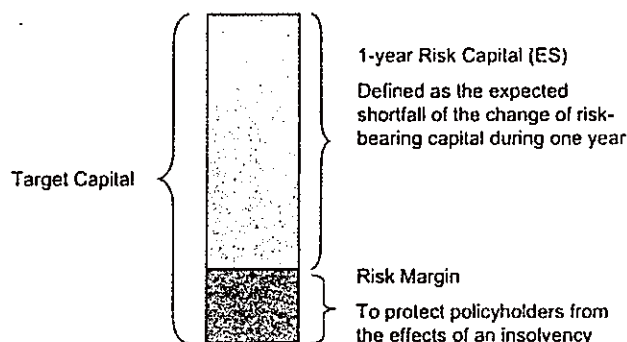


Figure 4. Target capital as the sum of 1-year risk capital and the risk margin

図 3: バリュアットリスクと期待ショートフォール

リスク金利で割り引くことになる。保険契約者が合理的に行動しないこともモデル構築上考慮する。年金等では政治的な要因に大きく左右されることもあるが、幾つかの合理的仮定を置いて話を進める。結局は best effort estimates ということを目指しているものと思われる。

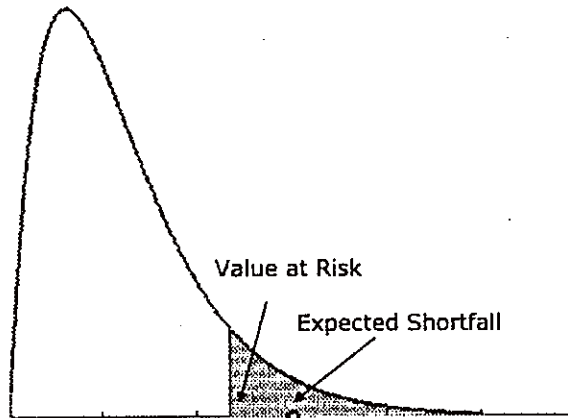
ともあれ、 $RBC < \text{目標資本}$  となった場合に、改善計画の徴求や新契約引受の停止など何らかの監督上の措置がとられる。

### 3 標準モデル (再び)

ここでは SST で与えられる標準モデルをより詳しく見ていこう。

#### 3.1 資産モデル

標準モデルにおける資産モデルは、基本的には RiskMetrics に準じたものである。SST 白書の p.19 に列挙されているような、23 個のリスクファクターからなっている。全てのリスク・ファクターに関し、その変化率は平均 0 の正規分布と仮定され、複数のリスク・ファクターの同時的振る舞いは共分散行列で記述する。リスク生成資本 (RBC) の変化率は、各リスク・ファクターの変化率の線形関数と仮定する。従ってトータルでのボラティリティは、共分散と傾きから決まる。デリバティブ等に起因する非線形性は標準モデルの範囲外である。



**Figure 7. Comparison of Value at Risk and Expected Shortfall**

図 4: バリュアットリスクと期待ショートフォール

### 3.2 生保リスク

生保リスクにおけるリスク・ファクターは、SST 白書 p.21 に列挙されている。各リスクファクターの変化率は正規分布と仮定し、各リスクファクターに関する RBC の感度 (回帰係数) を計算する。各リスクファクターの相関行列とボラティリティは、適切なデータが殆どないので、有識者との協議を行い、意見を集約した上で監督官庁が定める。SST 白書 p.22 を参照。

### 3.3 損保モデル

損保モデルは、これまで述べてきたようなファクターモデルではない。換言すれば、正規分布ではない確率分布を明示的に仮定している。実収保険料、コスト、将来発生する請求 (クレーム)、準備金 (引当金) 等の technical result に起因する RBC の年次変化をモデル化する。また、将来損失のモデル化も重要であるが、これは巨額損失と少額損失とでモデル化が別に行われる。大災害による損失はシナリオでモデル化することになる。また、準備金の変化は各商品 (LoB: Lines of Business) ごとにモデル化する。核事故、飛行機事故、ダム災害等に対する損失プールは、関係する会社により明示的にモデル化する。



将来の少額損失の当期計測については、商品 (LoB) ごとに、将来の実収保険料とそれに対する損失率の変動を推定する。上記を元に、商品間の相関行列を算出する。全商品合算での少額損失の平均と分散を求める。これらの損失をガンマ分布でモデル化する。その際、モーメントマッチングでガンマ分布をあてはめる。(24 ページを見ると対数正規分布か?) 内部モデルで分布型の仮定は変更して良いことになっている。

将来の巨額損失の当期計測については、各商品 (LoB) ごとに、複合ポアソン分布でモデル化する。請求件数はポアソン分布に従うものとする。請求の強度 (severity) はパレート分布に従うとし、そのパラメータは事前に与えておく。パレート分布に関しては切断型の裾を使っても良いが、切断する点の適切さに関しては、監督官庁のガイドラインに従う。

前年度発生損失については、準備金の増減に関する確率分布を求めることが必要となる。商品ごとにヒストリカル・ボラティリティを計算し、商品毎準備金の独立性を仮定、分散の和を求める。上記の分散を持つ逆対数正規分布 (ただし平均がゼロになるようシフト) を準備金増減の確率分布とする [SST 白書 p.23, Figure 13]。将来の支払いパターンを推定した上で、所与のリスクフリー金利でキャッシュフローを割り引く。

損失分布の集計 (aggregation) は、以下のような手順で行う。最初に、少額損失の分布と、年次準備金変化の分布とを集計する。所与の相関行列で和の平均と分散を考え、それにマッチさせる形で対数正規分布をあてはめる。巨額損失は少額損失・準備金変動とは独立に発生していると仮定し、複合ポアソン-パレート分布を畳み込みで集計する。(SST 白書 p.24, Figure 14)

### 3.4 その他の側面

標準モデルに関するその他の側面について簡単にコメントしておく。まず医療保険に関しては、個人対象の商品と、グループ対象商品とを区別し、相関をとって集計する。信用リスクについては、オペレーショナル・リスクを除き、Basel II の標準的なアプローチを採用している。再保険については、複雑多岐にわたるため、標準モデルは存在しない。

## 4 シナリオ分析

シナリオ分析は、標準モデルが表現しきれない tail behavior を補うものであり、企業を取り巻く金融環境に対する負のインパクトに対応する多数の逆シナリオから構成されている。その意味では、単一のリスクファクターを強調するストレステストとは異なるものであり、SST を構成する重要な要素である。

シナリオ分析には、定性的シナリオ、定量的シナリオが考えられるが、定量的シナリオでは生起確率が必要となる。これについては監督官庁が与えある場合もあるし、保険会社に所属する保険数理人が算出してもよい。

シナリオの具体例としては、以下のようなものが挙げられる。

- 工場の爆発
- 流行病
- 雹を伴う嵐
- 集中的水害
- 傷害
- 日当
- 再保険会社の破綻
- 金融危機
- 準備金、引当金
- 逆選択
- テロリズム
- 長生きリスク
- 歴史的金融危機シナリオ

SSTとは確率モデルとシナリオ分析のハイブリッドモデルであり、シナリオと標準モデルの統合が行われる。シナリオで想定しているイベントが起きるということは、追加的損失を意味するので、結果的にRBCは減少する。シナリオによっては、関心を集中しているリスク以外の他のリスクは一定と仮定する場合もある。

シナリオの定量的表現は、当然確率分布による。シナリオにあるイベントが引き起こすものは、例えば分布のシフトであったり、標準モデルとは形状の異なる確率分布を仮定する場合もある。最終的には、標準モデルから得られる分布と加重平均を取り、シナリオ効果を集計する [SST 白書 p.28, Figure 15].

## 5 内部モデル

SSTでは、各保険会社の事業内容や資産構成に応じた内部モデルの開発が奨励されているわけだが、その狙いは、皆が標準モデルに甘んじることで生

じる、システミックリスクを回避する点にある。ただ、内部モデルとはいっても、全く好き勝手にモデルを組み立ててよいはずはなく、監督官庁の許可を要する点、必要要件がある。それらは定性的要請であったり、定量的要請であったり、あるいは組織面での要請であったりする。内部モデル開発に対するインセンティブは、標準モデルのみではリスクを高めに見積もるようなデザインになっているところから生じるものと考えられている。例えば、再保険会社の破綻に対する備えはその一例になっている。

最終的に、保険会社は自己のリスクポジションを要約したものを SST レポートとして提出する。これには CEO がサインし、年次で監督官庁に報告する義務を負う。具体的には、SST 目標資本とその担保を含む、保険会社のリスクポジションに関する報告と、リスク管理報告書とからなる。

## 6 オペレーショナル・リスク

オペレーショナル・リスクについては、データの蓄積が浅いこともあって、現段階での定量化は難しく、恣意的になりがちである。ただ、Basel II に向けて銀行業界はデータが貯まりつつあるので、保険業界でも同様の流れを意識した方向に進んでいることは確かである。もっとも、定量化に向けては、コーポレート・ガバナンスに関する何らかの指標が必要である。SST 白書が書かれた 2004 年の時点では、自己評価アンケートの返送による定性的評価が中心に据えられているが、昨今オペレーショナル・リスクに関心が集まる中、徐々に定量的把握に向けて進歩が期待されるところである。

## 7 まとめ

Swiss FOPI が 2004 年に公表した SST 白書の内容をおおまかに紹介することで、リスクベースのソルベンシー基準が将来的にどのように構成されるのか、その見通しを概観した。銀行業は既に Basel II の導入によって、当局管理型の監督から、自己管理と市場規律を中心とした監督に移行しつつある。Solvency II、それをふまえての SST も、理念としては同じ流れに沿うものであるが、銀行業と保険業では、国際的統一基準への統合速度がかなり異なり、Solvency II はまだ議論の途上にあるとも言える。

SST の背後にあるのは、本文中でも言及した国際アクチュアリー協会や、あるいは保険監督者国際機構 (International Association of Insurance Supervisors) であるが、SST 白書は IAIS が 2005 年 10 月に採択した保険会社の健全性評価のための「コーナーストーン」に先駆けて提示されている。つまり、SST はスイス国内仕様のようなものでありながら、リスクベースのソルベンシー評価に関しては議論をリードする役割を担っているようにも見える。

また、SST 白書の巻末には、プロジェクトに協力した研究者の一覧が付されているが、確率論・統計学で著名な研究者の名前が多数記載されている。我が国でのリスクベースの監督基準が確立していく過程においては、仮に標準モデルの内容を理解するにせよ、内部モデルを構築するにせよ、統計科学の知識は必須であろう。リスクベースのソルベンシーは、統計科学の研究者が今後どのような役割が果たせるか検討しておくだけの価値のある分野と思われる。

## 参考文献

- [1] Swiss Federal Office of Private Insurance, *White Paper of the Swiss Solvency Test*, November 2004.
- [2] Keller, P., *A Primer for Calculating the Swiss Solvency Test "Cost of Capital" for a Market Value Margin*, April 2006.