

保険と金融の統計学 II ¹

国友直人 編²

2008年2月

¹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2007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。

²東京大学経済学研究科教授

概要

この研究報告書は2007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、研究報告などの一部分をまとめたものである。2007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まとまった形で日本語では利用可能でない保険分野や金融分野に係る統計学・確率論からのリスク分析が扱われている。今年の報告内容としては、日本の損害保険データの分析に関する増田・国友による報告論文「Lasso 分位点回帰の理論と損害保険への応用」、渋谷政昭氏（慶応大学工学部名誉教授）による講義録「極値理論とその応用」、西山陽一氏（統計数理研究所）による講義録「マルチンゲール理論の基礎」、Guus Balkema 氏（Amsterdam 大学名誉教授）による講義録「High Risk Scenarios and Extremes」、が具体的内容である。これらの話題は研究会に参加・聴講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、と判断した。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おり、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。

2008年2月

編者

目次

第 I 部：損害保険の統計学

1. 「Lasso 分位点回帰の理論と損害保険への応用」 増田智巳・国友直人

第 II 部：金融リスクの確率論・統計学：連続講義録

2. 「極値理論とその応用」 渋谷政昭
3. 「マルチンゲール理論の基礎」 西山陽一
4. 「High Risk Scenarios and Extremes」 Guus Balkema

Lasso 分位点回帰の理論と損害保険への応用*

増田智巳・国友直人†

2008年2月26日

概要

分位点回帰 (Quantile Regression) の方法を損害保険におけるリスク要因のデータ解析へ応用した。分位点回帰における推定問題を大規模な線形計画問題として問題を定式化した上で数値的に Lasso 法を実現し、自動車保険の保険請求 (クレーム) 額の分析を行った。分位点に依存する説明変数の分析より高額クレームのリスク要因を特定化できることが分かった。

1 はじめに

統計的分析を利用する多くの応用分野と同様に金融リスク管理や保険リスク管理で遭遇する様々な場面においても回帰分析 (regression analysis) がしばしば利用されている。特にある変数 (被説明変数と呼ぶ) のリスク要因分析においては説明変数に回帰して求められた回帰係数をリスクに対する感応度と見なしてしばしば統計的な解析が行われている。統計学ではよく知られているように、線形回帰モデルは被説明変数の期待値 (確率分布の意味での平均値) を定数を含む幾つかの説明変数の線形結合により説明する統計的モデルである。実際的な問題では被説明変数のある種の平均的傾向を説明することに大きな意味がある問題も少なくないが、金融や保険におけるリスク分析、特に損害保険データによる保険リスク分析においてはその利用における基本的問題も存在すると考えられる。例えば本稿で事例として用いる損害保険の損失データは大きく右に歪んでいるのが一般的である。またリスク管理の問題としては損失額リスクを損失額分布 (loss distribution) という確率分布で表現するとき、少額の損失と多額の損失とではリスクの意味合いもかなり異なっている。したがって、こうした応用例においては通常回帰分析のように被説明変数の期待値を説明するだけでなく、より一般的な方法により確率分布を説明変数により説明する統計的モデルや統計的方法の開発が求められている。

*本稿は東京大学公共政策大学院・損保ジャパン寄付講座・研究プロジェクトの報告として書かれた未定稿の原稿である。本稿で利用したデータを提供してくれた損保ジャパンの足立尚人氏に感謝する。なお、本稿の内容に関する責任は著者のみにあり、損保ジャパンは無関係である。また草稿に対する加藤賢吾氏の有益なコメントに感謝する。

†東京大学大学院経済学研究科

本稿では被説明変数の分位点を説明する回帰モデルを考察するが、この種の統計モデルは分位点回帰 (Quantile Regression) 問題と呼ばれている。分位点回帰モデルは最小絶対偏差法 (Minimum Absolute Deviation Method) の発展型として Koenker & Bassett (1978) によって提唱され、発展している統計的モデルである。例えば線形回帰モデルの元では誤差の絶対偏差最小化により得られる係数は中央値回帰 (Median Regression) に対応していることはよく知られている。このことは特に説明変数が定数のみであれば、被説明変数の中央値 (median) が定数の推定値となることから類推すると分かり易いであろう。分位点回帰では中央値を含む任意の分位点における説明変数の効果が同一とは限らないことが一つの特徴である。そこで分位点回帰 (Quantile Regression) の理論と損害保険分野への応用を考察するが、これまでに分位点回帰モデルの統計的な推定方法としては Koenker が開発した一般化絶対偏差最小化法がよく知られており、実際に経済の実証分析へ応用された例もある。一般化絶対偏差最小化の推定方法は漸近的にはよい性質を持つことが知られているが、実データを用いるモデルの推定では、大規模な線形計画問題として定式化し、数値的に解く必要が生じる。このように分位点回帰の理論は標準的な統計学の教科書で議論されている通常の線形回帰分析の議論とはかなり異なる側面があるので、理解可能な形で説明することにも大きな意義があると考えられよう。さらに線形回帰問題における実際的に重要な問題として説明変数の選択問題が古くから議論されてきている。分位点回帰問題においては説明変数のリストが分位点にも依存する、というさらなる問題が存在する。そこで Tibshirani (1996) が線形回帰モデルに対する応用を提唱している Lasso (least absolute shrinkage and selection operator) と呼ばれる統計的方法を分位点回帰に応用した Lasso 分位点回帰法を考察する。

本稿の主たる目的は分位点回帰の統計的方法を説明すると共に、損害保険のリスク解析に Lasso 分位点回帰を応用することである。保険の分野においてはリスクの多様化に対応して様々な保険商品が登場してきているが、保険会社は適切に支払保険金を見積もり、適切な保険料を定める必要がある。特に損害保険では損失分布は右に大きく歪んでいるので、分位点回帰を用いることで支払い保険金、つまりクレーム額のデータ分析を行うことが重要であり、特に被保険者の特性に依存した支払い保険金の分布をより正確に推定することが必要となる、と考えられる。すなわち、保険リスクのデータ分析において分位点回帰は広い応用が期待できるのである。ここでは実例として自動車保険のクレーム額についてのデータ分析を行い、その分析結果も報告する。我々の分析により特に分位点回帰に Lasso 罰則を付与することにより、高分位点の分位点回帰における変数選択を行うことで、保険支払リスクの要因を詳細に分析することが可能であることが分かった。

ここであらかじめ本稿の構成を説明しておこう。第2章では分位点回帰モデルと推定の漸近理論など必要な理論を説明する。次に第3章では Lasso 分位点回帰法について説明し、第4章ではその方法を用いて実際に自動車保険データについての実証分析を行った結果を説明し、第5章では結論を述べる。

2 分位点回帰の理論

分位点回帰は Koenker & Bassett (1978) により導入された統計モデルであり、被説明変数の条件付分位点を推定する方法と見なすことができる。通常の回帰分析においては二乗損失関数を用いて統計モデルの母数を推定し、条件付平均関数を求める最小二乗法を利用するのが一般的である。これに対して分位点回帰では絶対偏差加重和という損失関数を用いて母数を推定すると、 $(0, 1)$ 上の任意の下側 τ 点に対応する分位点について近似する関数が推定できる。このことから条件付分布の中心についての統計的推論が可能であるだけでなく、さらに確率分布の裾の挙動についても統計的解析を行うことができる。分位点回帰法に関する様々な問題については Koenker (2005) がかなり包括的に議論しているが、本稿では応用に興味を持つ読者を想定して、まず分位点回帰法における重要な論点を要約しておく。

2.1 分位点回帰モデル

確率変数 Y の分布関数を $P(Y \leq y) = F_Y(y)$ 、 Y の τ ($0 < \tau < 1$) 分位点を $Q_\tau(Y) = \inf\{y | F_Y(y) \geq \tau\}$ としよう。さらに説明変数ベクトル $X = (X_k)$ が与えられた時の確率変数 Y の条件付分布関数を $P(Y \leq y | X) = F_Y(y | X)$ 、条件付 τ 分位点を $Q_\tau(Y | X) = \inf\{y | F_Y(y | X) \geq \tau\}$ とする。このとき分位点回帰モデルは

$$(1) \quad \begin{aligned} Q_\tau(Y | \mathbf{X}) &= \beta_0(\tau) + \beta_1(\tau)X_1 + \cdots + \beta_p(\tau)X_p, \\ &= \beta_0(\tau) + \mathbf{X}'\boldsymbol{\beta}(\tau). \end{aligned}$$

と表現される。ただし $\boldsymbol{\beta}(\tau) = (\beta_1(\tau), \dots, \beta_p(\tau))'$ は定数項を表す母数 $\beta_0(\tau)$ を除く未知母数ベクトル、 $\mathbf{X}' = (X_1, \dots, X_p)$ は説明変数ベクトル、 $Q_\tau(y | \mathbf{x})$ は $\mathbf{X} = \mathbf{x}$ が与えられた時の y の条件付 τ 分位点関数である。ここで確率変数 U を

$$(2) \quad U = Y - [\beta_0(\tau) + \mathbf{X}'\boldsymbol{\beta}(\tau)]$$

により定義すれば、分位点回帰モデルは

$$(3) \quad Y = \beta_0(\tau) + \mathbf{X}'\boldsymbol{\beta}(\tau) + U$$

とも表現できる。この表現は統計的線形回帰モデルに類似しているが、右辺の母数の係数ベクトルが τ に依存し、誤差項 U の分布関数を $F_U(u)$ とすると

$$(4) \quad F_U(0) = P\left(Y \leq \beta_0(\tau) + \mathbf{X}'\boldsymbol{\beta}(\tau)\right) = \tau$$

となる。

ここで被説明変数 Y と説明変数ベクトル \mathbf{X} について互いに独立な n 個のデータの組が得られる状況を想定しよう。さらに $\mathbf{y} = (y_1, \dots, y_n)'$ を被説明変数ベクトル ($n \times 1$)、 $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})'$ ($i = 1, \dots, n$) を説明変数ベクトル ($p \times 1$)、 $\mathbf{u} = (u_1, \dots, u_n)'$ を $F_{u_i}^{-1}(\tau) = 0$

を満たす独立同一分布 F_u に従う誤差ベクトル ($n \times 1$) とする。本稿では議論の簡単化の為に分位点回帰モデルにおいて条件 $n \geq p + 1$ が成立し、被説明変数の分布関数は Lebesgue 測度に関して絶対連続の場合を考察する。この仮定の下では密度関数が存在するので分位点関数は分布関数の逆関数として一意的に定義でき、統計的分析がかなり簡単化される。ただし損害保険など実際に生じる多くの応用上ではそれほど大きな問題は生じない。

被説明変数 Y と説明変数ベクトル \mathbf{X} の n 組のデータより母数ベクトル $((p + 1) \times 1)$ $\beta^*(\tau) = (\beta_0(\tau), \beta'(\tau))'$ の推定問題を考えよう。 ($\beta_0(\tau)$ は定数項の係数である。) この問題については Koenker & Bassett(1978) は損失関数として

$$(5) \quad L(u) = \rho_\tau(u) = (\tau - 1(u < 0))u$$

を用いた (ここで $1(\omega)$ は ω が成立すれば 1, その他は 0 という指示関数とする)。そこで n 組のデータより評価基準

$$(6) \quad \min_{\beta_0, \beta(\tau)} \sum_{i=1}^n \rho_\tau(y_i - \beta_0(\tau) - \mathbf{x}'_i \beta(\tau)) .$$

を最小化する推定方法を考察しよう。ここで最小化問題の解を

$$(7) \quad \hat{\beta}^*_n(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}'(\tau))' = \arg \min \sum_{i=1}^n \rho_\tau(y_i - \beta_0(\tau) - \mathbf{x}'_i \beta(\tau))$$

と表す。また定数項を含む説明変数ベクトルを $\mathbf{x}^* = (1, \mathbf{x}')$ 、その説明変数ベクトルのデータを $\mathbf{x}^*_i = (1, \mathbf{x}'_i)$ ($i = 1, \dots, n$) と表しておくとなら分析がよりわかりやすい。

損失関数 $L(u)$ より得られる期待損失を $E[\rho_\tau(Y - \mathbf{x}^* \beta^*(\tau))]$ とする。このとき分位点回帰で得られる推定値 $\mathbf{x}^* \tilde{\beta}^*$ はこの期待損失を最小にする母数 $\beta^*(\tau)$ に対応すると見なすことができる。このことは自明ではないので確認しておこう。

まずケース I として $\Pr(Y \leq \mathbf{x}^* \tilde{\beta}^*(\tau)) \geq \tau$, $\Pr(Y \geq \mathbf{x}^* \tilde{\beta}^*(\tau)) \geq 1 - \tau$, (また一般性を失うことなく) $0 < \tau \leq 0.5$ となる場合を考察しよう。特に $\mathbf{x}^* \tilde{\beta}^*(\tau) < \mathbf{x}^* \beta^*(\tau)$ となる時には

$$\begin{aligned} & E[\rho_\tau(Y - \mathbf{x}^* \beta^*(\tau)) - \rho_\tau(Y - \mathbf{x}^* \tilde{\beta}^*(\tau))] \\ &= \int_{y < \mathbf{x}^* \beta^*(\tau)} (\tau - 1)(y - \mathbf{x}^* \beta^*(\tau)) dP(y) + \int_{y \geq \mathbf{x}^* \beta^*(\tau)} \tau(y - \mathbf{x}^* \beta^*(\tau)) dP(y) \\ &\quad - \int_{y < \mathbf{x}^* \tilde{\beta}^*(\tau)} (\tau - 1)(y - \mathbf{x}^* \tilde{\beta}^*(\tau)) dP(y) - \int_{y \geq \mathbf{x}^* \tilde{\beta}^*(\tau)} \tau(y - \mathbf{x}^* \tilde{\beta}^*(\tau)) dP(y) \\ &= \mathbf{x}^* (\beta^*(\tau) - \tilde{\beta}^*(\tau)) \left[(1 - \tau) \Pr(Y \leq \mathbf{x}^* \tilde{\beta}^*(\tau)) - \tau \Pr(Y > \mathbf{x}^* \tilde{\beta}^*(\tau)) \right] \\ &\quad + \int_{\mathbf{x}^* \tilde{\beta}^*(\tau)}^{\mathbf{x}^* \beta^*(\tau)} (\mathbf{x}^* \beta^*(\tau) - y) dP(y) \end{aligned}$$

となる。したがって $\Pr(Y \leq \mathbf{x}^* \tilde{\beta}^*(\tau)) \geq \tau$ より $\Pr(Y > \mathbf{x}^* \tilde{\beta}^*(\tau)) \leq 1 - \tau$ となることを用いれば

$$(8) \quad E[\rho_\tau(Y - \mathbf{x}^* \beta^*(\tau)) - \rho_\tau(Y - \mathbf{x}^* \tilde{\beta}^*(\tau))] \geq 0$$

が得られる。ここで等号は

$$(9) \Pr(Y \leq \mathbf{x}'\tilde{\beta}^*(\tau)) = \tau, \Pr(Y > \mathbf{x}'\tilde{\beta}^*(\tau)) = 1 - \tau, \Pr(\mathbf{x}'\tilde{\beta}^*(\tau) < Y < \mathbf{x}'\beta^*(\tau)) = 0$$

のときに成立する。次にケースIIとして $\mathbf{x}'\tilde{\beta}^*(\tau) > \mathbf{x}'\beta^*(\tau)$ となる場合を考察しよう。ケースIの議論と同様な議論により

$$\begin{aligned} & E[\rho_\tau(Y - \mathbf{x}'\beta^*(\tau)) - \rho_\tau(Y - \mathbf{x}'\tilde{\beta}^*(\tau))] \\ &= \mathbf{x}'(\tilde{\beta}^*(\tau) - \beta^*(\tau)) \left[(1 - \tau) \Pr(Y \leq \mathbf{x}'\tilde{\beta}^*(\tau)) - \tau \Pr(Y > \mathbf{x}'\tilde{\beta}^*(\tau)) \right] \\ & \quad + \int_{\mathbf{x}'\beta^*(\tau)}^{\mathbf{x}'\tilde{\beta}^*(\tau)} (\mathbf{x}'\beta^*(\tau) - y) dP(y) \end{aligned}$$

となる。(なお等号が成立する条件はケースIの場合と同様である。) したがって、ケースIとケースIIを合わせると主張が確認される。

2.2 推定量の漸近的性質

分位点回帰モデルでは母係数ベクトルの推定が重要な統計的問題である。母係数ベクトル $\beta^*(\tau) = (\beta_0(\tau), \beta'(\tau))'$, 推定量ベクトル $\hat{\beta}_n^*(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}'_n(\tau))'$, 真の母数値ベクトル $\beta^{**}(\tau) = (\beta_0^{(0)}(\tau), \beta^{(0)'(\tau)})'$ としておこう。分位点回帰推定量は標本数 n が大きいときには一貫性 (consistency) と漸近正規性 (asymptotic normality) を持つことが知られている。この推定量の漸近的性質は次のようにすると比較的容易に導くことができる。ここでは確率変数列 (y_i, u_i, \mathbf{x}_i) (ただし $\mathbf{x}_i^* = (1, \mathbf{x}_i)'$ $i = 1, \dots, n$) について次のような標準的な正則条件¹を考察しよう。

- (A1): $F_u(u)$ (\mathbf{x}_i を所与とする u_i の分布関数) が絶対連続かつ $f_u(u) > 0$ となる。
- (A2): $(y_i, \mathbf{x}_i) \sim \text{i.i.d}$ (i について互いに独立に同一分布にしたがう)。
- (A3): $0 < E[\mathbf{x}_i \mathbf{x}_i'] < \infty$ 。
- (A4): $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^* \mathbf{x}_i^{*'} = \mathbf{C}$ が存在する。
- (A5): $f_u(0)\mathbf{C}$ は正値定符号行列となる。

まず評価関数の差

$$(10) \quad Z_n = \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^{*'} \beta^*(\tau)) - \sum_{i=1}^n \rho_\tau(u_i)$$

¹条件(A1)と条件(A5)は被説明変数の分布関数 $F_Y(y|X)$ についての条件とみれば、密度関数 $f_Y(Q_\tau(y|X))$ についての条件に対応する。関連して2.3節の議論を参照されたい。むしろ通常の回帰分析での議論と同様にここで利用している正則条件(A1)-(A5)を弱めたり、あるいは様々な方向の条件の下での議論に拡張できる。

を考えよう。さらに $\delta(\tau) = \beta^*(\tau) - \beta^{**}(\tau)$ (ここで $\beta^{**}(\tau)$ は真の母数ベクトル)、 $\alpha_i = \mathbf{x}_i' \delta(\tau)$ とする。ここで等式

$$(11) \quad \rho_\tau(u-v) - \rho_\tau(u) = -v[\tau - \mathbf{1}(u < 0)] + \int_0^v [1(u \leq s) - 1(u \leq 0)] ds$$

を利用する(ただし $\mathbf{1}(\omega)$ は指示関数である)。そこで

$$\begin{aligned} S_n(\delta(\tau)) &= \frac{1}{n} Z_n(\delta(\tau)) \\ &= \frac{1}{n} \sum_{i=1}^n (-1) \alpha_i [\tau - \mathbf{1}(u_i < 0)] + \frac{1}{n} \sum_{i=1}^n \int_0^{\alpha_i} [1(u_i \leq s) - 1(u_i \leq 0)] ds \\ &\equiv S_{1n} + S_{2n}, \end{aligned}$$

と表現できるので、右辺の各項の収束を考えることができる。分位点の仮定より $E[\tau - \mathbf{1}(u_i \leq 0)] = 0$ 及び $E\|\mathbf{x}_i\|^2 < \infty$ となるので項 $\sum_{i=1}^n \alpha_i [\tau - \mathbf{1}(u_i < 0)]$ は(二乗可積分)マルチンゲールである。したがって、マルチンゲール収束定理より

$$(12) \quad S_{1n} \xrightarrow{P} 0$$

となる。また S_{2n} の各項は有界な確率変数であるので $\delta(\tau)$ に関して一様に収束して

$$(13) \quad S_n(\delta(\tau)) \xrightarrow{P} S(\delta(\tau)) = E_G \left[\int_0^{\mathbf{x}_i' \delta(\tau)} (F_u(s) - F_u(0)) ds \right]$$

が得られる。ただし G は \mathbf{x}_i の分布関数をあらわすものとした。ここで右辺で定義される極限の評価関数の性質を調べると

$$(14) \quad \frac{\partial S(\delta(\tau))}{\partial \delta} = E_G \left[F_u(\mathbf{x}_i' \delta(\tau)) \mathbf{x}_i^* - F_u(0) \mathbf{x}_i^* \right],$$

$$(15) \quad \frac{\partial^2 S(\delta(\tau))}{\partial \delta \partial \delta'} = E_G \left[f_u(\mathbf{x}_i' \delta(\tau)) \mathbf{x}_i^* \mathbf{x}_i^{*'} \right] > 0$$

となる。特に $\delta_n(\tau) = \mathbf{0}$ のときには $\frac{\partial S(\delta(\tau))}{\partial \delta} = \mathbf{0}$ である。また $f_u(u) > 0$ より二階微係数行列は正定値行列となる。したがって、評価関数 $S(\delta(\tau))$ は凸関数となり、かつ $[\delta](\tau) = \mathbf{0}$ のとき最小値をとる。そこで漸近理論の標準的議論(例えば Amemiya (1985) の定理 4.1.1)より推定量の一致性を導くことができる。

次に推定量の漸近分布を考察しよう。まず $Z_{1n} = nS_{1n}$, $Z_{2n} = nS_{2n}$ として $Z_n(\delta(\tau)) = Z_{1n}(\delta(\tau)) + Z_{2n}(\delta(\tau))$ と表現しよう α_i をあらためて $\alpha_{in} = [\mathbf{x}_i/\sqrt{n}] \sqrt{n}(\beta^*(\tau) - \beta^{**}(\tau))$ と表現しておく。仮定 (A2)・仮定 (A3) より $\max_i \|\alpha_{in}\| \xrightarrow{P} 0$ である事に注意する。また $\alpha_{in} = [\mathbf{x}_i/\sqrt{n}] \delta_n^*(\tau)$ および $\delta_n^*(\tau) = \sqrt{n}(\beta^*(\tau) - \beta^{**}(\tau))$ とおき、 $\delta_n^*(\tau) \rightarrow \delta^*$ となる系列を考察する。

ここで第1項 Z_{1n} の極限を考察しよう。条件 (A2)-(A4) を用いて Lindeberg 条件下の中心極限定理を応用すると

$$(16) \quad Z_{1n}(\delta(\tau)) \xrightarrow{d} -\delta^*(\tau)' W$$

が成り立つ。ただし $W \sim \mathcal{N}(\mathbf{0}_{p+1}, \tau(1-\tau)\mathbf{C})$ であるが

$$\mathbb{E}[(\tau - 1(u_i < 0))^2] = \mathbb{E}[\tau^2 - 2\tau\mathbb{E}(1(u_i < 0)) + \mathbb{E}(1(u_i < 0))] = \tau(1-\tau)$$

より共分散項の表現が得られる。さらに

$$(17) \quad Z_{2n}(\delta(\tau)) = \sum_{i=1}^n \mathbb{E}[Z_{2ni}(\delta(\tau))] + \sum_{i=1}^n (Z_{2ni}(\delta(\tau)) - \mathbb{E}[Z_{2ni}(\delta(\tau))])$$

と表現しておく。 $n \rightarrow \infty$ のとき

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[Z_{2ni}(\delta(\tau))|X] &= \sum_{i=1}^n \int_0^{\alpha_{in}} [F_u(s) - F_u(0)] ds \\ &= \sum_{i=1}^n \int_0^{\alpha_{in}} [s f_u(0)] ds + o(1) \\ &= \frac{1}{2n} f_u(0) \sum_{i=1}^n \delta_n^{*'}(\tau) \mathbf{x}_i^* \mathbf{x}_i^{*'} \delta_n^*(\tau) + o(1) \\ &\rightarrow \frac{f_u(0)}{2} \delta^{*'}(\tau) \mathbf{C} \delta^*(\tau) \end{aligned}$$

となる。ここでさらに $V(Z_{2n})$ を評価すると被積分関数が有界で

$$\mathbb{E} \left[\int_0^{\alpha_{in}} [1(u_i \leq s) - 1(u_i \leq 0) - \mathbb{E}(1(u_i \leq s) - 1(u_i \leq 0))] ds \right] \leq \mathbb{E} \left[\int_0^{\alpha_{in}} ds \right]$$

となるので

$$(18) \quad V(Z_{2n}(\delta(\tau))|X) \leq \frac{1}{\sqrt{n}} \max |\mathbf{x}_i^{*'} \delta_n^*(\tau)| \sum_{i=1}^n \mathbb{E}[Z_{2ni}(\delta(\tau))] \xrightarrow{p} 0$$

である。したがって、条件 (A4) より

$$(19) \quad Z_n(\delta(\tau)) \xrightarrow{d} Z_0(\delta(\tau)) = -\delta^{*'}(\tau)W + \frac{f_u(0)}{2} \delta^{*'}(\tau) \mathbf{C} \delta^*(\tau)$$

となる。ここで右辺の極限確率変数 $Z_0(\delta^*(\tau))$ は凸関数であり、最小値を与える解は一意であるので、推定量の確率変数列は $n \rightarrow \infty$ のとき

$$(20) \quad \sqrt{n}(\hat{\beta}_n^*(\tau) - \beta^{**}(\tau)) = \arg \min Z_n \xrightarrow{d} \arg \min Z_0 = \frac{1}{f_u(0)} \mathbf{C}^{-1}W$$

となる。以上の議論より次のような結果が得られる。

定理 1: 分位点回帰推定量 $\hat{\beta}_n^*$ は漸近的 ($n \rightarrow \infty$) に次の性質を持つ。

(i) 条件 (A1), (A2), (A3) のもとで

$$(21) \quad \hat{\beta}_n^*(\tau) = \begin{bmatrix} \hat{\beta}_0(\tau) \\ \hat{\beta}(\tau) \end{bmatrix} \xrightarrow{p} \beta^{**}(\tau) = \begin{bmatrix} \beta_0^{(0)}(\tau) \\ \beta^{(0)}(\tau) \end{bmatrix}$$

が成り立つ。

(ii) 条件 (A1), (A3), (A4), (A5) のもとで

$$(22) \quad \sqrt{n} \left(\hat{\beta}_n^*(\tau) - \beta^{**}(\tau) \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}_{p+1}, \frac{\tau(1-\tau)}{f_u(0)^2} \mathbf{C}^{-1} \right)$$

が成立する。

ここで推定量の漸近正規性の導出については比較的簡単な方法を説明したが、この証明方法は Knight (1998) が絶対偏差最小化推定量についての議論を参考としている。この方法これまで知られていた証明方法よりもかなり一般的でありかつ簡明である²。

2.3 漸近共分散の推定

分位点回帰法においては、漸近分布の共分散行列に誤差の密度関数が表れるので実際に分析を行う際には漸近共分散を推定する必要が生じる。漸近共分散を直接推定する方法としてはブートストラップ法など統計的リサンプリング法の利用も考えられる。またカーネル密度推定を用いて密度関数を推定することで共分散を推定する方法 (Powell (1991)) も提唱されている。本稿ではより直観的にも分かり易い Hall & Sheather (1988) の方法を利用している。

ここで分位点回帰問題では (ある意味での誤差項の) 確率変数 U の分布関数 $F_U(0)$ は確率変数 Y の分布関数 $F_Y(Q_\tau(y)|X)$ に対応していることに注意しよう。(以下では煩雑さを避ける目的で分布関数 $F_Y(\cdot)$ 、密度関数 $f_Y(\cdot)$ により表現しておく。) 密度関数の逆数

$$(23) \quad s(\tau) = [f_Y(F_Y^{-1}(\tau))]^{-1}$$

を推定することを考える。ここで $F_Y(F_Y^{-1}(\tau)) = \tau$ より

$$(24) \quad \frac{d}{d\tau} F_Y^{-1}(\tau) = s(\tau)$$

となる。例えば Siddiqui (1960) は被説明変数 Y の経験分布関数 \hat{F}_n を用いているが、ここでも $s(\tau)$ を

$$(25) \quad \hat{s}_n(\tau) = [\hat{F}_n^{-1}(\tau + h_n) - \hat{F}_n^{-1}(\tau - h_n)] / 2h_n$$

により推定するのが統計的に頑健な方法と考えられる。なおこうした推定方法を漸近的に正当化するには h_n はバンド幅で $h_n \rightarrow 0$ ($n \rightarrow \infty$) とする必要がある。こうした幾つかの数学的仮定の下でエッジワース展開を評価した上で Hall & Sheather (1998) はバンド幅の選択

$$(26) \quad h_n = n^{-1/3} z_\alpha^{2/3} \left[1.5 \frac{s(\tau)}{s''(\tau)} \right]^{1/3}$$

²特に評価関数が凸関数である場合には極小値推定量 (extremum estimator) の収束や漸近分布の導出についてより一般的な証明方法の展開も可能である。

を提案している。ここで α は有意水準、標準正規分布の分布関数 $\Phi(\cdot)$ 、 z_α は $\Phi(z_\alpha) = 1 - \alpha/2$ を満足する。ここで $s(\tau)$ と $s''(\tau)$ についてはデータより

$$(27) \quad \frac{s}{s''} = \frac{f^2}{2(f'/f)^2 + (f'/f - f''/f)}$$

と推定すればよい。さらに Koenker (2005) では正規分布を用いると $s(t)/s''(t)$ が位置-尺度変換について不変であることから、 F として標準正規分布を用いている。以上の理由より本稿で報告する実証分析でもこうした標準的な推定方法を採用した。

2.4 線形計画問題としての分位点回帰法

分位点回帰問題 (6) の推定は線形計画問題として書き表すことができる。例えば $\tau = 0.5$ の場合、つまり最小絶対偏差 (MAD) 回帰については古くから研究されている。例えば Barrodale & Roberts (1973) は最小絶対偏差回帰の性質を利用し、単体法 (simplex method) による最適化問題を効率的に解く方法を提案している。彼らのアルゴリズムはその後、単体法を用いて分位点回帰問題を解く場合にも広く使われることになっている。さらに線形計画問題においては 1980 年代に内点法 (interior point method) が登場したことにより、大規模計画問題に対する計算速度も大幅に改善できるようになった³。分位点回帰モデルにおける推定問題 (6) は線形計画問題

$$P_{qr} \left\{ \begin{array}{l} \min \quad \tau \mathbf{1}'_n \mathbf{u} + (1 - \tau) \mathbf{1}'_n \mathbf{v} \\ \text{subject to } \mathbf{y} - \mathbf{1}_n \beta_0 - \mathbf{X} \boldsymbol{\beta} = \mathbf{u} - \mathbf{v}, \\ \mathbf{u} \geq \mathbf{0}_n, \mathbf{v} \geq \mathbf{0}_n \end{array} \right.$$

として表現できる。ここで $\mathbf{0}_n$ と $\mathbf{1}_n$ はそれぞれ 0 と 1 を n 個並べたベクトルである。主問題では P_{qr} の目的関数を $(\beta_0, \boldsymbol{\beta}', \mathbf{u}', \mathbf{v}')$ と言う $2n + p + 1$ 個の変数について最小化する。ここで $\mathbf{X}^* = [\mathbf{1}_n, \mathbf{X}]$, $\boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}')$ とおく。問題 P_{qr} を単体法により解く場合には、初期実行可能基底解として

$$\boldsymbol{\beta}^*(h) = \mathbf{X}^*(h)^{-1} \mathbf{y}(h), \quad h \in \mathcal{K} \text{ ただし } \mathcal{K} = \{h \mid |\mathbf{X}^*(h)| = p + 1\},$$

$$\mathbf{u}(h) = \mathbf{v}(h) = \mathbf{0}_{p+1}, \quad \mathbf{u}(\bar{h}) = (\mathbf{y} - \mathbf{X}^* \boldsymbol{\beta}^*(h))_+, \quad \mathbf{v}(\bar{h}) = (\mathbf{y} - \mathbf{X}^* \boldsymbol{\beta}^*(h))_-$$

と置くことで問題が解決できる。さらに主問題 P_{qr} の双対問題は

$$D1_{qr} \left\{ \begin{array}{l} \max \quad \mathbf{y}' \mathbf{d} \\ \text{subject to } \mathbf{1}'_n \mathbf{d} = 0, \\ \mathbf{X}' \mathbf{d} = \mathbf{0}_p, \\ \mathbf{d} \in [\tau - 1, \tau]^n \end{array} \right.$$

³こうした問題については例えば Portnoy & Koenker (1997) を参照されたい。

で与えられるが⁴、この表現は Koenker (2005) の定理 2.1 に対応している。ここで変数を定義し直し、 $\mathbf{a} = \mathbf{d} + (1 - \tau)\mathbf{1}_n$ とすれば線形計画問題は

$$D2_{qr} \left\{ \begin{array}{l} \max \quad \mathbf{y}'\mathbf{a} \\ \text{subject to} \quad \mathbf{1}'_n \mathbf{a} = (1 - \tau)n, \\ \quad \quad \quad \mathbf{X}'\mathbf{a} = (1 - \tau)\mathbf{X}'\mathbf{1}_n, \\ \quad \quad \quad \mathbf{a} \in [0, 1]^n \end{array} \right.$$

なる有界変数問題になり、初期値である内点 \mathbf{a}_0 を $(1 - \tau)\mathbf{1}_n$ と選んでやれば、主双対内点法⁵ で扱われている形式となり、数値的最適化の扱いが容易となる。主問題としての定式化では非負条件に加えて制約条件数は標本数 n と等しくなっている。したがって標本数 n が大きくなると⁶計算負荷量が増大する。他方、双対問題では制約条件数は説明変数 $p + 1$ 程度であるので推定に必要な計算量は遙かに小さくなる。

分位点回帰問題を解くアルゴリズムについて単体法ベースの Barrodale-Roberts 法に対して、Frisch-Newton 法として Portnoy & Koenker (1997) で提案されている。分位点回帰問題では $(0, 1)$ 上の任意の τ に対して線形計画問題を解く必要がある。この為に線形計画法において効率的計算方法として知られているパラメトリック線形計画法の利用も考えられる。ここで議論している計画問題ではパラメトリック線形計画法を用いて $(0, 1)$ 上のすべての最適基底解を求めることが可能である。例えば Koenker & Orey (1987) では、 $(0, 1)$ 上のすべての τ に対してパラメトリック線形計画法と分位点回帰問題の最適基底解についての性質 (Koenker (2005) の定理 2.1) を利用し、解 (最適基底解) を連続的に得るアルゴリズムを提案している。計画問題の解が連続的に得られる点ではこの方法が良いと思われるが、 n が大きくなると解が変化する場合の数も多くなることが応用上の問題になってくる。

本稿で扱った問題を検討する過程では説明変数と誤差の両方について、すべて相関の無い標準正規乱数を用いてシミュレーション実験により数値計算の効率性を確かめてみた。例えば $n = 3,000$ 、 $p = 8$ とした場合には解が変化する回数は約 6,000 回だった。また p は変えずに $n = 10,000$ とした場合は約 20,000 回の解の変化が確認された。 n がこの程度の大きさならば計算時間についてはあまり考慮する必要は無いが、例えば $n = 500,000$ 程度になると解が変化する回数とともに計算時間もかなり大きくなると予想される。ここでこのような状況は損害保険会社が実際のデータで分析を行う際には考えられなくもない状況であり、こうしたときには計算は非効率となる。したがって、大きなデータの場合には実用上は τ について一定の幅をとりながら内点法により解を得る方法が良いと云えよう。

⁴線形計画法における主問題と双対問題の定式化や双対定理 (duality theorem) については竹内 (1966)5 章の説明が分かりやすい。主問題 I 「 $\mathbf{x} \geq 0, \mathbf{A}\mathbf{x} = \mathbf{b}$ の下で $z = \mathbf{c}'\mathbf{x}$ の最大化」は双対問題 II 「 $\mathbf{A}'\mathbf{y} \geq \mathbf{c}$ (\mathbf{y} は任意) の条件下で $w = \mathbf{b}'\mathbf{y}$ の最小化」と同等であり、 $\max\{z\} = \min\{w\}$ となる。ただし \mathbf{A} は行列、 $\mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}$ はベクトルをそれぞれ表している。なお z の最大化は $-z$ の最小化、 w の最小化は $-w$ の最大化に対応する。

⁵内点法については小島・土屋・水野・矢部 (2001) で分かりやすく説明されている。単体法が実行可能多面体の頂点を辿りながら最適解を探索することに対し、内点法では多面体の内部を通って最適解を探す。内点法は単体法と同様に初期値 (実行可能内点) を探すことも難しい問題であったが、近年は初期値に実行可能内点を見つけなくても問題を解けることが知られている。

⁶例えば 4 節で報告する実証分析で利用したデータは $n = 6,113$ であるが、実務的な応用まで考慮すると n は遙かに大きくなる。

3 Lasso 分位点回帰

3.1 Lasso

Lasso とは Tibshirani (1996) により提案された統計的アルゴリズムであり、開発の目的は線形回帰分析における予測精度の改善と変数選択であった。Lasso の罰則を加えた線形回帰分析とは、モデルの定数項以外の母数に絶対値 (L_1) ノルムの罰則をつけたものである。そして Lasso 推定量 $\hat{\beta}^{lasso}(\lambda)$ は最小化問題

$$(28) \quad \hat{\beta}^{lasso}(\lambda) = \arg \min \sum_{i=1}^n (y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta})^2 + \lambda \sum_{i=1}^p |\beta_i|, \quad \lambda \geq 0$$

の解として定義される。ここで λ はチューニング・パラメータと呼ばれている。ここでは式 (28) を λ 形式の Lasso 回帰と呼ぶことにする。次に λ 形式に対して t 形式の Lasso 回帰を定義しておこう。 t 形式の Lasso 推定量 $\hat{\beta}^{lasso}(\lambda)(t)$ は最小二乗推定量を制約条件の凸多面体領域に射影するという意味で次の問題の解として定義される。

$$(29) \quad \hat{\beta}^{lasso}(t) = \arg \min \sum_{i=1}^n (y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta})^2 \quad \text{subject to} \quad \sum_{i=1}^p |\beta_i| \leq t, \quad t \geq 0.$$

このような Lasso 問題における λ 形式と t 形式の最小化問題は同値である。こうした Lasso 回帰と類似の統計的方法としては Ridge 回帰が知られている。この Ridge 推定量は最小二乗推定量を球面上に射影する方法であり、Lasso 法と同様に制約条件の下でパラメータの値を徐々に縮小していく意味では同一である。ここで Lasso 回帰と Ridge 回帰の大きな違いは、パラメータの値を数値的に正確にゼロにすることができるか否かという点である。実は Lasso アルゴリズムではそのようなことが可能であり、必然的に変数選択問題とも密接に関係があるのである。ここで比較上の参考として Ridge 回帰問題は

$$(30) \quad \hat{\beta}^{ridge}(\lambda) = \arg \min \sum_{i=1}^n (y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta})^2 \quad \text{subject to} \quad \sum_{i=1}^p \beta_i^2 \leq t, \quad t \geq 0$$

と定式化できることを指摘しておく。Ridge 回帰は元々説明変数行列が線形独立でない場合に線形回帰分析を行う実用的な目的のために開発されたことは興味深い。

3.2 Lasso 罰則を加えた場合の分位点回帰

分位点回帰と Lasso を組み合わせることにより分位点回帰問題において変数選択が可能となる。Lasso の罰則が L_1 ノルムであることから、Lasso 分位点回帰もまた線形計画問題として定式化できることに注意する必要がある。ここで Lasso の罰則を加えた λ 形式の分位点回帰問題は

$$(31) \quad \min_{\beta_0, \boldsymbol{\beta}} \sum_{i=1}^n \rho_\tau(y_i - \beta_0 - \mathbf{x}'_i \boldsymbol{\beta}) + \lambda \sum_{i=1}^p |\beta_i|$$

である。この (31) は線形計画問題

$$P_{\text{lasso}}(\lambda) \left\{ \begin{array}{l} \min \quad \tau \mathbf{1}'_n \mathbf{u}_1 + \mathbf{1}'_p \mathbf{u}_2 + (1 - \tau) \mathbf{1}'_n \mathbf{v}_1 + \mathbf{1}'_p \mathbf{v}_2 \\ \text{subject to} \quad \mathbf{y} - \mathbf{1}_n(\beta_{0+} - \beta_{0-}) - \mathbf{X}(\beta_{+\setminus 0} - \beta_{-\setminus 0}) = \mathbf{u}_1 - \mathbf{v}_1, \\ \quad \quad \quad -\lambda \boldsymbol{\beta} = \mathbf{u}_2 - \mathbf{v}_2, \\ \quad \quad \quad (\boldsymbol{\beta}_0, \boldsymbol{\beta}') \in \mathcal{R}^{p+1}, \mathbf{u}_1 \geq \mathbf{0}_n, \mathbf{v}_1 \geq \mathbf{0}_n, \mathbf{u}_2 \geq \mathbf{0}_p, \mathbf{v}_2 \geq \mathbf{0}_p \end{array} \right.$$

として表現することで最適化を行った。この線形計画問題の主問題に対する双対問題は有界変数問題

$$D_{\text{lasso}}(\lambda) \left\{ \begin{array}{l} \max \quad \mathbf{y}' \mathbf{a}_1 \\ \text{subject to} \quad \mathbf{1}'_n \mathbf{a}_1 = (1 - \tau)n, \\ \quad \quad \quad \mathbf{X}' \mathbf{a}_1 + \lambda \mathbf{a}_2 = (1 - \tau) \mathbf{X}' \mathbf{1}_n + \lambda \mathbf{1}_p, \\ \quad \quad \quad \mathbf{a}_1 \in [0, 1]^n, \mathbf{a}_2 \in [0, 2]^p \end{array} \right.$$

と同一になる。この問題も分位点回帰問題と同様に、主双対内点法のソフトウェアでよく扱われている形式の線形計画問題になっていることがわかる。初期値 \mathbf{a}_{10} 、 \mathbf{a}_{20} が必要な場合もそれぞれ $(1 - \tau) \mathbf{1}_n$ 、 $\mathbf{1}_p$ とすればよい。

4 自動車保険の分析

4.1 自動車保険について

自動車保険では自動車の所有・使用・管理に関連して生ずる損害を填補する目的の保険であり、損害保険会社において中心的な保険である。また自動車保険の担保種目については対人賠償保険・自損事故保険・搭乗者傷害保険・対物賠償保険・車両保険・無保険車傷害保険の6種類があるが、本稿ではその中でも総額の支払い保険金が比較的大きい対物賠償保険について、分位点回帰を用いたデータ分析を行う。ここで対物賠償保険とは、“自動車の所有、使用、管理に起因して他人の財物を滅失、破損または汚損すること(対物事故)により、法律上の賠償責任を負担することによる損害を填補する保険”を意味する。損害保険料率算出機構がまとめた自動車保険統計⁷によると、2006年度における対物賠償保険の支払保険金は約6,800億円であり、対人賠償保険・搭乗者傷害保険・車両保険を抑えて最大であった。強制保険である自賠責保険では対物事故については支払は行われないので、対物賠償保険に関するリスクは損害保険会社にとって重要な分析対象であろう。

4.2 データについて

本稿のデータ分析では $n = 6,113$ 個の自動車保険対物事故について、ある期間を無作為に選び、各曜日について同じ数だけ無作為抽出したデータを用いている。被説明変数としてクレーム額、事故や個人に関する質的データを説明変数として用いた。クレーム額が1

⁷損害保険料率算出機構のHP (<http://www.nliro.or.jp/>) 上で公開されている。

万円以下のデータについては省いたが、これはリスク分析とは直接的に関係のない要因による支払いと見なした為である。利用可能な説明変数としては、運転者の性別や年齢、事故が起こった曜日と時間帯、車種などで、すべてダミー変数である。ここで被説明変数・説明変数についての情報をまとめておこう。

- Pay: クレーム額 (保険請求額)。
- Male: 男性ならば 1、その他は 0。
- Car3: 用途車種が自家用普通乗用車 (3 ナンバー) ならば 1、その他は 0。
- Holiday: 事故が起こった曜日が土日なら 1、その他は 0。
- Midnight: 事故が起こった時間帯が 23 時か 5 時までの間ならば 1、その他は 0。
- Age20s: 運転者年齢が 20 歳代ならば 1、その他は 0。
- Age40s: 運転者年齢が 40 歳代ならば 1、その他は 0。
- Gold: ゴールド免許所持者ならば 1、その他は 0。
- IUnlim: 対物保険金額 (事故が発生した場合に損害保険会社が支払う保険金の限度額) が無制限ならば 1、その他は 0。
- IUplim500: 対物保険金額が 500 万円以下ならば 1、その他は 0。

表 1 と図 1 はそれぞれクレーム額 (Pay) についての記述統計量、ヒストグラムを表し単位は 10 万円である。図より明らかなようにクレーム額の分布は非対称であり、右裾が重い分布になっている。また各説明変数が与えられた場合 (値が 1 となる) の個体数とクレーム額の平均、中央値、最大値、標準偏差を表 2 にまとめておく。例えば 4 行目より深夜ダミー変数の影響は平均的にクレーム額が他よりかなり大きくなっていることや、8 行目より対物保険金額は無制限の人の割合が多いことがわかる。

ここで説明変数ベクトル x を用いた分位点回帰の計測モデルを

(32)

$$Q_{\tau}(\text{Pay}|x) = \beta_0(\tau) + \beta_1(\tau)\text{Male} + \beta_2(\tau)\text{Car3} + \beta_3(\tau)\text{Holiday} + \beta_4(\tau)\text{Midnight} \\ + \beta_5(\tau)\text{Age20s} + \beta_6(\tau)\text{Age40s} + \beta_7(\tau)\text{Gold} + \beta_8(\tau)\text{IUnlim} + \beta_9(\tau)\text{IUplim500}.$$

と表しておこう。統計モデル (32) を用いて主に分位点 $\tau \in (0, 1)$ 全体の傾向、中位点、上側分位点 $\tau \in [0.9, 0.995)$ などについてデータ分析を行った。

すべての説明変数がダミー変数であったので Lasso 分位点回帰を用いる場合には説明変数の基準化を考える必要はないと、考えられよう。したがってチューニング・パラメータ λ に関する推定量の大きさを判断する際の測定単位は同一と見なして良い。本稿の分析では主に $\tau = 0.5, 0.95$ について Lasso 分位点回帰を行った分析結果を説明するが、それぞれの分位点 τ に対し図 12・図 13 が対応する。この図はチューニング・パラメータ λ を横軸にとったときのパラメータ $\beta^{lasso}(\tau, \lambda)$ 推定値のグラフである。それぞれの折れ線に対応する変数のインデックス i を図の λ 軸と β 軸近くに印しており、 λ の幅は 0.25 とした。

表 1: クレーム額 (Pay) の記述統計量。

最小値	標準偏差	1st-Qu	中央値	平均	3rd-Qu	最大値	歪度	尖度
0.101	2.544	0.696	1.361	2.088	2.527	43.58	43.98	4.795

図 1: クレーム額のヒストグラム

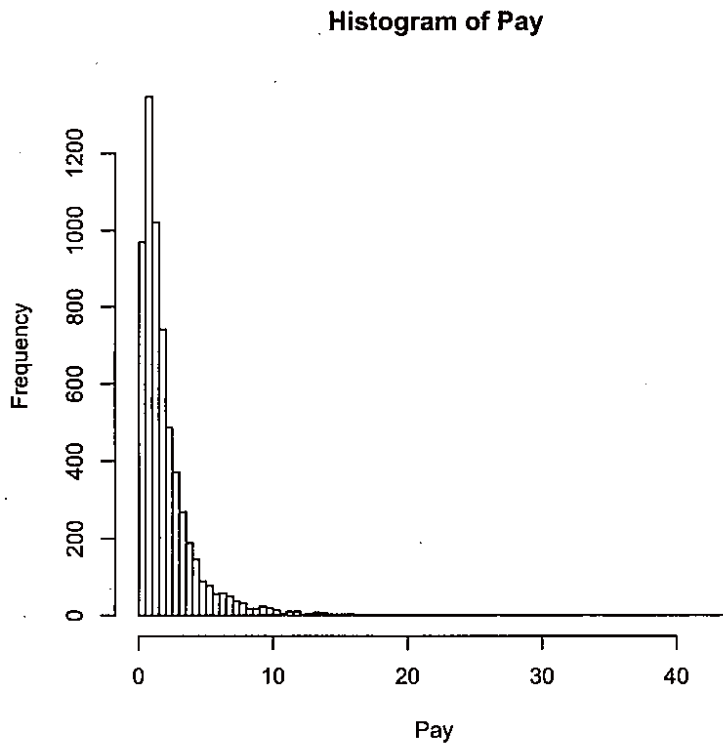


表 2: 説明変数を与えたときのクレーム額の記述統計量など。

変数名	件数 (固体数)	平均	中央値	最大値	標準偏差
Male	3252	2.250	1.449	43.58	2.764
Car3	1714	2.343	1.479	43.58	3.059
Holiday	1758	2.258	1.468	28.35	2.580
Midnight	328	3.216	1.872	34.40	4.194
Age20s	1282	2.257	1.546	27.13	2.567
Age40s	1063	1.876	1.247	18.42	2.017
Gold	2568	1.958	1.286	33.24	2.335
IUnlim	5367	2.071	1.352	43.58	2.526
IUplim500	86	2.477	1.514	19.84	3.224

4.3 実証結果

分位点回帰の推定結果を見ておこう。表3は最小二乗法による回帰分析の結果と $\tau = 0.05, 0.1, 0.5, 0.9, 0.95$ についての分位点回帰の結果であり、数字の上側がパラメータの推定値でその下の括弧内の数字が標準誤差である。図2-図12では、横軸が τ で縦軸が推定値である。図の中の横に横断する灰色の直線が最小二乗推定量でその上下にある同色の点線が90%信頼区間を表しており、黒い点を直線で結んだ線は分位点回帰推定値、周りの灰色の領域がその90%信頼区間である。これらの推定結果を踏まえて通常の線形回帰分析についての結果を見ると、運転者が男性(Male)、車種が3ナンバー車(Car3)、土日(Holiday)、深夜(Midnight)、40歳代(Age40s)、ゴールド免許所持(Gold)については最小二乗推定量は90%以上の有意性を示しているが、他の説明変数については有意と云えないことがわかる。以降「有意性を示す」とは90%有意のことを意味する。

まず性別について観察してみよう。一般に女性に比べ男性の方がクレーム額が大きくなると云われている。我々の分析結果を見ても、男性については全体的に推定値が正で有意になっていることから、男性効果が確認できた。最小二乗法による推定でも有意、上側の分位点回帰モデルでも有意であり、他の説明変数に比べてリスクとしては大きくなる。

車種についてであるが、自家用普通乗用車については分位点回帰推定量が男性の場合と同じようなパスを描いている。上側については男性の場合程顕著な有意性を示さないが、それでも十分にリスク要因といえるだろう。

平日と土日については、運転の目的が大分変わってくる。結果を見ると土日についてはクレーム額の分布が全体的に右にずれていることが確認できる。上側でもほぼすべての分位点について推定値は正で有意となっている。したがって土日に起こる事故が平日のそれよりクレーム額は大きくなることが予想される。

時間帯効果について深夜効果のみていく。深夜効果についての最小二乗推定量はすべての説明変数の中で最も有意な値になっている。分位点回帰でも深夜の事故のクレーム額の分布は、そうでない場合の事故に比べて全体が大きく右にずれている。したがってこの説明変数は最も大きなリスク要因だと考えられる。

年齢について20歳代と40歳代を見てみよう。20歳代はクレーム額が大きくなる傾向がある。最小二乗法では有意性を示さなかったが、分位点回帰の結果を見れば20歳代のクレーム額分布は下側から中央、そして $\tau = 0.8$ 付近にかけて正で有意となっていることが分かった。しかし、上側をみると有意性は示さないものの負の値をとっている。このことから、そのクレーム額分布は平均的に右よりだが、右裾は厚くないといえるであろう。次に40歳代を見てみよう。30歳以上の運転者についてはしばしば保険金額が安くなる傾向がある。我々の分析結果を見てみると最小二乗法では40歳代という要因は負で有意性を示している。分位点回帰によれば $\tau = 0.8$ 付近で少し有意性が確かめられるものの、特別にリスクが小さくなるとは云い難い。

ゴールド免許所持者に関しては τ について全体的に推定値が負に有意になっている。上位分位点においても観察されているので、ゴールド免許所持者についてのクレーム額はやはり小さくなると云えるであろう。

対物支払保険については、対物支払保険金額を無制限にしている人は事故に対するリス

クを大きく見積もっているのではと考えられる。最小二乗法では有意性が確かめられなかったが、分位点回帰においては上側、特に $\tau \in (0.9, 0.95)$ あたりで推定値が負で有意になっている。逆に対物保険金額を 500 万円としている人については上側の $\tau \in (0.94, 0.985)$ 付近で正で有意となっている。つまり他の説明変数を固定した場合の対物保険金額を 500 万円としている人のクレーム額の分布は、そうでない人に比べ右裾が重くなっていることが推測される。すなわち事故についてのリスクを小さく見積もる人は、高額を支払を要求しているのではないかと推測できよう。

以上で説明したように分位点回帰を用いるとこれまでの線形回帰分析のみを用いた場合に比べ、リスク要因についてのより詳しい結果が得られる。

さらに次に Lasso 分位点回帰により説明変数の選択を行った簡単な結果を紹介しておこう⁸。まず $\tau = 0.95$ (図 13) について推定結果を見ておこう。 $\lambda = 40$ 付近で男性が最初に正の値として推定された。つまりリスク要因として最初に現れたと言えよう。徐々に λ を小さくしていくと、次に深夜がリスク要因として検出され、さらに λ を小さくしていくと各推定値の絶対値はより大きくなる。その後、自家用普通乗用車、ゴールド免許所持者、休日、対物保険金額が無制限の順で推定値がゼロでなくなった。それぞれの推定値の符号は正、負、正、負となっており、特にゴールド免許所持者と対物保険金額が無制限の被保険者はリスクが小さくなると云えそうである。年齢については図を見る限り、そのリスク要因としての重要性はそれほど大きくない。また対物保険金額が 500 万円以下の被保険者については推定値の値が大きいことと、ゼロでなくなる最初の λ が他に比べ小さいことをみれば、より慎重な議論が必要そうである。以上の議論から判断すれば、特に男性変数と深夜変数はリスク要因として重要であると云えるであろう。

続いて中央値 ($\tau = 0.5$) についての結果 (図 12) にも簡単なコメントを加えておこう。最初にリスク要因として検出されるのは、上側の場合と同様に男性であった。次に 20 歳代、休日、自家用普通自動車推定値が正で検出された。上側に比べ、結果が大分違ってくるのがわかる。特に 20 歳代が上側の分位点に比べ中央値において、変数の重要性が増していると云えるであろう。また、ここでは対物保険金額についての変数は重要性が小さいように思われる。

以上、簡単に Lasso 分位点回帰の結果を図を参照してみたが、客観的にリスク要因の抽出として変数選択を行うにはチューニング・パラメータの選択や漸近理論などの考察が重要になってくるが、なお解決すべき統計的課題も少なくない。

5 おわりに

本稿では分位点回帰と Lasso 分位点回帰の理論を説明すると共に、実際に自動車対物賠償保険のクレーム額に対して応用した結果を報告した。特に被説明変数であるクレーム額の

⁸図 12・図 13 では数値は λ を動かしたときの各変数の係数推定値の変化を示している。 $\lambda = 0$ の時は制約のない場合の推定値を示し、 λ の値が大きくなるにつれ (制約が厳しくなり) 推定値の絶対値はより小さくなる。

表 3: 分位点回帰の結果

Covariates	0.05	0.10	0.50	0.90	0.95	LS
(Intercept)	0.231 (0.026)	0.312 (0.034)	1.195 (0.071)	4.207 (0.280)	6.076 (0.446)	1.907 (0.116)
Male	0.015 (0.015)	0.017 (0.020)	0.167 (0.040)	0.757 (0.160)	1.116 (0.256)	0.066 (0.271)
Car3	0.026 (0.016)	0.053 (0.022)	0.181 (0.044)	0.564 (0.176)	0.632 (0.280)	0.287 (0.072)
Holiday	0.035 (0.016)	0.081 (0.021)	0.150 (0.044)	0.451 (0.172)	0.797 (0.275)	0.202 (0.071)
Midnight	0.045 (0.032)	0.097 (0.043)	0.388 (0.088)	2.893 (0.350)	3.579 (0.558)	1.076 (0.145)
Age20s	0.025 (0.019)	0.039 (0.025)	0.179 (0.052)	-0.220 (0.206)	-0.365 (0.328)	0.038 (0.085)
Age40s	-0.013 (0.020)	-0.031 (0.026)	-0.078 (0.054)	-0.277 (0.213)	-0.057 (0.339)	-0.197 (0.088)
Gold	0.004 (0.015)	0.010 (0.020)	-0.089 (0.042)	-0.267 (0.165)	-0.664 (0.262)	-0.150 (0.068)
IUnlim	-0.022 (0.023)	-0.001 (0.031)	-0.022 (0.063)	-0.544 (0.251)	-0.722 (0.400)	-0.095 (0.104)
IUplim500	0.000 (0.065)	0.001 (0.086)	0.127 (0.176)	0.106 (0.698)	2.804 (1.114)	0.384 (0.289)

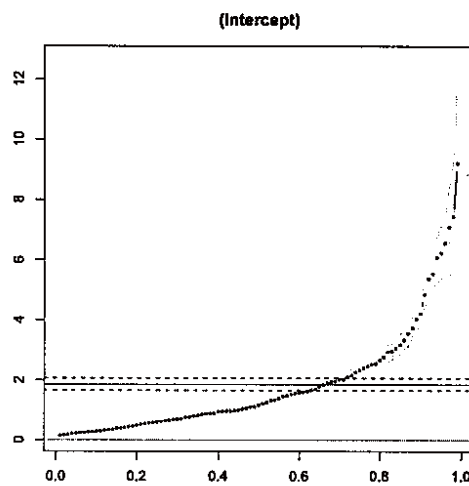


図 2: 切片

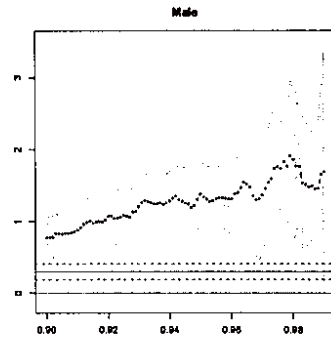
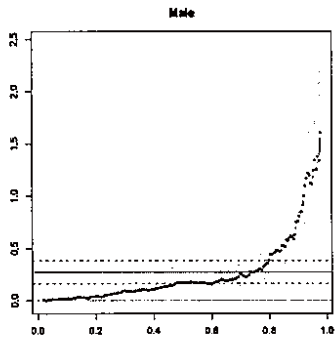


図 3: 男性

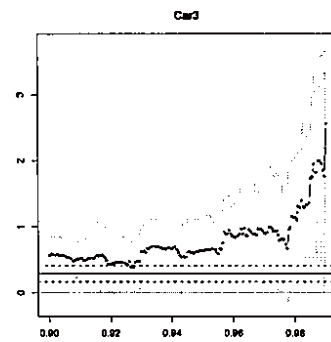
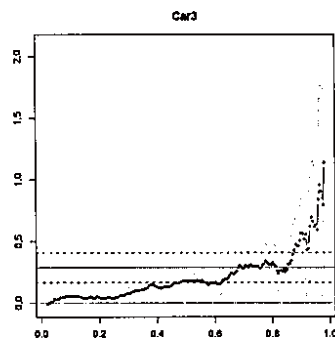


図 4: 自家用普通乗用車 (3ナンバー)

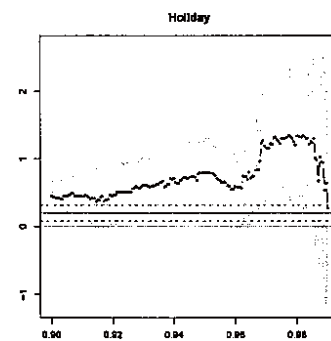
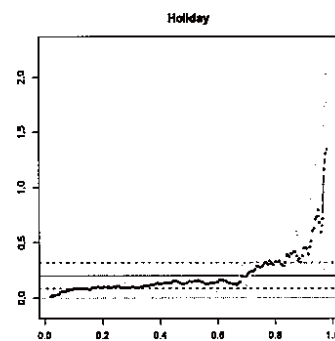


図 5: 土日

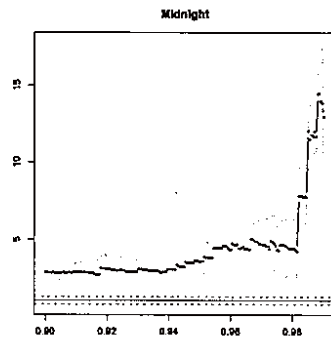
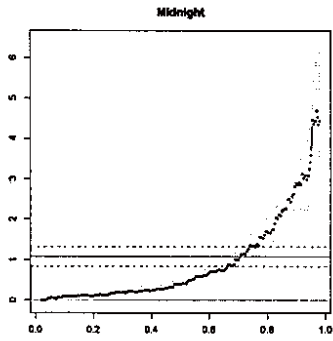


図 6: 深夜

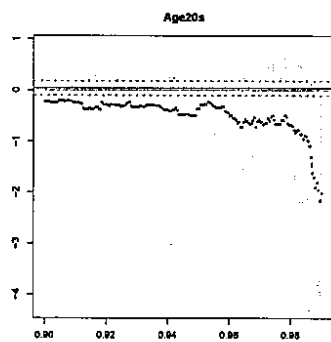
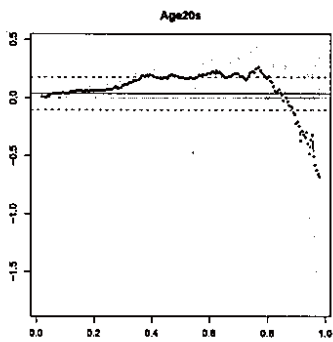


図 7: 年齢が 20 歳代

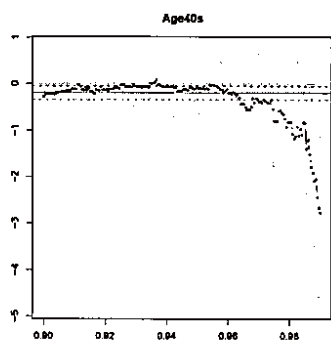
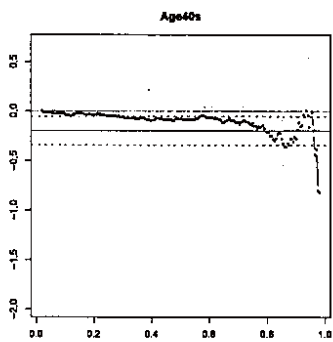


図 8: 年齢が 40 歳代

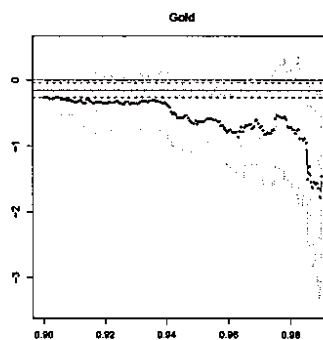
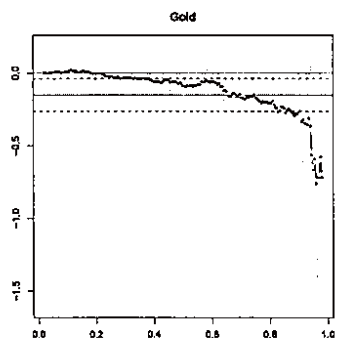


図 9: ゴールド免許所持者

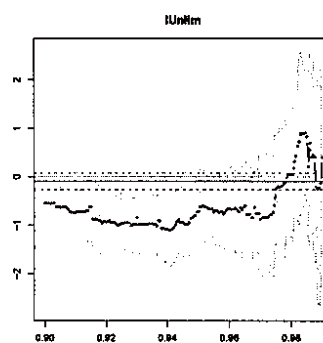
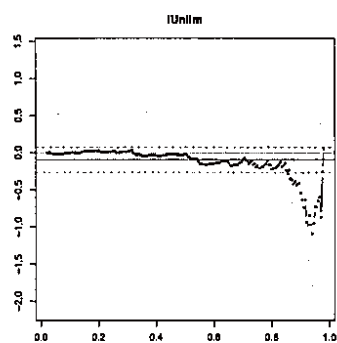


図 10: 対物保険金額が無制限

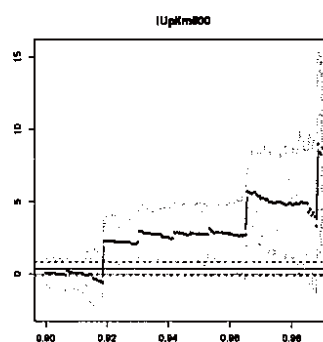
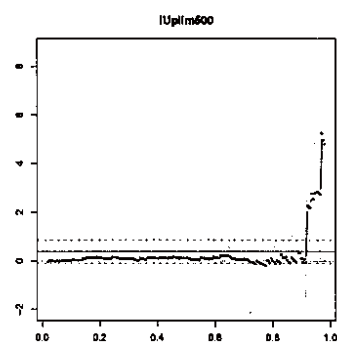


図 11: 対物保険金額が 500 万円以下

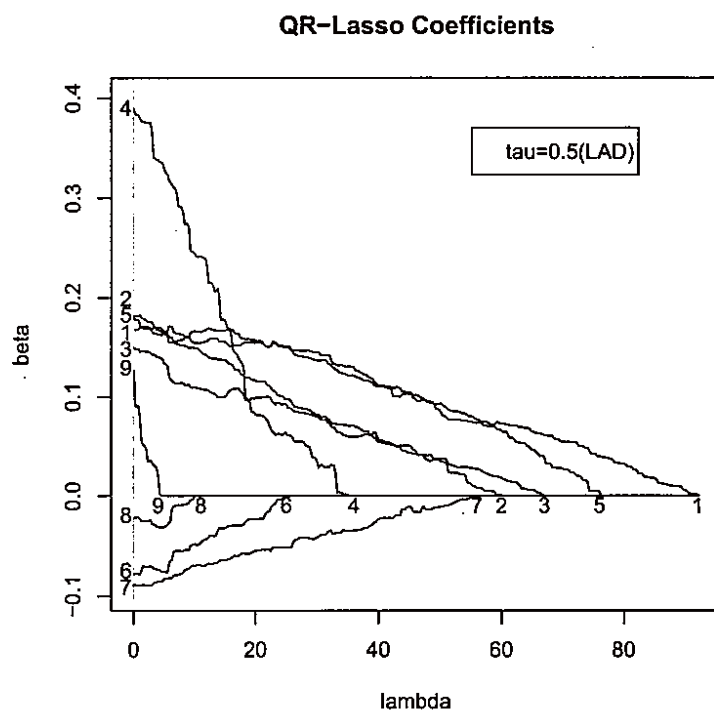


図 12: Lasso 分位点回帰の結果 ($\tau = 0.50$)

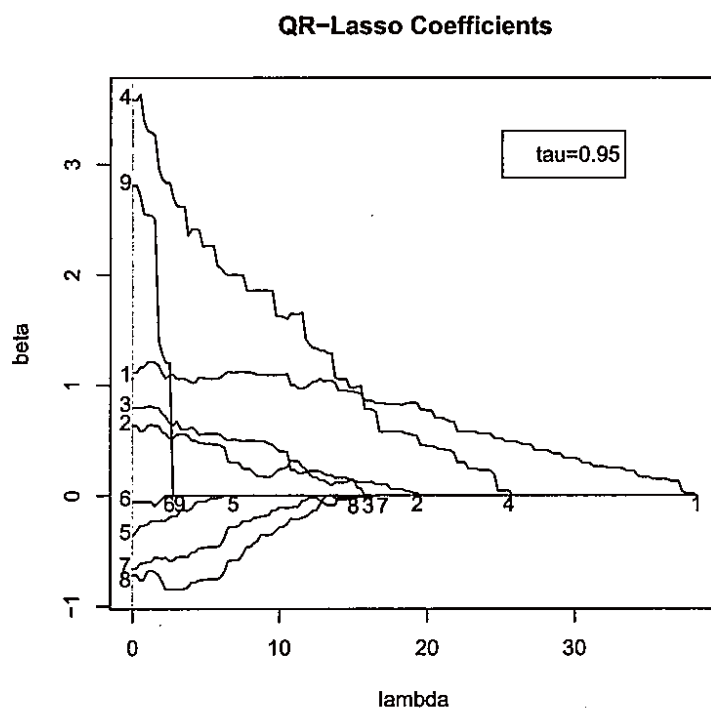


図 13: Lasso 分位点回帰の結果 ($\tau = 0.95$)

高分位点のリスク要因について興味ある結果が得られたと言う意味では、通常の線形回帰モデルよりも分位点回帰モデルの方が保険リスク分析に適しているという結論が得られた。

ところで分位点回帰は本稿で議論した自動車保険をはじめとする損害保険に限らず、生命保険や第三分野保険などの保険分野、あるいはより広い金融分野で利用可能である。近年では金融機関におけるリスク管理においても標準的なリスク指標として VaR(Value-at-Risk)がある。VaR は金融資産の収益率分布の下側分位点であり、こうした分位点回帰を用いた一つの応用が Engle & Manganelli (2004) によって報告されている。こうした、金融機関におけるリスク管理問題などにおいては分位点回帰法の有用性が期待される。また損害保険の分野においては近年、リスク細分型保険といわれる保険契約も登場している。例えばリスク細分型自動車保険では運転者年齢や地域、走行距離、目的などによりリスク要因を細かくして保険料を定めていると思われる。こうしたリスク要因のクレーム頻度との関係の分析を行えば、分位点回帰はさらに有用な統計的分析法になるとと思われる。

最後になるが、本稿で取りあげた分位点回帰や Lasso 分位点回帰については更に検討すべき様々な理論的問題や計算上の問題があることを指摘しておく。例えば理論面では有限標本において漸近的議論がどれほど有効であるかはまだよく分かっていない。あるいはデータ数や変数の数が極めて大きい場合の説明変数の選択などの議論も近年の応用上では重要であろう。

参考文献

- [1] Amemiya, T. (1985), *Advanced Econometrics*, Blackwell.
- [2] Barrodale, I. and F. Roberts (1973), "An Improved Algorithm for Discrete l_1 Linear Approximation," *SIAM Journal of Numerical Analysis*, 10, 839-848.
- [3] Engle, R. and S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics*, 22, 367-381.
- [4] Hall, P. and S. Sheather (1988),"On the Distribution of a Studentized Quantile," *Journal of the Royal Statistical Society, Series B*, 50, 381-391.
- [5] Knight, K. (1998),"Limiting Distributions for L_1 Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755-770.
- [6] Koenker, R. (2005), "Quantile Regression," *Cambridge University Press*.
- [7] Koenker, R. and G. Bassett (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.
- [8] Koenker, R. and V. d'Orey (1987), "Computing Regression Quantiles," *Applied Statistics*, 36, 383-393.

- [9] 小島政和・土屋隆・水野眞治・矢部博 (2001), "内点法", 朝倉書店.
- [10] 今野浩 (1987), "線形計画法," 日科技連.
- [11] Magnanti, T. and J. Orlin (1988), "Parametric Linear Programming and Anti-Cycling Pivoting Rules," *Mathematical Programming*, 41, 317-325.
- [12] Portnoy, S. and R. Koenker (1997), "The Gaussian hare and the Laplacian tortoise: computability of squared-error vs absolute error estimators," *Statistical Science*, 12, 279-300.
- [13] Powell, J. L. (1991), "Estimation of Monotonic Regression Models under Quantile Restrictions," in *Nonparametric and Semiparametric Models in Econometrics* (ed. by W. Barnett, J. Powell, and G. Tauchen), Cambridge Cambridge University Press.
- [14] Siddiqui, M. (1960), "Distribution of Quantiles from a Bivariate Population," *Journal of Research of the National Bureau of Standards*, 64, 145-150.
- [15] 竹内啓 (1966), "線形数学," 培風館.
- [16] Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267-288.

極値理論とその応用

渋谷政昭 (慶應義塾大学, 理工学部, 名誉教授)

sibuyam@1986.jukuin.keio.ac.jp

東京大学経済学部 June 2007 (12-20)

概要

McNeil, Frey and Embrechts (2005) は, 金融・保険業におけるリスクを market risk, credit risk, operational risk に分類して議論し, Quantitative Risk Management を新しい数理科学的原理として構成することを提案し試みている. その中で, 極値理論の役割が重視され, 10 章のうちの 1 章を割り当てている. この章を理解することがよい入門であろうと予想して, この講義の目標とする.

極値理論の目的は分布の裾で生じる事象についての推測である. 通常のパラメトリックな推測ではロバストな結果が得られない. ノンパラメトリックな推測では十分な精度が得られない. 歴史的には洪水・暴風・高波などの自然災害, 機械・電気製品の信頼性の分野で必要とされ, 発展してきた. 今日では地球環境の現実課題に挑戦するための道具を開発しなければならない. これらにも触れたい.

目次

1	極値理論入門	4
1.1	極値理論 (まえがき)	4
1.2	区分最大値 (古典理論)	4
1.2.1	最大値安定性	5
1.2.2	最大値吸引領域	6
1.2.3	最大吸引領域の諸例	7
1.2.4	極値分布と一般極値分布, von Mises-Jenkinson 形	8
1.3	水準超過観測値 (POT)	9
1.3.1	一般 Pareto 分布の性質	10
1.4	他の枠組み	12
1.4.1	順序統計量	12
1.4.2	最大 r 個の同時極限分布	13
1.4.3	関連の話題, 新記録 (records 数と値)	14
2	推測	15
2.1	統計的補外	15
2.2	一般極値分布の推定	16
2.2.1	ブートストラップ法	17
2.2.2	ベイズ推定	17
2.3	一般 Pareto 分布の当てはめ	17
2.4	参考書とソフトウェア	19

3	応用	20
3.1	Delta 計画	21
3.2	MV Derbyshire 号遭難事件	21
3.3	Venezuela 大洪水	21
3.4	損害保険への応用	21
3.5	Extreme Returns in Asset Prices	22
3.6	金属疲労と “Wicksell’s corpuscle problem”	22
3.7	“超高齢者” の寿命と表データ	23
4	基礎理論	24
4.1	最大値分布収束の図示	24
4.2	一般極値分布の吸引領域に関する十分条件	24
4.3	再現水準と裾確率点関数	25
4.3.1	再現水準と裾確率点関数	25
4.3.2	裾確率点関数の収束, 最大吸引領域	26
4.4	Hill 推定量	27
4.5	点過程モデル	28
4.5.1	水準超過と Poisson 過程	28
4.5.2	自励型点過程	30
4.6	補い	31
4.6.1	広義逆関数と確率点関数	31
4.6.2	正則変動関数	31
4.6.3	最大値安定性より 3 分布の導出の概略	32
5	多変量極値理論	33
5.1	多変量極値 (まえがき)	33
5.2	多変量極値分布 (区分最大値)	33
5.3	極値接合分布関数	33
5.3.1	準備: 接合分布関数 (copula)	33
5.3.2	極値接合関数	37
5.3.3	極値接合関数と単純極値分布との関係	37
5.4	単純多変量極値分布	38
5.4.1	定理の説明	39
5.4.2	指数測度 (平均測度)	39
5.4.3	関連定理	40
5.4.4	独立性と従属性	40
6	2 変量極値分布	41
6.1	2 変量単純極値分布, 従属性関数	41
6.2	2 変量極値分布の例	43
6.3	推測	44
6.4	裾の漸近的独立性・従属性	45
6.5	Sibuya の従属性関数	45
7	多変量極値モデル	47
7.1	種々の標準化 (周辺分布の選択)	47
7.1.1	指数周辺分布	47
7.2	単純多変量極値分布の例	47
7.2.1	非対称ロジスティックモデル	49
7.2.2	h の別表現	49
7.3	スペクトル表現	50

7.4	ノンパラメトリック推定	51
8	移動最大値過程 (強定常時系列)	52
8.1	強定常時系列	52
8.2	極値時系列 (M3/M4 process)	53
8.3	A.I/D.M4 process	54
8.4	MCM4	55
9	応用	55
10	多変量極値分布の新定義	56
11	付録	59
11.1	多変量正規変動関数	59
11.2	超単純多変量極値分布	60
11.2.1	2変量独立 Fréchet 分布	60
11.2.2	超単純多変量極値分布	62
11.2.3	多変量指数分布	65

1 極値理論入門

1.1 極値理論 (まえがき)

統計学は伝統的にばらつきの中央部分に関心を持っている。中央を規範とみなす心情、過半数の利益を重視する社会構成、などを背景とし、“中心的な” 中心極限定理が理論的基礎を与えた。この流れでは、たとえば外れ値を“ごみ” のように扱う風習が生ずる。しかし統計学の流れの中で少数が無視されていたわけではない。ポアソン分布を裏付ける“少数の法則”を代表として、稀な事象への関心があった。事故、故障、自然災害などの解析である。

1950-60 年代に E. J. Gumbel が“極値統計学 Statistics of Extremes (1958)” の応用に努力し、気象学、水文学、諸工学 (材料、腐食、土木、建築、信頼性) などに普及した。1980 年代より国際的フォーラムが開かれるようになったが統計学の中では少数派であり、“extremist” と自称する人もいる。雑誌 Extremes が 1998 年に創刊された。最近では環境の解析に利用されている。もっとも弱く貧しい人が、もっとも強い災害・汚染により被害を受けている。そして投機推奨時代では最大利益を追う人が利用を始めている。

推測としての極値理論は、“確率分布の右裾への統計的な補外” であり、極値分布に基づくセミパラメトリックなモデル構築である。製品の寿命、漏水などでは最小値が問題になるが、 $\min(x_1, \dots, x_n) = -\max(-x_1, \dots, -x_n)$ であるから、習慣として最大値だけを議論する。正の値だけ扱うのであれば $\min(x_1, \dots, x_n) = 1/\max(1/x_1, \dots, 1/x_n)$ と変換してもよい。

極値理論は利用可能なデータにより、大きく二つの枠組みに分かれる。区分最大値 (block maxima) データと、水準超過値 (threshold exceedances) データである。前者は気象・自然災害における年間最大値に関心があったために始まった枠組みで、極値分布を中心とする“古典理論” である。利用できるデータは乏しい。信頼できる年間最大値が 100 年以上継続しているデータは稀少である。後者は利用する観測数を増やすという当然の要求から生じた枠組みで、ある基準値を設けてこれを超える観測値を全て利用する方法である。水文工学の分野で“Peaks Over Threshold, POT” という名前で始まった。2 種類の型のデータの関係が明らかになり、極値分布に対応する一般 Pareto 分布がより扱いやすいために、新しい枠組みとして関心が移った。“一般極値分布” と“一般 Pareto 分布” が、極値理論入門のキーワードである。

1.2 区分最大値 (古典理論)

$(X_n)_{n=1}^{\infty}$ を i.i.d. 確率変数列、その分布関数を $F(x) = P\{X_n \leq x\}$ とする。標本最大値 $Z_n = \max(X_1, \dots, X_n)$ の分布関数 $F^n(x)$ の漸近分布を求める。 F に依存する適当な数列 $(a_n > 0)_{n=1}^{\infty}, (b_n)_{n=1}^{\infty}$ と非退化分布関数 $H(x)$ にたいして、 $(Z_n - b_n)/a_n$ の確率分布が

$$F^n(a_n x + b_n) \xrightarrow{d} H(x), \quad n \rightarrow \infty,$$

を満たすとき、“ F は H の最大値吸引領域 (maximum domain of attraction) に属する” と言い、 $F \in MDA(H)$ と記す。 (a_n, b_n) は F の“吸引係数 (attraction coefficients)”, あるいは“正規化係数 (norming constants)” と呼ばれている。

F がある吸引係数 (a_n, b_n) にたいして $F \in MDA(H)$ であれば、吸引係数を変えても H の位置、尺度を変えたもの (同じ型の分布) にしか収束しない。実際

$$\begin{cases} F^n(a_n x + b_n) \xrightarrow{d} H(x), & \text{and} \\ F^n(\alpha_n x + \beta_n) \xrightarrow{d} H^*(x), \end{cases}$$
$$\iff \alpha_n/a_n \rightarrow a, \quad (\beta_n - b_n)/a_n \rightarrow b, \quad n \rightarrow \infty, \quad H^*(x) = H(ax + b).$$

したがって $H(x)$ を標準化しておけば、吸引係数として許される範囲は、

$$H^*(x) = H(x) \iff \alpha_n/a_n \rightarrow 1, \quad (\beta_n - b_n)/a_n \rightarrow 0, \quad n \rightarrow \infty.$$

Extremal Types Theorem

このような最大値の極限分布となるものは次の 3 つの型 (位置、尺度をパラメータとする族) に限られる。

これらの分布を 極値分布 (Extreme Value Distributions) と呼ぶ. 最大値極限分布と呼ぶのが適当であろうが, 最大値, 最小値を合わせて極値と呼び, 極限の意味も含ませている.

“Gumbel 分布” Λ

$$H_0(x) = \exp(-\exp(-x)), \quad -\infty < x < \infty. \quad (1)$$

“Fréchet 分布” Φ_α

$$H_{1\alpha}(x) = \begin{cases} \exp(-x^{-\alpha}), & x > 0, \\ 0, & x \leq 0, \alpha > 0. \end{cases} \quad (2)$$

“負の Weibull 分布” Ψ_α

$$H_{2\alpha}(x) = \begin{cases} 1, & x > 0, \\ \exp(-(-x)^\alpha), & x \leq 0, \alpha > 0. \end{cases} \quad (3)$$

分布関数の表現法にはいくつかの習慣がある. たとえば上記の $\Lambda, \Phi_\alpha, \Psi_\alpha$. ここでも後で別の表現法を使う. 分布範囲の違いに注意すること. 信頼性理論では製品寿命の最小値に関心があり, Weibull 分布 $1 - H_{2\alpha}(-x), x > 0$, がもっとも人気のある分布である. これと区別するために, (3) を ‘負の Weibull 分布’, あるいは ‘極値 Weibull 分布’ と呼ぶ.

問 1. 3 種の極値分布族に従う乱数を生成するプログラムを書け.

問 2. (i) 確率変数 Y が H_{11} (または H_{21}) に従うとき, $Y^{1/\alpha}$ (または $-(-Y)^{1/\alpha}$) は $H_{1\alpha}$ (または $H_{2\alpha}$) に従う. この結果をより一般的にせよ.

(ii) 次の表は, 指数確率変数と 3 つの分布 H_0, H_{11}, H_{21} に従う確率変数の間の関数関係である. これを確かめよ.

表 1: $Y = \varphi(X)$

	$X \setminus Y$	$1 - \exp(-x)$	H_0	H_{11}	H_{21}
exponential	$1 - \exp(-x)$	-	$-\log X$	$1/X$	$-X$
Gumbel	H_0	$\exp(-X)$	-	$\exp X$	$-\exp(-X)$
Fréchet	H_{11}	$1/X$	$\log X$	-	$-1/X$
neg. Weibull	H_{21}	$-X$	$-\log(-X)$	$-1/X$	-

1.2.1 最大値安定性

3 種分布に限られることの説明: (X_1, \dots, X_{mn}) を前節の条件を満たす, 大きさ mn の標本とする. これを m 組の, 大きさ n の標本に分け, 各組の最大値を

$$W_j := \max(X_{(j-1)n+1}, \dots, X_{jn}), \quad j = 1, \dots, m,$$

とすると $(W_j - b_n)/a_n$ の分布関数は

$$F^n(a_n w + b_n) \rightarrow H(w), \quad n \rightarrow \infty,$$

を満たす.

$$W = \max(W_1, \dots, W_m) = \max(X_1, \dots, X_{mn})$$

であり $(W - b_{mn})/a_{mn}$ の分布関数が

$$F^{mn}(a_{mn} w + b_{mn}) \rightarrow H(w), \quad n \rightarrow \infty$$

を満たすから, 適当な $(a_m^*, b_m^*)_{m=1}^\infty$ にたいして

$$H^m(a_m^* w + b_m^*) = H(w)$$

を満たさなければならない. これが **最大値安定 (max-stability)** の条件である. この “関数方程式” を解いて 3 分布を導く. Fisher and Tippett, Gnedenko, たちの仕事である. 補い 4.6.3 参照.

$H_0, H_{1\alpha}, H_{2\alpha}$ は, 次のように最大値安定である.

$$H_0^n(x) = H_0(x - \log n), H_{1\alpha}^n(x) = H_{1\alpha}(n^{-1/\alpha} x), H_{2\alpha}^n(x) = H_{2\alpha}(n^{1/\alpha} x), \quad \forall n \in \mathbb{Z}_{>0}, \forall x \in \mathbb{R}.$$

問 1. 最大値安定性よりも歴史が古いのが, 和に関する安定性である. 両者の間の理論的な相似性を調べよ. たとえば Embrechts, Klüppelberg and Mikosch (1997).

1.2.2 最大値吸引領域

ある分布関数が極値分布の最大値吸引領域に属する, つまり $F \in MDA(H)$, ($H = H_0, H_{1\alpha}, H_{2\alpha}$), であるための十分条件として, ハザード関数 (災厄関数) を用いるもの (von Mises 1936) が理解しやすいであろう.

定理 1 (Fréchet 分布). F がすべての $x > x_0$ にたいして正の密度関数 f をもち, ある $\alpha > 0$ にたいして

$$\lim_{x \rightarrow \infty} \frac{x f(x)}{1 - F(x)} = \alpha,$$

ならば $F \in MDA(H_{1\alpha})$ である. ($b_n = 0, 1 - F(a_n) = 1/n$, とすればよい.)

次の命題が成り立ち, ハザード関数 f/\bar{F} は $O(1/x)$, $x \rightarrow \infty$, より早く 0 に近づくことはできない. 生物・製品の寿命分布のハザード関数は単調に増加する. 定理の条件はこのような意味で裾が重い (厚い) 特殊な分布を指定する.

命題 $h(x)$ を任意のハザード関数とすると ‘ $v < x < \infty$ で $h(x)x^{1+\varepsilon}$ が単調減少する’ ような $v, \varepsilon > 0$ は存在しない.

典型例 Pareto 分布

$$\begin{aligned} \bar{F}(x) &= x^{-\alpha}, \quad x > 1; \quad f(x) = \alpha x^{-\alpha-1}; \quad h(x) = \alpha/x, \quad xh(x) = \alpha, \\ f'(x) &= -\alpha(\alpha+1)x^{-\alpha-2}, \quad f'(x)/f(x) = -(\alpha+1)/x, \quad \frac{f'}{f} \Big/ \frac{f}{\bar{F}} = -1 - \frac{1}{\alpha}. \end{aligned}$$

定理 2 (Weibull 分布). F の密度関数 f がある区間 (x_0, x_1) で正, $x > x_1$ で 0 であり, ある $\alpha > 0$ にたいして

$$\lim_{x \uparrow x_1} \frac{(x_1 - x)f(x)}{1 - F(x)} = \alpha,$$

ならば $F \in MDA(H_{2\alpha})$ である. ($b_n = x_1, 1 - F(x_1 - a_n) = 1/n$, とすればよい.)

定理 1 で注意したことは逆に, ハザード関数 f/\bar{F} が $1/(x_1 - x)$ より遅く発散することはない. これより早く発散すると定理 3 の場合になる.

典型例 ベータ分布 $\text{Be}(1, \alpha)$

$$\begin{aligned} \bar{F}(x) &= (1-x)^\alpha; \quad f(x) = \alpha(1-x)^{\alpha-1}; \quad h(x) = \alpha/(1-x), \quad (x_1 - x)h(x) = \alpha, \quad x_0 = 0, x_1 = 1, \\ f'(x) &= -\alpha(\alpha-1)(1-x)^{\alpha-2}; \quad f'(x)/f(x) = -(\alpha-1)/(1-x), \quad \frac{f'}{f} \Big/ \frac{f}{\bar{F}} = -\frac{\alpha-1}{\alpha} = -1 + \frac{1}{\alpha}. \end{aligned}$$

定理 3 (Gumbel 分布). F の密度関数 f がある区間 (x_0, x_1) , $x_1 \leq \infty$, で微分可能 $f'(x) < 0$ であり, $x_1 < \infty$ ならば $f(x) = 0$, $x > x_1$ とする.

$$\lim_{x \uparrow x_1} \frac{f'(x)(1-F(x))}{(f(x))^2} = -1,$$

ならば $F \in MDA(H_0)$ である. ($1-F(b_n) = 1/n$, $a_n = \varphi(b_n)$, $\varphi(t) \sim (1-F(t))/f(t)$, $t \rightarrow x_1$, とすればよい.)

定理の条件から

$$\frac{f(x)}{1-F(x)} \sim \frac{-f'(x)}{f(x)}, \quad (x \uparrow x_1),$$

であるが, 多くの分布がこの条件を満たす. $x_1 = \infty$ の場合も $x_1 < \infty$ の場合もある. そのため議論がより複雑となる.

典型例 $x_1 = \infty$, 指数分布: $1-F(x) = f(x) = -f'(x) = e^{-x}$.

典型例 $x_1 < \infty$, 負の Fréchet 分布. Y が指数分布に従うとき $X = -1/Y$ の確率分布.

$$\begin{aligned} \bar{F}(x) &= \exp(1/x), \quad x < 0; \quad f(x) = x^{-2} \exp(1/x); \quad h(x) = 1/x^2; \\ f'(x) &= \left(-\frac{2}{x^3} - \frac{1}{x^4}\right) \exp\left(\frac{1}{x}\right); \quad \frac{f'}{f} = -\frac{1+2x}{x^2}, \\ \frac{f'}{f} / \frac{f}{\bar{F}} &= -(1+2x) \rightarrow -1, \quad x \uparrow 0. \end{aligned}$$

注意 上記 3 定理をまとめた十分条件を第 4 節で述べ証明する.

問 1. ハザード関数を与えれば, 分布関数, 確率密度関数が定まる.

別紙図一般極値分布参照.

1.2.3 最大吸引領域の諸例

以下の G_γ は次の小節で導入する一般極値分布の記号法である.

McNeil et al. (2005) は, 裾が十分に重くても Gumbel 分布の吸引領域に属するものがあることから, ファイナンスでの極値理論が Fréchet 分布 だけではすまないと注意している.

1. $MDA(H_{1\alpha})$, G_γ , $\gamma = 1/\alpha > 0$. Fréchet 分布

- Pareto $x^{-\alpha}$, $x > 1$; $\alpha > 0$.
- Burr $1/(1+x^\tau)^\lambda$, $x > 0$; $\tau, \lambda > 0$, $\alpha = \tau\lambda$.
- 自由度 (ϕ_1, ϕ_2) の F 分布. $\alpha = \phi_2/2$.
- 自由度 ϕ の t 分布. $\alpha = \phi$.
- $X \sim \text{Ga}(\alpha)$ (ガンマ分布) のとき, $1/X, \log X$ の分布 α .
- Fréchet 分布自身

2. $MDA(H_{2\alpha})$, G_γ , $\gamma = -1/\alpha < 0$. 極値 Weibull 分布

- ベータ分布. $\text{Be}(a, b)$ $\alpha = b$.
- 極値 Weibull 分布自身

3. $MDA(H_0)$, G_0 . Gumbel 分布

- ガンマ分布 $\text{Ga}(\alpha)$. $b_n = \log n + (\alpha - 1) \log \log n - \log \Gamma(\alpha)$, $\varphi(t) = 1$, $a_n = 1$.
- 標準正規分布. $b_n = (2 \log n)^{1/2} - (\log \log n + \log 4\pi)/(2(2 \log n)^{1/2})$, $\varphi(t) = 1/t$, $a_n = 1/b_n$.
- 標準対数正規分布. 上記の正規分布の吸引係数を b_n^* , a_n^* と置き直すと, $b_n = \exp(b_n^*)$, $a_n = a_n^* b_n$.
- ロジスティック分布
- 双曲分布, 一般双曲分布
- 負の Fréchet 分布

1.2.4 極値分布と一般極値分布, von Mises-Jenkinson 形

三つの極値分布をまとめて次のように表わすことができる。ただし, $(1 + \gamma x)^{-1/\gamma} \rightarrow e^{-x}$, $\gamma \rightarrow 0$, を利用している。

$$G_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & x > -1/\gamma, \quad \gamma > 0, \\ \exp(-e^{-x}), & -\infty < x < \infty, \quad \gamma = 0, \\ \exp(-(1 + \gamma x)^{-1/\gamma}), & x < -1/\gamma, \quad \gamma < 0. \end{cases} \quad (4)$$

$\gamma > 0$ は $H_{1\alpha}$ ($\gamma = 1/\alpha$), $\gamma = 0$ は H_0 , $\gamma < 0$ は $H_{2\alpha}$ ($\gamma = -1/\alpha$) に対応している。

この表現は“von Mises-Jenkinson form”と呼ばれている。しばしばまた一般極値分布, (generalized extreme-value distributions), と呼ばれているが, 三つの極値分布の統一型であり, 拡張しているわけではない。パラメータ γ は裾指数 (tail index, extreme value index) と呼ばれている。 γ が増加すると, より裾が重い分布となる。

$\gamma \neq 0$ のときには

$$G_\gamma(x) = \exp(-(1 + \gamma x)_+^{-1/\gamma}), \quad (w)_+ = \max(w, 0). \quad (5)$$

とまとめて表すことができる。 $\gamma = 0$ は $\gamma \rightarrow 0$ の場合とみなせばよい。

一般極値分布導入の意義

形式 (4) は統一した型であるというだけでなく, 推測上の意味がある。第 1 に, どの最大値吸引領域に属するかについての知識が不確かな場合が多い。第 2 に, 極値分布への収束が必ずしも早くない。たとえば正規分布の H_{30} への収束は非常に遅い。定理 3 の関数 φ を用いると

$$n(1 - F(a_n x + b_n)) \approx (1 + \gamma_n x)^{-1/\gamma_n}, \quad \gamma_n = \varphi'(b_n), \quad n \rightarrow \infty,$$

より

$$F^n(a_n x + b_n) \approx \exp(-(1 + \gamma_n x)^{-1/\gamma_n}) = G_{\gamma_n}(x), \quad n \rightarrow \infty,$$

となる。 $G_\gamma(x)$ でなく $G_{\gamma_n}(x)$ による近似を penultimate (極限の一步手前の) 近似と呼んでいる。つまり γ をデータにより推定する方が望ましいことがある。ただし推測の精度を議論するときには新たな困難を生じることになる。

第 3 に, 3 種の型の区別を, 裾指数の推定に帰着し, 位置・尺度パラメータと同時に推定することになる。一般極値分布の分布範囲が裾指数に依存するために, 非正則な最尤推定となるが, Smith (1985) は, Akahira and Takeuchi (1981) の議論を適用して, $\gamma > -1/2$ ならば最尤推定可能であることを示した。この議論は, 次小節で述べる一般 Pareto 分布でも成り立つ。このために裾指数, 位置, 尺度パラメータのすべてを説明変数の関数とすることができ, 極値理論の適用範囲を拡張した。

別紙図一般極値分布参照。

一般極値分布の吸引領域に関する次の十分条件は諸分布の判定に便利である。

定理 4 (一般極値分布の吸引領域)。

$$\phi(x) := \frac{1 - F(x)}{f(x)}, \quad \phi'(x) = \frac{d}{dx} \phi(x), \quad x^* = x^*(F) := \sup\{x : F(x) < 1\}.$$

とすると,

$$\lim_{x \rightarrow x^*} \phi'(x) = \gamma \implies F \in MDA(G_\gamma)$$

逆に, F が 2 階微分可能で, $F \in MDA(G_\gamma)$ であれば $\lim_{x \rightarrow x^*} \phi'(x) = \gamma$.

1.3 水準超過観測値 (POT)

一般に

$$\begin{aligned}\log G(x) &= \log(1 - \bar{G}(x)) \approx -\bar{G}(x) = -1 + G(x), \quad \bar{G}(x) \rightarrow 0, \\ G(x) &\approx 1 + \log G(x) =: W(x),\end{aligned}$$

である。

G を G_γ (または $H_{i\alpha}$) とするとき、変換

$$W(x) = 1 + \log G(x), \quad \log G(x) > -1, \quad (6)$$

で導かれる分布関数 ($W(x) \geq 0$ となるように左側を打ち切った) を調べる。

まず $H_0, H_{1\alpha}, H_{2\alpha}$ を変換した結果は

$$\begin{cases} 1 - e^{-x}, & x > 0, & \text{指数分布} \\ 1 - x^{-\alpha}, & x > 1, \alpha > 0, & \text{Pareto 分布} \\ 1 - (-x)^\alpha, & -1 < x < 0, \alpha > 0, & X \sim \text{Be}(1, \alpha) \text{ のとき } X - 1 \text{ の分布} \end{cases} \quad (7)$$

となる。

これらを統合し、あるいは (4) の G_γ を (6) により変換して

$$\begin{aligned}W_\gamma(x) &= \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma}, & 0 < x < \infty, \quad \gamma > 0, \quad (\gamma = 1/\alpha), \\ 1 - e^{-x}, & 0 < x < \infty, \quad \gamma = 0, \\ 1 - (1 + \gamma x)^{-1/\gamma}, & 0 < x < -1/\gamma, \quad \gamma < 0, \quad (\gamma = -1/\alpha), \end{cases} \\ &= 1 - (1 + \gamma x)_+^{-1/\gamma}, \quad 0 < x < \infty. \end{aligned} \quad (8)$$

を導く。最後の式は (5) 参照。これを一般 Pareto 分布 (generalized Pareto distributions) と呼ぶ。パラメータ γ は一般極値分布の裾指数に対応し、同じ名前と呼ばれる。

分布関数 G, W が (6) の関係を満たすとき、それらの生存関数 \bar{G}, \bar{W} は

$$\bar{G}(x) = 1 - \exp(-\bar{W}(x)) = \bar{W}(x) (1 - 2^{-1}\bar{W}(x) + O(\bar{W}^2(x))), \quad \bar{W}(x) \rightarrow 0,$$

である。一般の分布関数 G について

$$P\{u < X \leq x | u < X\} =: G(x|u) = \frac{G(x) - G(u)}{1 - G(u)} = 1 - \bar{G}(x)/\bar{G}(u), \quad x \geq u,$$

となることに注意すると、

$$|\bar{G}_\gamma(x|u) - \bar{W}_\gamma(x|u)| = O(\bar{W}_\gamma(x|u)), \quad \bar{W}_\gamma(u) \rightarrow 0,$$

となる。つまり、 $F \in MDA(G_\gamma)$ のとき $F^n(a_n x + b_n)$ を G_γ で近似するかわりに、 $F(a_n x + b_n | u_n)$ をより簡単な W_γ で近似することができる。より正確には次の定理が成り立つ。(Balkema and de Haan, 1974; Pickands, 1975)

定理 5.

$$F \in MDA(G_\gamma) \iff \lim_{\bar{F}(u) \rightarrow 0} \frac{\bar{F}(a(u)x + u)}{\bar{F}(u)} = \bar{W}_\gamma(x), \quad \forall x \geq 0, \bar{F}(a(u)x + u) > 0.$$

ただし、 $a(\cdot)$ は吸引係数 a_n を連続化した適当な正の関数である。

1.3.1 一般 Pareto 分布の性質

一般 Pareto 分布 (8) に尺度パラメータを加え, 分布の下限を 0 としたものを改めて

$$F(x) = F(x; \gamma, a) = \begin{cases} 1 - (1 + \gamma x/a)_+^{-1/\gamma}, & \gamma \neq 0, \\ 1 - e^{-x/a}, & \gamma = 0, \end{cases} \quad (9)$$

とする. その分布範囲は

$$\begin{cases} 0 < x < \infty, & \gamma \geq 0, \\ 0 < x < a/(-\gamma), & \gamma < 0, \end{cases}$$

である. (9) の分布を GPrt (γ, a) で表わす. その期待値, ハザード関数は,

$$E(X) = a/(1 - \gamma), \quad \gamma < 1; \quad h(x) = \frac{f(x)}{1 - F(x)} = \frac{1}{a + \gamma x}, \quad \forall \gamma \in \mathcal{R},$$

である. $X > x_0$ の条件の下で, $X \sim \text{GPrt}(\gamma, a)$ の生存関数は

$$P\{X - x_0 > x | X > x_0\} = \frac{1 - F(x_0 + x)}{1 - F(x_0)} = \left(1 + \frac{\gamma x}{a + \gamma x_0}\right)^{-1/\gamma}, \quad x \geq 0; \quad \gamma \neq 0,$$

ただし $\gamma < 0$ のとき $x < a/(-\gamma) - x_0$. つまり $X - x_0 | X > x_0 \sim \text{GPrt}(\gamma, a + \gamma x_0)$ であり, 左側を打ち切ったとき同じ型の分布に従い, 尺度が x_0 の 1 次式として変わる. $\gamma = 0$ の場合は指数分布が 'memoryless である' と呼ばれる性質である. 従って条件付期待値 (平均余命 mean residual life, 平均超過関数 mean excess function) は $E(X)$ の式の a を $a + \gamma x_0$ に変えた

$$m(x_0) := E(X - x_0 | X > x_0) = \frac{a + \gamma x_0}{1 - \gamma}, \quad \gamma < 1, \quad 0 < x_0,$$

である.

別紙図一般 Pareto 分布参照.

定理 6. $X \sim \text{GPrt}(\gamma, a)$, $\gamma < 1$ のとき, $X - x_0 | X \geq x_0$ の確率点は x_0 の 1 次関数である. 上側確率 q の確率点を $Q(q, x_0)$ とすると, 平均超過関数が特定の q にたいする確率点として表わされる:

$$m(x_0) = Q((1 - \gamma)^{1/\gamma}, x_0), \quad 0 < x_0,$$

である. ($\gamma \geq 1$ のときは $m(x_0)$ が存在しない.)

証明.

GPrt (a, γ) の上側確率 q の点は

$$Q(q, 0) = a(q^{-\gamma} - 1)/\gamma.$$

条件付分布の上側確率 q の点は a を $a + \gamma x_0$ に変えて

$$Q(q, x_0) = (a + \gamma x_0)(q^{-\gamma} - 1)/\gamma = Q(q, 0) + (q^{-\gamma} - 1)x_0$$

となる. これを $m(x_0)$ と比較すれば

$$q^{-\gamma} - 1 = \gamma/(1 - \gamma) \quad \text{or} \quad q = (1 - \gamma)^{1/\gamma}$$

である.

定理 7. 正値確率変数 X の平均超過関数 $m(x_0)$ が x_0 の 1 次関数ならば, X は一般 Pareto 分布に従う.

証明.

$$m(x_0) = \frac{1}{\bar{F}(x_0)} \int_{x_0}^{\infty} t f(t) dt - x_0, \quad x_0 > 0,$$

であるから、もし $m(x_0) = \alpha + \beta x_0$, $x_0 > 0$, であると

$$\int_{x_0}^{\infty} t f(t) dt = (\alpha + (\beta + 1)x_0) \int_{x_0}^{\infty} f(t) dt.$$

両辺を x_0 について微分すると

$$(\alpha + \beta x_0) f(x_0) = (\beta + 1) \bar{F}(x_0)$$

したがって

$$\frac{f(x_0)}{\bar{F}(x_0)} = \frac{\beta + 1}{\alpha + \beta x_0} = \frac{1}{a + \gamma x_0}; \quad \alpha = \frac{a}{1 - \gamma}, \quad \beta = \frac{\gamma}{1 - \gamma},$$

これを解けば

$$\bar{F}(x_0) = \left(1 + \frac{\gamma}{a} x_0\right)^{-1/\gamma}$$

γ の符号

データに (9) を当てはめたとき、 γ の符号が興味の焦点となることもある。それを見る簡便な方法は次の標本平均余命 m のグラフである。これは閾値を決定するときの判断にも利用される。

$$\hat{m}(u) := \frac{\sum_{j=1}^n x_j I[x_j \geq u]}{\sum_{j=1}^n I[x_j \geq u]}, \quad (I[\text{true}] = 1, I[\text{false}] = 0).$$

$\gamma < 0$ の場合には、

$$F(x; \eta, \omega) = 1 - (1 - x/\omega)^\eta, \quad 0 \leq x < \omega; \quad \eta = -1/\gamma, \quad \omega = a/(-\gamma), \quad (10)$$

となる。つまり X/ω がベータ分布 $\text{Be}(1, \eta)$ に従う。期待値は $E(X) = \omega/(1 + \eta)$, 条件付期待値 (平均余命) は

$$\begin{aligned} m(u) &= E(X - u | X \geq u) = \frac{\gamma}{1 - \gamma} \left(u + \frac{a}{\gamma}\right) \\ &= \frac{-\gamma}{1 - \gamma} (\omega - u) = \frac{\omega - u}{1 + \eta}, \end{aligned}$$

である。負の勾配

$$\alpha := \frac{-\gamma}{1 - \gamma} = \frac{1}{1 + \eta} = E\left(\frac{X - u}{\omega - u} \mid X \geq u\right),$$

は生続率 (perseverance parameter) と呼ばれている。災厄関数は

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\eta}{\omega} \left(1 - \frac{x}{\omega}\right),$$

である。

寿命分布に上限 $\omega < \infty$ があれば、極限分布で $\gamma < 0$ となるとは限らない。有界の上限でハザード関数 $h(x)$ が必ず無限大となるが、その速度が遅いときだけ $\gamma < 0$ となる。人口学では災厄関数 (人口学では “force of mortality”) が指数的に増加する、無限区間上に分布する Gompertz 曲線が用いられている。これは負の Gumbel 分布で、 $\gamma = 0$ の “最小値” 極限分布のひとつである。高齢者の寿命分布に限ると極値理論の観点から Gompertz 曲線は不自然である。

一般 Pareto 分布から極値分布へ

K を期待値 λ のポアソン分布で、零を打ち切ったときの正整数確率変数とする。つまり

$$P\{K = k\} = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \frac{\lambda^k}{k!}, \quad k = 1, 2, \dots$$

表 2: 一般 Pareto 分布の性質

γ の範囲	$(-\infty, -1]$	$(-1, -1/2)$	$[-1/2, 0)$	$\{0\}$	$(0, 1)$	$[1, \infty)$
分布の別名	ベータ分布			指数分布	Pareto 分布	
分布の範囲	有限区間			上限がない		
期待値	有限期待値					期待値が存在しない
最尤推定量	存在しない	有効性未解決	有効			

$(X_k)_{k=1}^{\infty}$ を分布関数 $F(x)$ をもつ i.i.d. 確率変数列とする。このとき

$$P\{\max(X_1, \dots, X_K) \leq x\} = \frac{e^{-\lambda}}{1 - e^{-\lambda}} (\exp(\lambda F(x)) - 1) = \frac{1}{1 - e^{-\lambda}} (\exp(-\lambda(1 - F(x))) - e^{-\lambda}).$$

特に $F(x) = W_{\gamma}(x)$, (8), とすると

$$\begin{cases} \frac{1}{1 - \exp(-\lambda)} (\exp(-\lambda \exp(-x)) - \exp(-\lambda)), & \lambda = 0, \\ \frac{1}{1 - \exp(-\lambda)} (\exp(-\lambda(1 + \gamma x)^{-1/\gamma}) - \exp(-\lambda)), & \lambda \neq 0, \end{cases}$$

となる。分布範囲は W_{γ} と同じである。これらは一般極値分布の原点における下側打ち切り分布となっている。このモデルは気象、事故などで予想されるものであり、区分最大法を古典理論として無視することはできない。

問題 1. 古典理論に基づく推測法を、この場合に適用できるように修正せよ。

1.4 他の枠組み

1.4.1 順序統計量

連続な分布関数 $F(x)$ からの確率標本 (X_1, \dots, X_n) を大きさの順に並べた

$$X_{1:n} \leq \dots \leq X_{n:n}$$

を順序統計量 (order statistics), $X_{r:n}$ を第 r 順序統計量と呼ぶ。最小値, 最大値はその特別な場合である。 $X_{r:n}$ の分布関数を $G(x)$ とすると,

$$X_{r:n} \leq x \iff \sum_{i=1}^n I[X_i \leq x] \geq r,$$

であるから,

$$G(x) = \sum_{i=r}^n \binom{n}{i} F^i(x) (1 - F(x))^{n-i},$$

$$\frac{d}{dx} G(x) = \frac{n!}{(r-1)!(n-k)!} F(x)^{r-1} (1 - F(x))^{n-r} f(x),$$

$1 \leq r \leq n$, $r = r(n) = ns + O(1)$, $n \rightarrow \infty$, ($0 < s < 1$ は定数) のとき, 標準化した $X_{r(n):n}$ の極限分布は $N(0, 1)$ である。これに対して r を固定したとき, 第 r 最大値 $X_{n-r+1:n}$ の極限分布は極値分布に近い。これについては次に議論する。

$r(n)$ が n に依存するとき, 複雑な状況になる。たとえば松縄規, 中村好延 (2004) サンプルサイズの増加と共に変動する複合極値統計量の分布の情報近似, 統計数理, 52-1, 63-82.

第 r 最大値

第 r 最大値 $X_{n-r+1:n}$ ($n \rightarrow \infty$) の極限分布を求める. ($X_{n-r+1:n}, \dots, X_{n:n}$) の同時分布については次の小節で議論する. $F(x) \in (G_\gamma)$ として F の吸引係数を a_n, b_n とする.

$$Y_n(x) := \sum_{i=1}^n I[X_i > a_n x + b_n],$$

とすると, これは確率 $1 - F(a_n x + b_n)$ の 2 項分布に従い, $n \rightarrow \infty$ のとき,

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\log G_\gamma(x) =: \lambda_\gamma(x),$$

であるから

$$P\{X_{n-r+1:n} \leq a_n x + b_n\} \rightarrow G_\gamma(x) \sum_{i=0}^{r-1} \frac{(\lambda_\gamma(x))^i}{i!} = \frac{1}{(r-1)!} \int_{\lambda_\gamma(x)}^{\infty} e^{-t} t^{r-1} dt,$$

が得られる. つまり第 r 最大値も最大値と同じ吸引係数により, ポアソン確率で修正した極値分布に収束する.

1.4.2 最大 r 個の同時極限分布

分布 F からの確率標本 Y_1, Y_2, \dots, Y_n の順序統計量を $Y_{1:n} \geq Y_{2:n} \geq \dots \geq Y_{n:n}$ とし, 分布 F が一般極値分布 G_γ の吸引領域に属すると仮定する: すなわち 適当な数列 $a_n > 0, b_n \in \mathbb{R}, n = 1, 2, \dots$ が存在し

$$\lim_{n \rightarrow \infty} P\left(\frac{Y_{1:n} - b_n}{a_n} \leq z\right) = G_\gamma(z), \quad \forall z \in \mathbb{R}.$$

このとき, 上位 r 個の順序統計量の同時分布関数

$$P\left(\frac{Y_{1:n} - b_n}{a_n} \leq z_1, \frac{Y_{2:n} - b_n}{a_n} \leq z_2, \dots, \frac{Y_{r:n} - b_n}{a_n} \leq z_r\right), \quad z_1 \geq z_2 \geq \dots \geq z_r,$$

は同時密度関数

$$g_{\gamma, 12 \dots r}(z_1, z_2, \dots, z_r) = \frac{g_\gamma(z_1) \cdots g_\gamma(z_{r-1})}{G_\gamma(z_1) \cdots G_\gamma(z_{r-1})} g_\gamma(z_r), \quad g_\gamma(z) = \frac{dG_\gamma(z)}{dz},$$

を持つ分布関数 $G_{\gamma, 12 \dots r}$ に収束する. David and Nagaraja (2003) の 10.6 節を参照.

$$g_{\gamma, 12 \dots r}(z_1, z_2, \dots, z_r) = \begin{cases} \exp\left(-\sum_{j=1}^r z_j - e^{-z_r}\right), & \gamma = 0, \\ \left(\prod_{j=1}^r (1 + \gamma z_j)^{-1/\gamma-1}\right) \exp\left[-(1 + \gamma z_r)^{-1/\gamma}\right], & \gamma \neq 0, \end{cases}$$

である.

確率ベクトル (Z_1, Z_2, \dots, Z_r) が分布 $G_{\gamma, 12 \dots r}$ に従うとき, $Z_j, j \geq 1$ の周辺分布関数 $G_{\gamma, j}$ は r によらず,

$$G_{\gamma, j}(z) = \begin{cases} \sum_{r=0}^{j-1} \exp(-rz - e^{-z})/r!, & \gamma = 0, \\ \sum_{r=0}^{j-1} (1 + \gamma z)^{-r/\gamma} \exp\left[-(1 + \gamma z)^{-1/\gamma}\right]/r!, & \gamma \neq 0, \end{cases}$$

その周辺密度関数 $g_{\gamma, j}$ は

$$g_{\gamma, j}(z) = \begin{cases} \exp(-jz - e^{-z})/\Gamma(j), & \gamma = 0, \\ (1 + \gamma z)^{-j/\gamma-1} \exp\left[-(1 + \gamma z)^{-1/\gamma}\right]/\Gamma(j), & \gamma \neq 0, \end{cases}$$

となる.

1.4.3 関連の話題, 新記録 (records 数と値)

最大新記録

$(X_i)_{i=1}^{\infty}$ を連続な分布関数 $F(x)$ に従う独立な確率変数の系列とする. もしも $X_k = \max(X_1, \dots, X_k)$, であるとき, X_k は最大新記録である. X_1 も最大新記録であるとみなす. この事象の確率は $F(x)$ によらず,

$$P\{X_k = \max(X_1, \dots, X_k)\} = 1/k,$$

である. 最小新記録を考えても同じである. 実際 (X_1, \dots, X_k) を大きさの順に並べると, $k!$ 通りの順列が等確率で現れ, X_k が新記録となるのはそのうち $(k-1)!$ 通りである. (X_1, \dots, X_n) の中の最大新記録数を K とすると,

$$P\{K = k\} = \binom{n}{k} \frac{1}{n!}, \quad 1 \leq k \leq n,$$

ただし $\binom{n}{k}$ は符号なし第 1 種 Stirling 数で, 母関数

$$t(t+1) \cdots (t+n-1) = \sum_{k=1}^n \binom{n}{k} t^k$$

から定まる正整数である. X_n が最大新記録であるという条件の下で, それ以降の i.i.d. 系列 X_{n+1}, X_{n+2}, \dots において X_{k+m} が X_n より初めて大きい確率は幾何分布で,

$$\frac{1}{n+1} \left(\frac{n}{n+1} \right)^{m-1}, \quad m = 1, 2, \dots$$

つまり $\max(x_1, \dots, x_n)$ の再現期間は $n+1$ である.

新記録値

確率変数の i.i.d 系列の最大新記録だけではなく, その値 X_k の確率分布にも興味がある. つまり “記録値 (record values)” である. 当然その値は X_i の分布関数 $F(x)$ に依存する. たとえば B. C. Arnold, N. Balakrishnan and H. N. Nagaraja (1998) *Records*, Wiley 参照.

i.i.d. でないとき

従属確率変数 X_1, X_2, \dots が \mathbb{R}^d 値 iid 確率変数であるとき, n 変数の間の最小ユークリッド距離 $Y_n = \min_{1 \leq i < j \leq n} |X_i - X_j|$ について

$$P\{n^2 Y_n^d > t\} \rightarrow e^{-ct}, \quad n \rightarrow \infty, \quad c = \frac{\pi^{d/2}}{2\Gamma(1+d/2)} \int_{\mathbb{R}^d} f(x)^2 dx.$$

f は X_i の確率密度関数である. Onoyama, Sibuya and Tanaka (1984)

独立非同値確率変数の和極値分布の一つの Gumbel 分布が, 独立な確率変数の和の極限として得られる, という奇妙な現象がある.

期待値 n/i の独立な指数確率変数 $X_{i,n}$, の 2 重 3 角列 $1 \leq i \leq n, n = 1, 2, \dots$ の和 $Y_n = \sum_{i=1}^n X_{i,n}$ を考えると, $n \rightarrow \infty$ のとき $(Y_n - \log n)/n$ は標準 Gumbel 分布 (分布関数 $\exp(-\exp(-x))$) に分布収束する.

実際期待値 λ の指数分布の m.g.f. は $\lambda/(\lambda - z)$ で, Y_n の m.g.f. は

$$M_n(z) = \Gamma(n+1)\Gamma(1-nz)/\Gamma(n+1-nz)$$

である. $n \rightarrow \infty$ のとき

$$M_n(u/n)/(u \log n) \rightarrow \Gamma(1-u), \quad n \rightarrow \infty,$$

でこれは標準 Gumbel 分布の m.g.f. である.

成功確率 i/n の幾何分布の 2 重 3 角行列の和についても同じ結果が得られている.

Itoh and Mahmond (2005).

2 推測

2.1 統計的補外

極値理論を用いた統計的な補外 (statistical extrapolation) とは, $F \in MDA(G_\gamma)$ を前提として, F の分布の “大きな部分” を G_γ または W_γ で近似し, その位置・尺度, 裾指数 γ を推定し, これらのパラメータの関数も推定することである. パラメータ推定を行うものの, 前提とするモデルが広いという意味でセミパラメトリック法である. 利用可能なデータの特徴によって推測方法が異なる. 次の三つの場合を考える. A (区間最大), B(最大 r), C(水準以上).

A: 区分最大値 (block maxima)

観測数 m のデータが n 組あり, これら全体が確率標本とみなせるときに, 各組の最大値だけを利用する:

$$Z_j = \max(X_{j1}, \dots, X_{jm}), \quad j = 1, \dots, n,$$

$X_{jk} \in MDA(G_\gamma)$ を前提として, (Z_1, \dots, Z_n) を $G_\gamma((z - \xi)/\eta)$ からの確率標本とみなしてパラメータを推測する. パラメータ関数として特に上側 ϵ 確率点

$$z(\epsilon) := \xi + \eta G_\gamma^{-1}(1 - \epsilon)$$

の推測に関心がある. $z(\epsilon)$ にたいして $N = 1/\epsilon$ を “ N return period (再現時期)” と呼ぶことがある. 成功確率 ϵ のベルヌーイ系列で, 事象 $Z > z(\epsilon)$ が起こった時点 (1 シフト幾何分布) の期待値が N である.

Gumbel 分布を前提とし, 確率紙 (Gumbel QQ plot) に直線を当てはめ補外する簡便法が伝統的である. 最尤解をニュートン法で求めると, 当てはまりが非常に悪くない限り収束は早い.

γ が未知であると最尤法が不便であり, 簡便な方法として確率重み付きモーメント (Probability Weighted Moments) 法が便利である.

各組の標本の大きさ m が確率変数のときがある. 典型的なのは腐食で, ある表面面積を観測してもっとも深い腐食孔を計るときである. 腐食孔の数がポアソン分布であり, 腐食孔の深さが独立な指数分布に従えば最大値の分布が Gumbel になることを前に述べた.

B: 最大 r 個

A と同じデータ, 同じ条件の下で, 各組から最大 r 個を測定して利用する. 前節の同時極限分布について最尤推定を行う. この方法は A と C の中間的な性格を持っている. たとえば毎日のデータが得られるが, 季節変動などのために全体を i.i.d. とみなせない. 最大値だけではデータが不足であるときに利用できる. スポーツの記録では, 金賞だけでなく, 銀, 銅, あるいは上位 5 位が記録に残る. 実験データでも, ブロックの最大値だけではもったいないという心情が働くのであろう. しかし, r をどのように定めるか, という問題を生じる. しかも情報量が増えるが期待するほどではない. 高橋, 渋谷 (2004) 参照.

C: 水準超過値 (threshold exceedances)

一般 Pareto 分布と極値分布との関係に注目して, 確率標本の, ある閾値 u より大きな観測値全部にたいして $W_\gamma((y - u)/\eta)$ を当てはめる.

地震, 台風, などの自然災害の発生を点過程 $(T_n)_n$, そのときの (たとえば保険会社の) 損害額を $(X_n)_n$ とする. 確率過程 $(T_n, X_n)_{n=1}^\infty$ は一般に独立でないが, u を閾値として $X_n > u$ を満たすような $n = n(m)$, $m = 1, 2, \dots$ に限った $(T_m, X_m | u)_{m=1}^\infty$ は, $u \rightarrow \infty$ のときに変数付きポアソン過程となる. したがって $(X_m - u)_{m=1}^\infty$ を i.i.d. 確率変数列とみなすことができ, 一般 Pareto 分布を当てはめることができ

る。これが Peaks Over Threshold(POT) 法のひとつの裏付けであり、いろいろな方向への拡張を可能とする。水準超過値にたいしても確率重み付きモーメント法を適用できる。これを拡張して、順序統計量の適当な 1 次結合 (L 統計量) の期待値に基く、L モーメント推定量 (Hosking, 1990) も用いられている。

短い時間で強い相関を持つ時系列では、POT 法を用いても峰がかたまって山脈となってしまう、独立な標本を得ることができない。

2.2 一般極値分布の推定

Fisher 行列

Fisher 情報行列 $I_n(\xi, \eta, \gamma)$ は次のようになる (Prescott & Walden, 1980) :

$$\frac{n}{\eta^2 \gamma^2} \begin{pmatrix} \gamma^2 p & -\gamma\{p - 2\Gamma(2 + \gamma)\} & \eta\gamma \left(q - \frac{p}{\gamma}\right) \\ \cdots & 1 - 2\Gamma(2 + \gamma) + p & \eta \left[1 - \gamma_E + \frac{1 - \Gamma(2 + \gamma)}{\gamma} - q + \frac{p}{\gamma}\right] \\ \cdots & \cdots & \eta^2 \left[\frac{\pi^2}{6} + \left(1 - \gamma_E + \frac{1}{\gamma}\right)^2 - \frac{2q}{\gamma} + \frac{p}{\gamma^2}\right] \end{pmatrix}$$

ただし、 $\Gamma(\cdot)$ はガンマ関数、 $\psi(r) = d \log \Gamma(r) / dr$ で $p = (1 + \gamma)^2 \Gamma(1 + 2\gamma)$ 、 $q = \Gamma(2 + \gamma)(\psi(1 + \gamma) + (1 + \gamma)/\gamma)$ 、 γ_E は Euler の定数で、 $\gamma > -1/2$ のとき正則条件が満たされている。

PWM 法

分布関数 $G(x)$ または確率点関数 $G^{-1}(u)$ が陽に表わされるときにモーメントを拡張した確率加重モーメント

$$M_{kr,s} := E(X^k (G(X))^r (1 - G(X))^s), \quad k, r, s = 0, 1, 2, \dots$$

を用いた推定法が PWM である。

位置 ξ 、尺度 η の一般極値分布 $G_\gamma((x - \xi)/\eta)$ の場合

$$\beta_r := M_{1,r,0} = \frac{1}{r+1} \left(\xi + \frac{\eta}{\gamma} (\Gamma(1 - \gamma)(r+1)^\gamma - 1) \right), \quad r = 0, 1, 2, \quad \gamma < 1,$$

を用いると

$$\frac{2\beta_1 - \beta_0}{3\beta_2 - \beta_0} = \frac{2^\gamma - 1}{3^\gamma - 1}, \quad \eta = \frac{\gamma(2\beta_1 - \beta_0)}{\Gamma(1 - \gamma)(2^\gamma - 1)}, \quad \xi = \beta_0 + \frac{\eta(1 - \Gamma(1 - \gamma))}{\gamma}$$

となる。第 1 式から γ が定まるが、右辺が $\gamma = 0$ 付近でほとんど放物線であり、 $\gamma = 7.8590c - 2.9554c^2$ 、 $c = (\log 2 / \log 3) - (2\beta_1 - \beta_0) / (3\beta_2 - \beta_0)$ で近似できる。

極値データの順序統計量を $x_{(1)} \leq \dots \leq x_{(n)}$ とし、標準一様分布標本の順序統計量を $U_{(1)} \leq \dots \leq U_{(n)}$ とする。 β_r を

$$\hat{\beta}_r = \frac{1}{n} \sum_{i=1}^n E(U_{(i)}^r) x_{(i)}, \quad r = 0, 1, 2, \dots$$

により推定する。

$$E(U_{(i)}^r) = (n - i)^{(r)} / (n - 1)^{(r)},$$

である。上記の近似式を用いると、次の簡単な推定量が得られる。

$$\hat{\beta}_r = \frac{1}{n} \sum_{i=1}^n \left(\frac{i - 0.35}{n} \right)^r x_{(i)}, \quad r = 0, 1, 2,$$

$$\hat{\gamma} = 7.8590c - 2.9554c^2, \quad c = \frac{\log 2}{\log 3} - \frac{2\hat{\beta}_1 - \hat{\beta}_0}{3\hat{\beta}_2 - \hat{\beta}_0}.$$

確率加重モーメント推定量は漸近的に正規分布に従う。たとえば $\hat{\gamma} / \sqrt{0.5633/n}$ が標準正規分布に従うことから、 $\gamma = 0$ の検定を行なえる。

2.2.1 ブートストラップ法

極値理論にブートストラップ法を導入するときの困難と問題点について、Athreya and Fukuchi (1994), Strawderman and Zelterman (1998), Tajvidi (2003) などが議論している。

それにたいして、大きさ n の標本から大きさ b_n の標本を (復元/非復元) 抽出する subsampling 法により、推定するパラメータの信頼区間を構成することが可能であることを Bertail, Politis and Romano (2004) が示した。極値理論における subsampling 法およびその応用について岡 (2007) が詳しく報告している。

2.2.2 ベイズ推定

時間的空間的な拡がりを持つ現象についての極値モデルでは、裾指数、尺度・位置パラメータの恒等性は期待できず、適当な説明変数が利用できなければベイズ・アプローチを採ることになる。たとえば日最大雨量について Coles (2004) の解析がある。問題は裾指数である。裾指数の推定は困難であり、目立った傾向を示すことが少ないので、一定とすることが多い。

そもそも一般極値分布の裾指数について 2 つのアプローチがある。通常勧められているのは、 γ も一つのパラメータとして推定するのであるが、この方法では $\gamma = 0$ (Gumbel 分布) は決してモデルとして選択されないことになる。 $F \in MDA(G_0)$ であっても収束が遅ければ $G_\gamma, \gamma \neq 0$ がデータにより適合するので、それ自体はよいように見えるが、長い期間にたいする再現水準の推測では不適切となる。

もう 1 つのアプローチでは、3 つの仮説 $\gamma = 0, \gamma > 0, \gamma < 0$ のひとつを検定で選択することである。このアプローチでは $\gamma = 0$ が存在しうることを前提とするし、検定の不確実性が、その後の推測に及ぼす影響を考慮していない。

これらの問題の解決として、Coles and Tawn (2004a,b) は $\gamma = 0$ に正の確率を与える事前分布を提案している。例として年最高海面水位と日降水量のデータを解析している。後者ではに対する事前確率が 0 であった。後者のデータについて、別の論文、Coles and Tawn (2004a) では専門家の知識を取り入れる事前確率の導入法を議論している。

2.3 一般 Pareto 分布の当てはめ

一般 Pareto 分布を当てはめるときには適当な閾値を定めること、裾の指標を定めることが必要となる。いずれも容易ではなく、多くの研究の対象となっている。すべてのパラメータを経験ベイズ法で解く方法もある。探索的データ解析では次に述べる平均超過関数の標本版を両く ME プロットが勧められている。

閾値の決定

一般に生存関数 \bar{F} 、密度 f をもつ確率変数 X にたいして

$$m(u) := E(X - u | X > u) = \frac{1}{\bar{F}(u)} \int_u^\infty (x - u) f(x) dx = \frac{1}{\bar{F}(u)} \int_u^\infty \bar{F}(x) dx,$$

とする。最後の式は $x\bar{F}(x) \rightarrow 0, x \rightarrow \infty$ のとき成り立つ。 X が寿命の場合には時点 u における平均余命 (mean residual life) であるが、これを平均超過関数 (mean excess function) と呼んでいる。逆に $m(u)$ により確率分布が指定される。 \bar{F} が連続で $\bar{F}(0) = 1$ のとき

$$\bar{F}(x) = \frac{m(0)}{m(x)} \exp \left\{ - \int_0^x \frac{1}{m(u)} du \right\},$$

である。

一般 Pareto 分布 $W_\gamma(x/a)$ にたいする平均超過関数は、 $\gamma < 1$ のとき

$$m(u) = (a + \gamma u) / (1 - \gamma), \quad a + \gamma u > 0,$$

で u の一次関数である。ME プロットは、標本 $(X_i)_{i=1}^n$ にたいして

$$\hat{m}(u) := \sum_{i=1}^n (X_i - u) I[X_i > u] / \sum_i I[X_i > u],$$

を描くことになる。ただしこの図のぼらつきは大きい。

$\gamma = 0$ に限ると $X_1 \geq \dots \geq X_k \geq u$ が指数分布からの順序統計量とみなせるような u を決定することになる。いくつかの方法が提案されているが、効率のよい方法はない。

γ の最尤推定と PWM 法

一般 Pareto 分布の大きさ n の確率標本の対数尤度は

$$l = -\frac{1+\gamma}{\gamma} \sum_{i=1}^n \log(1 + \tau x_i) - n \log a, \quad \tau = \gamma/a,$$

である。これから最尤方程式は

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \log(1 + \tau x_i) &= \gamma, \\ \frac{1+\gamma}{\gamma} \frac{1}{n} \sum_{i=1}^n \frac{x_i}{1 + \tau x_i} &= \frac{1}{\tau}, \end{aligned}$$

となる。第1式を第2式に代入すれば τ についての最大化となる。Fisher 情報行列は

$$I(\gamma, a) = \frac{n}{(2\gamma+1)(\gamma+1)} \begin{pmatrix} 2 & 1/a \\ 1/a & (\gamma+1)/a \end{pmatrix}, \quad \gamma > -1/2,$$

その逆行列は

$$\text{Var}((\hat{\gamma}, \hat{a})) = \frac{\gamma+1}{n} \begin{pmatrix} \gamma+1 & -a \\ -a & 2a^2 \end{pmatrix},$$

となる。

一般 Pareto 分布 $W_\gamma(x/a)$ の場合の PWM 法は

$$\beta_s := M_{1,n,s} = \frac{a}{(s+1)(s+1-\gamma)}, \quad s = 0, 1,$$

を用いて

$$a = \frac{2\beta_0\beta_1}{\beta_0 - 2\beta_1}, \quad \gamma = 2 - \frac{\beta_0}{\beta_0 - 2\beta_1},$$

を標本確率加重モーメントと等しく置いて解く。

Hill 推定量と Pickands 推定量

平均超過関数は閾値 u の値を定める目安となるが、その勾配から γ を推定することができる。特に $X \sim F \in MDA(G_\gamma), \gamma > 0$, を仮定して、簡単のため $X > 1$ とすると、

$$E(\log X - \log t \mid \log X > \log t) \rightarrow \frac{1}{\alpha}, \quad t \rightarrow \infty,$$

となる。この事実を ME プロットに適用すると次の Hill 推定量が得られる。(Hill, 1975)

$F \in MDA(H_{1\alpha})$ からの標本の順序統計量 $X_{1:n} \geq \dots \geq X_{k+1:n} (\geq u)$ に基づいて

$$\hat{\alpha}_H = \left(\frac{1}{k} \sum_{j=1}^k \log X_{j:n} - \log X_{k+1:n} \right)^{-1},$$

により推定する。詳細を第4節で述べる。

これと類似している、ある吸引条件から導いた、Pickands 推定量がある。今度は $F \in MDA(G_\gamma)$ を前提とする:

$$\hat{\gamma}_P = \frac{1}{\log 2} \log \frac{X_{k:n} - X_{2k:n}}{X_{2k:n} - X_{4k:n}}.$$

これは $k \rightarrow \infty, k/n \rightarrow \infty$ のとき、 γ に確率収束する。(Pickands, 1975; Dekkers and de Haan, 1989).

Pareto プロット

X が Pareto 分布 $\text{Par}(\alpha)$ に従うとき、 $\alpha \log X$ は標準指数分布に従う。Pareto 分布からの標本の順序統計量を $x_1 \geq \dots \geq x_n > x_{n+1} = u$ とし指数 QQ プロット

$$\left(-\log \frac{n+1-i}{n+1}, \log x_{n+1-i} \right), \quad i = 1, \dots, n,$$

を描けばだいたい直接にのる。これが Pareto プロットであるが、ジッフ・プロットでもある。適当な重みの最小2乗法によりパラメータを推定する方法が提案されている。

Beirlant, et al. (1999) Tail index estimation and an exponential regression model, *Extremes*, 2, 177-200.

補外

$F \in MDA(G_\gamma)$ の大きさ n の確率標本より閾値 u を超えるデータにたいして一般 Pareto 分布 W_γ を当てはめ、 F の上側確率 $p (\ll 1)$ の確率点を推定する。

$$\bar{F}(u+y) = F(u)\bar{F}_u(y), \quad \bar{F}_u(y) \approx \bar{W}_\gamma(y/\eta)$$

であり、 $\bar{F}(u)$ は N_u/n , $N_u = \sum_{i=1}^n I[X_i > u]$, で推定できる。したがって第1式の標本版は

$$\widehat{\bar{F}}(u+y) = \frac{N_u}{n} \left(1 + \hat{\gamma} \frac{y}{\hat{\eta}} \right)^{-1/\hat{\gamma}},$$

となり、 $x(p)$, $\bar{F}(x(p)) = p$, の推定量として

$$\hat{x}(p) = u + \frac{\hat{\eta}}{\hat{\gamma}} \left(\left(\frac{n}{N_u p} \right)^{-\hat{\gamma}} - 1 \right),$$

を用いることになる。もしも $\hat{\gamma} < 0$ であるならば、分布の上限を

$$\hat{x}(0) = u - \hat{\eta}/\hat{\gamma},$$

で推定することになる。

2.4 参考書とソフトウェア

参考書

1. Beirlant, Goegebeur, Segers and Teugels (2004) 分厚い (490p.) だけに推測理論に詳しい。データ解析例も豊富である。11章中最後の3章、多変量、時系列、ベイズ法、は別の著者による新しい話題の説明で、それまでのように体系的ではない。本文で利用しているSプログラム、データが出版社ウェブサイトから利用可能である。
2. Coles (2001) 簡潔 (208p.) で読みやすい。数理統計学を省いてあるが、概念の意味は的確に説明しており、本書に沿ったS, Rのパッケージが利用可能であり、その使用例も説明されている。

3. Falk, Hüsler and Reiss (2004) 数学的体系的である。汎関数的小数法則を点過程に適用して、極値理論を展開している。
4. Finkenstädt and Rootzén (2004) Gothenburg で開かれたセミナーに招待された 7 人の、諸トピックスについての講義ノート。
5. de Haan and Ferreira (2006) 極値統計の確率統計理論として、もっとも新しく体系的である。著者自身が開発した正則変動関数理論に基づいている。
6. Kotz and Nadarajah (2000) パラメトリックモデルが、多変量極値分布も含めて数多く集められている。
7. McNeil, Frey and Embrechts (2005) まえがきで述べた QRM に関する 10 章中に、多変量モデル、時系列、接合関数と従属性、極値理論が 1 章ずつ当てられている。それぞれオリジナリティが高い。

ソフトウェア

フリー・ソフトウェア環境 R (www.r-project.org) のパッケージは CRAN (The Comprehensive R Archive Network, cran.r-project.org) から利用可能である。その中の極値理論関係パッケージには次のものがある。

- `evd`(2006-09-29)
A. Stephenson (Lancaster Univ.) が開発した諸分布関数の計算と 1, 2 変数の基本的諸方法のためのパッケージ。
- `ismev`(2006-03-10)
Coles (2001) の S プログラムを E. Gilleland, R. Katz, G. Young が R 用に再開発したものである。
- `extRemes`(2007-01-10)
`ismev` の初心者教育用に準備した GUI で、気象・自然現象のデータフレームが用意されている。
- `evir`(2007-04-28)
A. McNeil が開発した S 言語のための EVIS を A. Stephenson が R に書き直した extreme values in R. 極値理論のための多様な諸関数が用意されている。
- `evdbayes`(2006-10-19)
A. Stephenson and M. Ribatet. GEVD, 点過程などのモデルの MCMC によるベイズ推定パッケージ。
- `VaR`(2004-01-13)
T. Danilyarov (Heidelberg U.) による VaR 推定の諸関数。
- `fExtremes`(2007-0-35)
ファイナンス用ソフトウェアを開発している R metrics グループが開発した包括的なパッケージで、`ismev`, `evir` などのパッケージにある関数も取り入れている。

これらについて、また S その他のパッケージについては Stephenson and Gilleland (2006) の解説を参照。

3 応用

あるときから、リスクという英語がいくつかの意味をもって、いろいろな場面で使われるようになり、研究者が研究費を獲得するためのキーワードにもなっている。統計学に深く関連するリスク問題は次のように分けられる。

1. 自然災害
2. 事故・故障 (設計時における信頼性, 使用時における安全管理)
3. 産業・生活による自然環境破壊, その中心である地球温暖化
4. 生産・消費活動を脅かす経済・社会の不安定性

これらの問題すべてに役立つ特効薬を期待することはできない。個々の問題はそれぞれの考察を必要とする。

3.1 Delta 計画

1953年1月31日北海の嵐による堤防の破壊でオランダ南部 Zeeland 州など Delta 地帯 1500 km² が水没, 1853 人が死亡した. 国家的 Delta Project 1968-1987 が計画実行され, 3.2 km の防波堤防が建設された. de Haan (1990) 参照.

Return Period 10,000 年を政府が提示した. 別の根拠として, 実際には嵐と最高潮位がずれていたのを, もしも一致していたらという最悪のシナリオから導かれた数値でもあるという.

この課題はさらに発展し, ヨーロッパ諸国の共同研究計画として取上げられ, 多変量極値統計量の理論が応用された (後述).

3.2 MV Derbyshire 号遭難事件

1980年9月9日, 台風 15 号 (Orchid) の際に, ケッククから川崎に鉄鉱石を輸送していたイギリス 9 万トン貨物船 MV Derbyshire 号が, 連絡のないまま, 乗船していた 44 人とともに日本の南東約 600 km で沈没した. 台風 15 号は暴風雨圏が非常に大きく, 長時間継続したのが特徴であった.

1987 年	原因不明の海難審査報告 遺族の強い要求により再調査
1994 年	4200 m の海底で沈没船発見
1997-98 年	135,774 枚のデジタル写真, 200 時間のビデオ撮影
報告書	前方甲板船倉ハッチカバーより浸水が一次的原因 (船員の不注意)
再調査	写真再検討, 水槽模型船実験, 実験データの統計的解析
2000 年 8 月	最終報告

海難審査報告では原因不明となったが遺族の強い要求により再調査が行われた. 沈没船のデジタル写真, ビデオ撮影により空気取入口からの浸水, バラストのアンバランス, から最終的に船倉のハッチカバーの破壊という原因が判明した. 問題は同型船の設計基準が諸運行条件で安全に航海できるものか, の検討となった. 船型, 積載量, 波にたいする進行方向, 台風の特徴, 波浪が主要因で, これらの種々の条件の下での水槽模型船実験が行われ, 実験データの統計的解析により, 問題部位への最大荷重の確率分布が推定された. その結果, 設計基準の改定勧告がなされ, 海洋波浪の衛星による推定が計画された.

Heffernan and Tawn (2001) 参照. 事件の詳細は www.mv-derbyshire.org.uk/

3.3 Venezuela 大洪水

1991年12月 Venezuela 大雨・洪水が起こり, 50,000 人が死亡し, 日本の援助で建てられたばかりの病院が崩壊した. 災害からの救出, 医療のために国際的な支援が行なわれた. Maiquetia 国際飛行場の降雨データによると, そのときの日降雨量 410.4 mm は, 過去 50 年間の最大日雨量約 150 mm を遙かに超えるものであった. Coles (2004) は過去 40 年間の Maiquetia 国際飛行場日雨量データを解析した. 1 年を 3 時期に分け, 各時期における一般極値分布パラメータが, 独立に変動するというベイズモデルを考えて MCMC を用い, その事後分布において降雨量 410.4 mm は不可能な値ではなく, 上側確率 0.67% の確率点であることを示している. 極値理論を適用するための条件が満たされていないと判断する場合に, モデルの不安定性を考慮する手段としてベイズが使われている.

3.4 損害保険への応用

損害保険における高額保険金支払いとして, たとえば Large Norwegian Fire Claim Data, Schnieper R (1993) の例がある. 再保険の方式には,

- excess-of-loss (XL) 個々の支払い請求について, 金額が一定値を超えたときに保証
- stop-loss or total loss ある portfolio にたいする支払い請求総額が一定値を超えたときに保証

- ECOMOR (Excédent du coût moyen relatif) ある portfolio 中の個別支払い請求の最大 k 件について保証

いずれにせよ、限界額を与えれば POT モデルを適用できる。再保険請求額について一般 Pareto 分布を当てはめ、諸パラメータ関数を推定する。あるいはパラメータ値を仮定してシミュレーションを行う。

破産の問題 (Ruin Problem) とすると, initial reserve : $s = s(q, T)$
 期間 T 内に破産が起こる確率を q とする s を決める。

損害保険については T. Rolski, H. Schmidli, V. Schmidt and J. Teugels (1998) Stochastic Processes for Insurance and Finance, Wiley,
 T. Mikosch (2006) Non-Life Insurance Mathematics, Springer.

3.5 Extreme Returns in Asset Prices

時点 $t = 0, 1, 2, \dots$ における投機商品価格を s_t としたとき $r_t := \log(s_t/s_{t-1})$ を log-returns (geometric returns), $\tilde{r}_t = (s_t - s_{t-1})/s_{t-1}$ を arithmetic returns と呼ぶ。次のようなことが “stylized facts” として主張されている。1. log-return の方が扱いやすい。2. r_t はマルチンゲールである。3. r_t は裾の重い分布である。4. 相関があるが強くなく、短期間である。5. tranquility / volatility の期間に分かれる。

裾と中心の両方を問題にしたいので Gaussian mixture, Student, sum-stable 分布などの分布をモデルとする。一方規制により VaR および CaR が主要な関心である。

Basel Committee of Banking Supervision: Basel Accord (I 1988, II 1993) the Value-at-Risk (VaR): the limit which is exceeded by the loss of a given speculative asset or a portfolio with a specified low probability.

the Capital-at-Risk (CaR): the amount which may be invested so that the loss exceeds a given limit with a specified low probability.

3.6 金属疲労と “Wicksell’s corpuscle problem”

金属疲労 (metal fatigue) は多くの事故の原因である。金属に荷重を加え、元に戻すと、弾性変形の範囲内では元の形に戻る。しかしこれを数百万回反復すると、表面・内部に傷が生じ亀裂が成長して破壊に到る。村上敬宜 (九大教授) は高品質鋼の疲労寿命が、応力集中部分にある非金属介在物の最大寸法で定まることを理論的・実験的に示した。介在物はほとんど球状で、鋼内部にランダムに点状している。鋼試験片を切断し切断面の円形介在物の寸法データから空間の球の寸法分布を推定しなければならない。(問題は平面に点状する円を直線で切断し、切断区間長から円の直径の分布を求めることと同等である。) この問題は “ステレオロジー (空間統計学)” の古典的な問題で **Wicksell’s corpuscle problem** と呼ばれ、逆問題の典型例として知られている。(Sibuya, 1999; Takahashi and Sibuya, 2002; Bortot, Coles and Sisson, 2007).

球の中心が強度 λ_V のポアソン過程に従い、大円面積 S_V がこれと独立で確率密度関数 $f_V(s)$ をもち、 $E(\sqrt{S_V}) = m$ とすると、切断面円形の中心は強度 $\lambda_A = 2m\lambda_V/\sqrt{\pi}$ のポアソン過程に従い、切断面の面積 S_A の確率密度関数は

$$f_A(t) = \frac{1}{2m} \int_t^\infty \frac{1}{\sqrt{s-t}} f_V(s) ds, \quad 0 < t < \infty,$$

となる。逆に

$$f_V(s) = -\frac{2m}{\pi} \frac{d}{ds} \int_s^\infty \frac{1}{\sqrt{t-s}} f_A(t) dt, \quad 0 < s < \infty,$$

である。生存関数についても同様の式が得られ、これらを総称して “Wicksell 変換” という。

f_V, f_A の分布関数 F_V, F_A の一方が一般極値分布の吸引領域 $DAM(G_\gamma)$ に入ると、他も吸引領域に入るが、極値指数が次のように変わる。問題は F_A の標本に基づく F_V の最大値の予測である。

$$F_V \in MDA(G_\gamma) \iff F_A \in MDA(G_{2\gamma/(2-\gamma)}).$$

$$F_V \in MDA(G_{2\gamma^*/(2+\gamma^*)}) \iff F_A \in MDA(G_{\gamma^*}).$$

3.7 “超高齢者”の寿命と表データ

日本人の平均寿命は男女とも世界最高であり、半世紀以上も平均寿命が確実に増加している。しかし、最近の超高齢の寿命分布を調べると、分布そのものはそれほど変化しておらず、高齢者数の急速な増加が平均寿命を延ばしている原因となっている。

利用可能なデータとして、人口動態統計データ(死亡表)を全国高齢者調査(生存統計)が利用できる。

人口動態統計における84歳以上、高齢者調査の100歳以上のデータについて、標本平均余命を計算すると、ほとんど直線的に減少し一般 Pareto 分布 ($\gamma < 0$) が当てはまると予想される。種々の解析の結果では、大体当てはまりがよく、寿命分布には有界な上限があると考えの方がよい。人口学ではハザード関数が指数的に増加する Gompertz 曲線がもっぱら利用されている。

調査統計は通常、表の形で公表されている。死亡統計はある年度に各年齢で死亡した人の数の表である。生存統計は毎年のある日に生存している各年齢の人の数である。(表を小さくするために、年齢を5歳区分にしたり、毎年でなく5年ごとに調査したりする。) このように表に区分されたデータに対し、一般 Pareto 分布をあてはめるための理論は十分に研究されていない(渋谷, 華山, 2004)。

4 基礎理論

4.1 最大値分布収束の図示

最大値の分布関数の一般極値分布への収束

$$F^n(a_n x + b_n) \rightarrow G_\gamma(x), \quad n \rightarrow \infty,$$

を図示するためには F の吸引係数 (a_n, b_n) が必要であるが、 F の確率点関数 F^{-1} が既知であれば、汎用のプログラムを作ることができる。 $F_n, n = 1, 2, \dots$ の位置・尺度パラメータを定めるのに、任意に選んだ 2 つの確率点 (たとえば第 1、第 3 四分位点) を G_γ の確率点と等しくとる。つまり

$$\begin{aligned} \xi: G_\gamma(\xi_i) = p_i, \quad i = 1, 2, \quad p_1 := 1/4, p_2 := 3/4, \\ F^n(a_n \xi + b_n) = p_i \iff a_n \xi_i + b_n = F^{-1}(p_i^{1/n}) =: \eta_{ni}, \quad i = 1, 2, \end{aligned}$$

これを解くと

$$a_n = \frac{\eta_{n2} - \eta_{n1}}{\xi_2 - \xi_1}, \quad b_n = \frac{\xi_1 \eta_{n2} - \xi_2 \eta_{n1}}{\xi_2 - \xi_1}, \quad n = 1, 2, \dots,$$

である。これより、たとえば最大値の pdf は

$$n a_n f(a_n x + b_n) F^{n-1}(a_n x + b_n)$$

により計算する。別紙図一般極値図にその例を示す。

4.2 一般極値分布の吸引領域に関する十分条件

定理 8 (von Mises, 1936). F を分布関数, f をその密度関数とし,

$$\phi(x) := \frac{1 - F(x)}{f(x)}, \quad \phi'(x) = \frac{d}{dx} \phi(x), \quad x^* = x^*(F) := \sup\{x : F(x) < 1\},$$

とすると,

$$\lim_{x \rightarrow x^*} \phi'(x) = \gamma \implies F \in MDA(G_\gamma).$$

このとき、吸引係数 (a_n, b_n) は $n(1 - F(b_n)) = 1, a_n = \phi(b_n)$ と取ればよい。逆に、 F が 2 階微分可能、 $F \in MDA(G_\gamma)$ であれば $\lim_{x \rightarrow x^*} \phi'(x) = \gamma$ 。

□

証明 $F(x)$ の累積ハザード関数を $H(x) = -\log(1 - F(x))$, ハザード関数を $h(x) = (d/dx)H(x) = f(x)/(1 - F(x)) = 1/\phi(x)$ とする。

$$\begin{aligned} H(u+v) - H(u) &= \int_0^v h(u+w)dw = \int_0^x h(u+s\phi(u))\phi(u)ds, \\ v &= x\phi(u), \quad w = s\phi(u). \end{aligned}$$

被積分関数の逆数を変形する。最後の等式は平均値の定理による。

$$\begin{aligned} \frac{1}{h(u+s\phi(u))\phi(u)} &= \frac{\phi(u+s\phi(u))}{\phi(u)} \\ &= 1 + \frac{\phi(u+s\phi(u)) - \phi(u)}{\phi(u)} = 1 + s\phi'(y), \quad u < y < u + s\phi(x). \end{aligned}$$

上の積分に代入すると

$$\begin{aligned} \int_0^x \frac{ds}{1+s\phi'(y)} &= \int_1^{1+x\phi'(y)} \frac{dt}{t} \frac{1}{\phi'(y)} = \frac{1}{\phi'(y)} \log(1+x\phi'(y)), \\ t &= 1+s\phi'(y), \quad u < y < u + s\phi(x). \end{aligned}$$

したがって

$$\begin{aligned}\frac{1 - F(u + x\phi(u))}{1 - F(u)} &= \exp(-(H(u + x\phi(u)) - H(u))) \\ &= (1 + x\phi'(u))^{-1/\phi'(u)}, \quad u < y < u + x\phi(x).\end{aligned}$$

$(a_n), (b_n)$ を $n(1 - F(b_n)) = 1, a_n = \phi(b_n)$, により定めると, 定理の条件から $\phi'(b_n) \rightarrow \gamma(b_n \uparrow x^*)$ であり,

$$\begin{aligned}n(1 - F(b_n + a_n x)) &= \frac{1 - F(b_n + a_n x)}{1 - F(b_n)} \approx (1 + \phi'(b_n)x)^{-1/\phi'(b_n)} \\ &\rightarrow (1 + \gamma x)^{-1/\gamma} = -\log G_\gamma(x), \quad (n \rightarrow \infty), \\ F^n(a_n x + b_n) &\rightarrow G_\gamma(x), \quad (n \rightarrow \infty).\end{aligned}$$

注意 この証明は, 水準超過値の一般 Pareto 分布への収束も同時に示している.

de Haan and Ferreira (2006) Chap. 1 に, この十分条件にたいする別のアプローチがある.

4.3 再現水準と裾確率点関数

成功確率 p のベルヌーイ系列で, 成功までの試行数を K とすると

$$P\{K > k\} = (1 - p)^k, \quad P\{K = k\} = p(1 - p)^{k-1}, \quad k = 1, 2, \dots, \quad E(K) = \frac{1 - p}{p} + 1 = \frac{1}{p}.$$

したがって分布関数 $F(x)$ をもつ i.i.d. 確率変数列 $(X_k)_{k=1}^\infty$ で,

$$1 - F(x_n) = 1/n, \quad \text{or} \quad x_n = F^{-}(1 - 1/n),$$

を満たす確率点 x_n を超える X_k が出現するまでの試行数 K_n の期待値が n である. 自然災害対策工事の安全基準で上側確率 $1/n$ の確率点 $F^{-}(1 - 1/n)$ を n 再現期間 (return period) と呼ぶ. 河川の堤防は 100 再現年 (return year) を基準とすることが多い. 確率点関数 $F^{-}(\cdot)$ については補い 4.6.1 参照.

$$P\{K_n > x_n\} = P\{\max(X_1, \dots, X_n) < x_n\} = (1 - 1/n)^n \approx e^{-1} = 0.274$$

であり, 再現が再現期間を過ぎる確率はかなり大きいことに注意.

応用範囲が拡がり, 必ずしも時間に限られないことから, $1 - F(x) = 1/t$ の解 $x(t) = F^{-}(1 - 1/t)$ を, 期間 t に対応する再現水準 t period return level と呼ぶようになった.

4.3.1 再現水準と裾確率点関数

分布関数 $F(x)$ からの大きさ n の確率標本の最大値の分布関数のある点 x_n における値は

$$F^n(x_n) = (1 - (1 - F(x_n)))^n \approx \exp(-n(1 - F(x_n))), \quad n \rightarrow \infty.$$

これが意味をもつのは

$$n(1 - F(x_n)) \approx 1 \iff F(x_n) \approx 1 - 1/n, \quad n \rightarrow \infty,$$

のときである. n の代わりに連続変数 t を用いて n と x_n との対応を $x = x(t)$ とする:

$$F(x) = 1 - \frac{1}{t} \iff t = \frac{1}{1 - F(x)} \iff x = F^{-}(1 - 1/t).$$

最後の $x = x(t)$ を改めて

$$U(t) = U(t; F) = \left(\frac{1}{1 - F(x)} \right)^- = F^{-} \left(1 - \frac{1}{t} \right), \quad 1 < t < \infty,$$

と表し, F の裾確率点関数 (tail quantile function) と呼ぶ. 期間 t に対する再現水準にほかならない. U は非減少関数で F より一意に定まる.

- 一様分布: $F(x) = x, U(t) = 1 - 1/t$
 - 指数分布: $F(x) = 1 - e^{-x}, U(t) = \log t$
 - パレートの分布: $F(x) = 1 - x^{-\alpha}, x > 1, U(t) = t^{-\alpha}, \alpha > 0$
- 一般極値分布, 一般パレートの分布の裾確率点関数は下記の表の通りである. 別紙図裾確率点関数参照.

表 3: 裾確率点関数

	G_γ	$U(t)$	W_γ	$U(t)$
$\gamma \neq 0$	$\exp(-(1 + \gamma x)^{-1/\gamma})$	$\frac{1}{\gamma}((\log \frac{t}{t-1})^{-\gamma} - 1)$	$1 - (1 + \gamma x)^{-1/\gamma}$	$\frac{t^\gamma - 1}{\gamma}$
$\gamma = 0$	$\exp(-\exp(-x))$	$-\log(-\log(1 - \frac{1}{t}))$	$1 - \exp(-x)$	$\log(t)$

4.3.2 裾確率点関数の収束, 最大吸引領域

$F \in MDA(H)$ つまり

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x),$$

とする. 両辺の対数を取り, $\log(y) \approx 1 - y, y \rightarrow 1$ を用いると

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\log H(x),$$

両辺の逆数の逆関数を考えれば

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = H^-(e^{-1/x}) =: D(x),$$

となる. したがって a_n, b_n を $a(t) := a_{[t]}, b(t) := b_{[t]}, [\cdot]$ は整数部分, に変えると, 次の収束 3 条件は同等となる.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x), \quad (11)$$

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\log H(x), \quad (12)$$

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x) = H^-(e^{-1/x}). \quad (13)$$

$H = G_\gamma$ に限ると $F \in MDA(G_\gamma)$ の必要十分条件が得られる.

定理 9. $F \in MDA(G_\gamma)$ と同等な条件が得られる.

(a) 適当な $a(t) > 0$ に関して

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \forall x > 0. \quad [R.H.S. = \log(x) \text{ if } \gamma = 0.]$$

(b) 適当な $a(t) > 0$ に関して

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \gamma x)^{-1/\gamma}, \quad \forall x: 1 + \gamma x > 0.$$

(c) 適当な $f(t) > 0$ に関して

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}, \quad \forall x: 1 + \gamma x > 0, \quad x^* = \sup\{x: F(x) < 1\}.$$

$f(t) = a(1/(1 - F(t)))$. とすることができる. また吸引係数を $b_n = U(n), a_n = a(n)$ とすることができる.

定理 10. $F \in MDA(G_\gamma)$ と同等な条件.

$\gamma > 0$ のとき

$$U(y) \in \mathbf{R}_\gamma \text{ or } \bar{F}(x) \in \mathbf{R}_{-1/\gamma}.$$

$\gamma < 0$ のとき

$$\bar{F}(x^* - 1/x) \in \mathbf{R}_{1/\gamma}, \quad x^* = \sup\{x : F(x) < 1\}.$$

\mathbf{R}_γ は指数 γ の正則変動関数の集合である. 補い 4.6.2 参照.

裾確率点関数を用いると, 極値統計学の数学理論がより扱いやすくなる. de Haan and Ferreira (2006) は裾確率点関数を “安定分布とその吸引領域における特性関数の役割を, 極値理論で果たすもの” と評価している.

4.4 Hill 推定量

分布関数 F が Fréchet 分布, すなわち一般極値分布 G_γ , $\gamma > 0$, の最大値吸引領域に属している ($F \in MDA(G_\gamma)$) とする. このことは, 次の条件

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad \gamma > 0,$$

と同値である. つまり生存関数が指数 $-1/\gamma$ の正則変動関数である. この条件はまた

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \lim_{t \rightarrow \infty} \frac{\int_t^\infty (\log u - \log t) dF(u)}{1 - F(t)} = \gamma,$$

と同値になる. 上式の等号は部分積分による.

この漸近的な条件から γ の推定量を構成する. Y_1, Y_2, \dots, Y_n を F からの確率標本とし, $Y_{1,n} \leq Y_{2,n} \leq \dots \leq Y_{n,n}$ をその順序統計量とする. 上式で t を $Y_{n-k,n}$, F を経験分布関数で置き換えると, Hill 推定量 (Hill, 1975)

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=0}^{k-1} \log Y_{n-i,n} - \log Y_{n-k,n},$$

が得られる.

最尤法や平均超過関数を用いても Hill 推定量を導出することができる. 上の説明から, 正則変動関数が本質的な他の分野でも, Hill 推定量を利用できる. Hill 推定量については, Embrechts et al. (1997), Beirlant et al. (2004), de Haan and Ferreira (2006) に詳しい説明がある.

Hill 推定量の性質

定理 Y_1, Y_2, \dots を分布 F からの確率標本とする.

(1) $F \in MDA(G_\gamma)$, $\gamma > 0$ で, $n \rightarrow \infty$, $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$ ならば

$$\hat{\gamma}_H \xrightarrow{P} \gamma.$$

(2) 逆に, $n \rightarrow \infty$ で, $k = k(n) \rightarrow \infty$, $k(n)/n \rightarrow 0$, $k(n+1)/k(n) \rightarrow 1$ のとき

$$\hat{\gamma}_H \xrightarrow{P} \gamma > 0,$$

ならば $F \in MDA(G_\gamma)$.

(3) 分布関数 F が, 次の 2 次の性質

$$\lim_{t \rightarrow \infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-1/\gamma}}{A\left(\frac{1}{1-F(t)}\right)} = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\gamma\rho},$$

(ここで, $\gamma > 0$, $\rho \leq 0$ で A は正または負の関数で $\lim_{t \rightarrow \infty} A(t) = 0$) を満たすならば, $n \rightarrow \infty$ で $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$ のとき

$$\sqrt{k}(\hat{\gamma}_H - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

ここで,

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda \text{ 有限.}$$

これらの結果から Hill 推定量は一致推定量であり, 漸近的に正規分布をするが偏りがあることがわかる. Hill 推定量に関して, 最適な k を決める, iid の条件をゆるめる, 改良する等の研究がある.

4.5 点過程モデル

4.5.1 水準超過と Poisson 過程

適当な空間 \mathcal{X} (たとえば \mathbb{R}^d) の上の確率変数の系列 $(Y_i)_{i \in \mathbb{Z}_{>0}}$ にたいして

$$N(A) := \sum_{i=1}^n I[Y_i \in A], \quad A \subset \mathcal{X},$$

とするとき $N(\cdot)$ を点過程と呼ぶ. A を任意に固定すると, Y_i が独立で同一分布法則に従うならば $N(A)$ は 2 項分布 $Bn(n, \rho(A))$, $\rho(A) = E(I[Y_i \in A]) = P\{Y_i \in A\}$ に従う. Borel 集合列 $A_1 \supset A_2 \supset \dots$ にたいして

$$n\rho(A_n) \rightarrow \Lambda(A_\infty), \quad A_n \rightarrow A_\infty,$$

ならば, $N(A_\infty)$ は Poisson 分布に従う.

任意の Borel 集合 A にたいして, $N(A)$ が平均値 $\Lambda(A)$ の Poisson に従い, 互いに交わらない Borel 集合 B_1, \dots, B_n , $n \in \mathbb{Z}_{>0}$ にたいして $(N(B_1), \dots, N(B_n))$ が独立のとき $N(\cdot)$ を Poisson 過程と呼ぶ. 非負測度 $\Lambda(A)$, $0 \leq \Lambda(A) \leq \infty$, は平均値測度と呼ばれる. 任意の Borel 集合 A にたいして

$$\Lambda(A) = \int_A \lambda(y) dy$$

であるとき, $\lambda(y)$ を Poisson 過程の強度と呼ぶ. $\lambda(y)$ が定数のとき Poisson 過程は一様であるという.

Rényi の定理. $\Lambda(A)$ が \mathbb{R}^d の上のノンアトミック測度で, 有界集合にたいして有限であるとする. \mathbb{R}^d の有限個の軸平行長方形の和集合 A にたいし

$$P\{N(A) = 0\} = \exp(-\Lambda(A))$$

ならば $N(\cdot)$ は Poisson 過程である. この確率を空白確率 (void probability) と呼ぶ.

定理 11. \mathbb{R}^d 上の点過程の系列 $N_n(\cdot)$ が Poisson 過程 $N(\cdot)$ に分布収束する, つまり任意の Borel 集合 $B_1, \dots, B_r \subset \mathbb{R}^d$ にたいして

$$(N_n(B_1), \dots, N_n(B_r)) \xrightarrow{d} (N(B_1), \dots, N(B_r)),$$

となるための十分条件は, 任意の半開長方形 $I = \prod_{i=1}^d (a_i, b_i]$ にたいして

$$\lim_{n \rightarrow \infty} EN_n(I) = EN(I),$$

および, 任意の半開長方形の任意個の和 $B = \bigcup_{k=1}^m I_k$ にたいして

$$\lim_{n \rightarrow \infty} P\{N_n(B) = 0\} = P\{N(B) = 0\}.$$

例. 1次元一様 Poisson 過程

$F \in MDA(H)$ の吸引係数を a_n, b_n とする. この条件は $\log(1-x) \sim -x, x \rightarrow 0$, から

$$\lim_{n \rightarrow \infty} n \log(1 - \bar{F}(a_n x + b_n)) = \log H(x),$$

あるいは

$$\lim_{n \rightarrow \infty} n \bar{F}(a_n x + b_n) = -\log H(x),$$

と同等であることを注意しておく. F の i.i.d. 系列 $(X_i)_{t \in \mathbb{Z}_{>0}}$ にたいして $\mathcal{X} = (0, 1)$ 上の点過程列を

$$N_n(A) = \sum_{i=1}^n I[Y_{i,n} \in A], \quad A \subset \mathcal{X}, \quad Y_{i,n} = \frac{i}{n+1} I[X_i > u_n(x)], \quad u_n(x) := a_n x + b_n,$$

によって定義すると, その極限 $N(\cdot)$ の平均測度は

$$\Lambda(A) = (t_2 - t_1)\lambda(x), \quad A = (t_1, t_2) \subset \mathcal{X}, \quad \lambda(x) = -\log H(x).$$

固定した閾値 x を超えた“時点” $t = i/n$ の極限は $(0, 1)$ 上の一様 Poisson 過程となっている.

例. 2次元非一様 Poisson 過程

同じように $F \in MDA(H)$ の i.i.d. 系列 $(X_i)_{t \in \mathbb{Z}_{>0}}$ にたいして $\mathcal{X} = (0, 1) \times (x, \infty)$ 上の点過程列を

$$N_n(A) = \sum_{i=1}^n I[(\frac{i}{n+1}, X_i) \in A], \quad A \subset \mathcal{X},$$

によって定義する. $u_n = a_n x + b_n$ とする.

$$\lim_{n \rightarrow \infty} P\{N_n((0, 1) \times (u_n, \infty)) = 0\} = \lim_{n \rightarrow \infty} P\{X_{\max} \leq u_n\} = \exp(-(1 + \gamma x)_+^{-1/\gamma}),$$

であるから極限 $N(\cdot)$ は Poisson 過程で $A = (t_1, t_2) \times (y, \infty)$ の平均測度は

$$\Lambda(A) = \int_{t_1}^{t_2} \int_y^\infty \lambda(v) dv dt = (t_2 - t_1)(1 - W_\gamma(y)),$$

となり, 水準超過モデルの別の導入が得られた. 強度関数は t によらず

$$\lambda(t, y) = \lambda(y) = \frac{d}{dy} W_\gamma(y) = (1 + \gamma x)_+^{-1/\gamma-1} = \frac{1}{1 + \gamma x} \bar{W}_\gamma(x), \quad \text{if } 1 + \gamma x > 0, \quad (14)$$

である. 直接点過程から出発して議論すると, ここでも生存関数の正則変動性が本質的である.

モデリング $F^n(a_n x + b_n) \approx G_\gamma(x)$ において, 吸引係数についての情報は通常得られないから, 未知の位置, 尺度パラメータを導入して,

$$F^n(x) \approx G_\gamma\left(\frac{x - \xi}{\eta}\right), \quad \text{or} \quad n \bar{F}(x) \approx -\log G_\gamma\left(\frac{x - \xi}{\eta}\right) = \bar{W}_\gamma\left(\frac{x - \xi}{\eta}\right),$$

と近似することになる. 同じ方針で閾値を超過するデータ (POT) の場合には,

$$\bar{F}(x) = \bar{F}(u)(\bar{F}(x)/\bar{F}(u)) \approx G_\gamma\left(\frac{u - \xi}{\eta}\right) \bar{W}_\gamma\left(\frac{x - \xi}{\eta}\right) / \bar{W}_\gamma\left(\frac{u - \xi}{\eta}\right) \quad (15)$$

$$= \exp(-\tau(u)) \left(1 + \gamma \frac{x - u}{a(u)}\right)^{-1/\gamma}, \quad \tau(u) := \bar{W}_\gamma\left(\frac{u - \xi}{\eta}\right), \quad a(u) := \eta + \gamma(u - \xi). \quad (16)$$

と近似する. 点過程に戻り, X_1, \dots, X_n, u が与えられ,

$$N_n(A) = \sum_{i=1}^n I[(\frac{i}{n+1}, X_i) \in A], \quad A \subset \mathcal{X} = (0, 1) \times (u, \infty)$$

で定義される点過程を強度関数

$$\lambda(t, x) = \lambda(x) = \frac{1}{\eta} \left(1 + \gamma \frac{x - \xi}{\eta} \right)_+^{-1/\gamma-1} \quad (17)$$

の Poisson 過程と仮定する。つまり平均測度を

$$\Lambda((t_1, t_2) \times (x, \infty)) = (t_2 - t_1) \overline{W}_\gamma \left(\frac{x - \xi}{\eta} \right),$$

と仮定する。(17) と (16) から (14) も参照

$$\lambda(u) = \frac{1}{\eta + \gamma(u - \xi)} \overline{W}_\gamma \left(\frac{u - \xi}{\eta} \right) = \frac{\tau(u)}{a(u)}$$

であることに注意すると強度関数の別表現が得られる:

$$\lambda(x) = \frac{\tau(u)}{a(u)} \frac{\lambda(x)}{\lambda(u)} = \frac{\tau(u)}{a(u)} \left(1 + \gamma \frac{x - u}{a(u)} \right)^{-1/\gamma-1} \quad (18)$$

X_1, \dots, X_n, u の超過値を $(\bar{X}_1, \dots, \bar{X}_{N_u})$, とする。 N_u は超過データ数である。超過値は一般パレート分布に従うとしてパラメータ $\gamma, a(u) = \eta + \gamma(u - \xi)$ を推定し, N_u から $\tau(u) = \overline{W}_\gamma \left(\frac{u - \xi}{\eta} \right)$ を推定し, 合わせて ξ, η を定めることができる。

あるいは直接 Poisson 過程の尤度

$$\exp(-\tau(u)) \prod_{j=1}^{N_u} \lambda(\bar{X}_j)$$

を数値的に最大化することもできる。水準超過点過程に基づく推測理論は Davison and Smith (1990) で詳しく議論されている。

4.5.2 自動型点過程

点過程によるモデリングの利点は、より柔軟な枠組みを構築する可能性を含んでいることである。特に時系列データで、極値理論を適用するために i.i.d. 系列を近似的に得る“白色雑音化”を避けられる。簡単のため 1 次元時系列に限り、閾値を越える時点だけに注目する 1 次元 Poisson 過程だけを考える。

データ X_1, \dots, X_n , と閾値 u が与えられたとき, 上のように $N_u = \#\{(i, X_i) : 1 \leq i \leq n, X_i > u\}$ とし, データを

$$\{(T_j, \bar{X}_j) : 1 \leq j \leq N_u\}, \quad T_j : \bar{X}_j = X_{T_j}.$$

と書き換える。 $\mathcal{X} = (0, n]$ の上の点過程

$$N_n(A) = \sum_{i=1}^n I[Y_i \in A], \quad A \subset \mathcal{X}, \quad Y_i := iI[X_i > u],$$

が Poisson 過程であると仮定し, その強度関数が過去の閾値超過に依存して

$$\lambda^*(t) = \tau + \psi \sum_{j: 0 < T_j < t} h(t - T_j, \bar{X}_j - u),$$

と変わる動的なモデルを考える。関数 h としては

$$h(s, x) = \exp(\delta x - \gamma s), \quad \delta, \gamma > 0,$$

$$h(s, x) = \exp(\delta x)(s + \gamma)^{-\rho+1}, \quad \delta, \gamma, \rho > 0.$$

などが考えられる。尤度関数は

$$\exp\left(-\int_0^n \lambda^*(s) ds\right) \prod_{j=1}^{N_u} \lambda^*(T_j)$$

となる。

これは地震学で用いられているモデルがファイナンスで利用できるのではないかという発想 (McNeil et al., 2005) であるが, 解析例はない。

4.6 補い

4.6.1 広義逆関数と確率点関数

非減少関数 $\varphi(x)$ の (左連続) 広義逆関数を次式で定義する.

$$\varphi^-(y) = \inf\{x : \varphi(x) \geq y\}$$

- φ が連続ならば, $\varphi(\varphi^-(y)) = y$, しかし $\varphi^-(\varphi(x)) \leq x$.
- φ が単調連続ならば $\varphi^-(y)$ は逆関数 $\varphi^{-1}(y)$ に等しい.
- $\varphi_1, \varphi_2, \dots$ を非減少関数の系列, φ も非減少関数であり, $x \in (a, b)$ で $\varphi(x)$ が連続で

$$\lim_{n \rightarrow \infty} \varphi_n(x) = \varphi(x), \quad x \in (a, b),$$

であるとする. すると $y \in (\varphi(a), \varphi(b))$ で $\varphi^-(x)$ が連続ならば

$$\lim_{n \rightarrow \infty} \varphi_n^-(y) = \varphi^-(y), \quad y \in (\varphi(a), \varphi(b)).$$

de Haan は記号 $\varphi^-(x)$ を用いている. また合成関数, 数式にたいしても $(1/(1 - \varphi(x)))^-$ のように使っている.

確率点関数 分布関数 $F(x) = P\{X \leq x\}$ の広義逆関数 $F^-(u) = \inf\{x : F(x) \geq u\}$ を F の (X の) 確率点関数 (quantile function) と定義する.

- U を区間 $(0,1)$ 上の一様確率変数 (乱数) とすると, $F^-(U)$ は X と同じ分布に従う.
- 確率変数 X, Y の分布関数をそれぞれ F, G とする. F, G が連続ならば,

$$G^-(F(X)) \stackrel{D}{=} Y.$$

4.6.2 正則変動関数

無限大の近傍における解析学の主要な道具が正則変動関数である. これは無限大における微分の性質とみなすことができる.

参照. 志村隆彰 (2006) 正則変動性と確率論, 東京大学大学院経済研究科, 講義ノート; de Haan and Ferreira (2006) Appendix B.

ある実関数 g の差分商

$$(g(y+h) - g(y))/h, \quad h \neq 0,$$

にたいして y を固定して $h \rightarrow 0$ の極限ではなく, h を固定して $y \rightarrow \infty$ の極限を考える.

もしも任意の h にたいして極限が存在すれば, 極限は h に依存しない. このとき, ある $g_0(y)$ 関数を用いて

$$g(y) = g_0(y) + o(1), \quad y \rightarrow \infty,$$

と表わせ, $g_0(y)$ は微分可能で,

$$\lim_{y \rightarrow \infty} g'_0(y) = \lim_{y \rightarrow \infty} \frac{g(y+h) - g(y)}{h},$$

を満たす. この極限が存在するならば, それを α とする. $f(t) = \exp g(\log t) \Leftrightarrow g(y) = \log(f(\exp(y)))$ と変換すると, f は次の定義を満たす.

定義. $f: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ が

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha, \quad \forall x \in \mathbb{R}_{>0}, \quad \alpha \in \mathbb{R}, \quad (19)$$

を満たすとき, ∞ で正則変動関数 (regularly varying function) であると言い, $f \in \mathbb{R}_\alpha$ で表わす. $f \in \mathbb{R}_0$ のとき, f は ∞ で緩慢変動関数 (slowly varying function) であると言う.

\mathbb{R}_0 に属する諸関数.

- $(\log x)_+^\alpha$, $\alpha \in \mathbb{R}$. $()_+$ は不要.
- $\prod_{k=1}^n (\log_k x)^{\alpha_k}$; $\log_{k+1} x := \log(\log_k x)$, $\log_1 x = \log x$, $\alpha_k \in \mathbb{R}$, $1 \leq k \leq n$.
- $\exp((\log x)^\beta)$, $\beta < 1$.
- $\forall \ell(x) : \lim_{x \rightarrow \infty} \ell(x) = c \in \mathbb{R}_{>0}$.

\mathbf{R}_α の基本性質.

- \mathbf{R}_0 は加算, 乗算, 除算に関して閉じている.
- $\ell \in \mathbf{R}_0 \implies \ell^\alpha \in \mathbf{R}_0$, $\alpha \in \mathbb{R}$.
- $f \in \mathbf{R}_\rho \implies 1/f \in \mathbf{R}_{-\rho}$, $\rho \in \mathbb{R}$.
- $f \in \mathbf{R}_\rho \implies f^- \in \mathbf{R}_{1/\rho}$, $\rho \in \mathbb{R}_{>0}$.

表現定理 (Karamata, 1930). $f \in \mathbf{R}_\alpha$ ならば関数 $a : \mathbb{R}_{>0} \rightarrow \mathbb{R}$, $c : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$;

$$\lim_{t \rightarrow \infty} a(t) = \alpha, \quad \lim_{t \rightarrow \infty} c(t) = c_0, \quad (0 < c_0 < \infty)$$

および $t_0 \in \mathbb{R}_{>0}$ が存在して

$$f(t) = c(t) \exp \int_{t_0}^t \frac{a(s)}{s} ds, \quad t > t_0,$$

と表わせる. 逆も成り立つ.

The de Bruyn conjugate. $\ell \in \mathbf{R}_0$ ならば, 下記の条件を満たす関数 $\tilde{\ell} \in \mathbf{R}_0$ が, 漸近的に一意に存在する.

$$\ell(x)\tilde{\ell}(x\ell(x)) \rightarrow 1, \quad x \uparrow \infty.$$

4.6.3 最大値安定性より 3 分布の導出の概略

この種の関数方程式の解法については, Aczél, J. and Dhombres, J. (1989) *Functional Equations in Several Variables*, Cambridge University Press. を参照.

$$H^n(a_n x + b_n) = H(x), \quad H^s(A(s)x + B(s)) = H(x) \quad (\text{連続化})$$

$s \log H(A(s)x + B(s)) = \log H(x)$, $-\log(-\log H(A(s)x + B(s))) - \log s = -\log(-\log H(x)) =: y$
 $-\log(-\log H(x))$ の逆関数を U とする.

$$x = H^{-1}(\exp(-\exp(-y))) =: U(y)$$

$$A(s)x + B(s) = U(y + \log s), \quad U(y + \log s) - B(s) = A(s)U(y)$$

$$y = 0 \implies U(\log s) - B(s) = A(s)U(0), \quad U(y + \log s) - U(\log s) = A(s)(U(y) - U(0))$$

$$\tilde{U}(y) := U(y) - U(0), \quad \tilde{A}(s) := A(e^s)$$

$$\tilde{U}(y+x) - \tilde{U}(x) = \tilde{A}(x)\tilde{U}(y), \quad \tilde{U}(y+x) - \tilde{U}(y) = \tilde{A}(y)\tilde{U}(x)$$

$$\tilde{U}(x+y) = \tilde{A}(x)\tilde{U}(y) + \tilde{U}(x) = \tilde{A}(y)\tilde{U}(x) + \tilde{U}(y), \quad \tilde{U}(x)(1 - \tilde{A}(y)) = \tilde{U}(y)(1 - \tilde{A}(x))$$

(a) $\tilde{A} \equiv 1 \implies \tilde{U}(x+y) = \tilde{U}(x) + \tilde{U}(y)$, $\tilde{U}(y) = ay$, $H(U(0) + ay) = \exp(-\exp(-y))$.

(b) $\exists x$ s.t. $\tilde{A}(x) \neq 1 \implies \tilde{A}(x) \neq 1, \forall x$, なぜならば, $\exists y : \tilde{A}(y) = 1, \tilde{U}(x)(1 - \tilde{A}(y)) = 0$, $\tilde{U}(y+x) - \tilde{U}(y) = 0, \forall y$ である. したがって

$$\frac{\tilde{U}(x)}{1 - \tilde{A}(x)} = \frac{\tilde{U}(y)}{1 - \tilde{A}(y)} = c \quad \text{const.} \quad \tilde{U}(x) = c(1 - \tilde{A}(x))$$

$$\tilde{A}(x)\tilde{U}(y) = \tilde{U}(x+y) - \tilde{U}(x) = c(\tilde{A}(x) - \tilde{A}(x+y)).$$

$$\tilde{A}(x+y) = \tilde{A}(x) \left(1 - \frac{1}{c}\tilde{U}(y)\right) = \tilde{A}(x)\tilde{A}(y), \quad \tilde{A}(x) = e^{\gamma x}, A(s) = s^\gamma, G_\gamma.$$

5 多変量極値理論

5.1 多変量極値 (まえがき)

極値観測データは、多くの統計データと同様に、根源的に多変量である。つまり、異なる時間、異なる地点において測定されたデータは、“独立に同一条件の下で測定された”保証がなく、時間的、空間的な従属性をもっている。たとえば、気象、自然災害、環境などの調査・研究ではこの従属性を無視できない。

1 変量確率標本 (独立同一分布) に関する極値理論は美しく汎用的であるが、多変量ではかなり複雑となる。2 変量の場合には視覚化可能であるため扱いやすいが、3 変量以上となると低次元でも種々の困難が生じる。そのために多変量極値理論は未完成で多くの挑戦が続いている。以下はそのような活動への入門である。

5.2 多変量極値分布 (区分最大値)

$(X_n = (X_{1n}, \dots, X_{dn}))_{n=1}^{\infty}$ を分布関数 F をもつ d 変量確率変数の i.i.d. 列とし、成分ごとの最大値を

$$M_n := (\max_{1 \leq i \leq n} X_{1i}, \dots, \max_{1 \leq i \leq n} X_{di})$$

とする。その同時分布関数は $P\{M_n \leq \mathbf{x}\} = F^n(\mathbf{x})$ である。

適当なベクトル列 $(\mathbf{a}_n > \mathbf{0})_{n=1}^{\infty}$, $(\mathbf{b}_n)_{n=1}^{\infty}$ により各成分を基準化したとき、非退化分布 H の連続点 \mathbf{x} において

$$\lim_{n \rightarrow \infty} F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) = H(\mathbf{x})$$

のとき、 H を多変量極値分布と呼び、 F は H の最大値吸引領域に属すると言い、 $F \in MDA(H)$ と表す。 $(\mathbf{a}_n, \mathbf{b}_n)_{n=1}^{\infty}$ を吸引係数ベクトル列 (あるいは規準化ベクトル列) と呼ぶ。 F の周辺分布関数 $F_J(\mathbf{x}_J)$, $\mathbf{x}_J = \{x_i : i \in J\}$, $J \subset \{1, \dots, d\}$, は $F_J \in MDA(H_J)$ を満たす。 H_J は、 H の周辺分布関数で多変量極値分布である。

特別な場合として H の周辺分布 $H_1(x) = H(x, \infty, \dots, \infty)$, \dots , $H_d(x) = H(\infty, \dots, \infty, x)$ のそれぞれは 1 変量極値分布に従い、 F の周辺分布を F_1, \dots, F_d とすると $F_i \in MDA(H_i)$ である。吸引係数列もそのまま利用できる。1 変量極値分布は一般極値分布に限られるので、 $H(\mathbf{x})$ の周辺分布によらない従属性が問題である。1 変量極値分布の型は確率変数の単調連続変換 (確率積分変換) により自由に変換できるから、 $H(\mathbf{x})$ の周辺分布を固定し、さらにそれに応じて F_i も (既知であるとして) 同じ分布に変換して、議論することができる。

2 つの周辺分布が代表的である。1 つは一様分布、もう 1 つは標準 Fréchet 分布である。これらについて順に述べる。いずれにしる多変量極値分布の全体は多様であり、1 変量極値分布のように有限個のパラメータですべてを表せるものではない。できるだけ汎用で、計算しやすく、推測可能なモデルの追求が続いている現状である。

5.3 極値接合分布関数

5.3.1 準備: 接合分布関数 (copula)

参照 McNeil et al. (2005) Chap. 5; Nelsen, R. B. (2006), Kolev (2006), 塚原英敦 (2004). “copula” は Sklar, 1959, の命名である。塚原は原義から“繋合関数”を提案したが、誰も読めないのが変更した。

定義 1 (接合分布関数 (copula)). d 変数接合分布関数、あるいは単に接合関数 $C(\mathbf{u}) = C(u_1, \dots, u_d)$ とは単位立方体 $[0, 1]^d$ 上の確率分布関数であって、その 1 変数周辺関数がすべて $[0, 1]$ 一様分布のものである。

$C(\mathbf{u})$ は各成分 u_i の非減少関数であり

$$C(1, \dots, 1, u_i, 1, \dots, 1) = u_i, \quad 0 \leq u_i \leq 1, \quad 1 \leq i \leq d,$$

を満たすが、それだけでは不十分であり、軸に平行な任意の直方体 $\prod_{i=1}^d [a_i, b_i]$, $0 \leq a_i < b_i \leq 1$, $1 \leq i \leq d$, に入る確率が非負の条件が必要である。つまり,

$$\sum_{i_1=1}^2 \cdots \sum_{i_d=1}^2 (-1)^{i_1+\cdots+i_d} C(u_{i_1}, \dots, u_{i_d}) \geq 0,$$

$$i_j \in \{1, 2\}, \quad u_{j1} = a_j, \quad u_{j2} = b_j, \quad 1 \leq j \leq d.$$

$C(\mathbf{u})$ が d 変数接合分布関数であればすべての k 変数周辺分布, $2 \leq k \leq d$, もまた $[0, 1]^k$ 上の接合分布関数である。

定理 12 (Sklar, 1959). F が周辺分布関数 F_1, \dots, F_d をもつ分布関数であると, ある接合分布関数 $C(\mathbf{u})$ が存在して

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

と表すことができる。周辺分布関数が連続であればこの表現は一意的である。

逆に C をある d 変数接合分布関数, F_1, \dots, F_d を 1 変数分布関数とすると $C(F_1, \dots, F_d)$ は F_1, \dots, F_d を周辺分布とする d 変数同時分布関数である。

分布関数が連続でないときにはいろいろと不都合なことがある。以下ここでは連続な場合だけを扱う。そうすると接合分布関数を次のように再定義できる。

定義 2 (F の接合分布関数). 確率変数 $X = (X_1, \dots, X_d)$ の同時分布関数を F , 周辺分布関数を F_1, \dots, F_d とするとき確率変数

$$(F_1(X_1), \dots, F_d(X_d))$$

の分布関数を F の接合分布関数と呼ぶ。

つまり F の接合分布関数は分布関数の, 周辺確率表示にしかすぎない:

$$C(\mathbf{u}; F) := F(F_1^-(u_1), \dots, F_d^-(u_d)).$$

定理 13. $X = (X_1, \dots, X_d)$ を連続な周辺分布関数, 連続な接合分布関数 C をもつ確率変数とする。 $\varphi_1, \dots, \varphi_d$ を任意の単調増加関数とすると $(\varphi_1(X_1), \dots, \varphi_d(X_d))$ の接合分布関数は C で変わらない。

•

$$C_0(\mathbf{u}) := \prod_{i=1}^d u_i,$$

は任意の独立連続分布関数の接合分布関数である。

•

$$C_+(\mathbf{u}) := \min(u_1, \dots, u_d),$$

\mathbb{R}^d の“上昇曲線”: $(f_1(t), \dots, f_d(t))$; f_i 非減少) の上に退化した完全正従属 (perfectly positively dependent, comonotone) 確率分布の接合分布である。

•

$$C_-(u_1, u_2) := \max(u_1 + u_2 - 1, 0),$$

\mathbb{R}^2 の“下降曲線” の上に退化した完全負従属 (perfectly negatively dependent, countermonotone) 確率分布の接合分布である。

- 相関行列 Λ の標準多変量正規分布関数 $\Phi_\Lambda(\mathbf{x})$ の接合分布関数は、 Φ を標準 1 変量正規分布の分布関数をとすると、

$$C_\Lambda(\mathbf{u}) = \Phi_\Lambda(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

これから、一般の正規分布、対数正規分布の分布関数は

$$\Phi_\Sigma(\mathbf{x}) = C_\Lambda(\Phi(\sigma_1 x_1 + \mu_1), \dots, \Phi(\sigma_d x_d + \mu_d))$$

これに対応する対数正規分布の接合分布関数も同じものである。

定理 14 (Fréchet-Hoeffding 限界). すべての d 変数接合分布関数は不等式

$$\max\left\{\sum_{i=1}^d u_i + 1 - d, 0\right\} \leq C(\mathbf{u}) \leq \min(u_1, \dots, u_d)$$

を満たす。

より一般に、周辺分布関数 F_1, \dots, F_d をもつ同時分布 F に対して

$$\max\left\{\sum_{i=1}^d F_i(x_i) + 1 - d, 0\right\} \leq F(\mathbf{x}) \leq \min(F_1(x_1), \dots, F_d(x_d))$$

が成り立つ。

Fréchet-Hoeffding 上限は接合分布関数 C_+ で下限は $d=2$ の場合にだけ、接合分布関数 C_- となる。

これまでの議論における下側確率をすべて上側確率に変えることにより、生存接合分布関数を定義できる。

定義 3 (F の生存接合分布関数). 確率変数 $X = (X_1, \dots, X_d)$ の同時生存関数を \bar{F} 、周辺生存関数を $\bar{F}_1, \dots, \bar{F}_d$ とするとき確率変数

$$(\bar{F}_1(X_1), \dots, \bar{F}_d(X_d))$$

の分布関数を F または \bar{F} の生存接合分布関数 (*survival copula*) と呼ぶ。

生存接合分布関数は分布関数である。接合分布関数の生存関数と区別しなければならない。 X の接合分布関数が C のとき、対応する生存分布関数を \hat{C} と表す。一つの定義は次の通りである。

$$\hat{C}(\mathbf{u}; F) := \bar{F}(F_1^-(1 - u_1), \dots, F_d^-(1 - u_d)).$$

たとえば $d=2$ のとき、

$$\begin{aligned} \hat{C}(u_1, u_2) &= P\{\bar{F}_1(X_1) \leq u_1, \bar{F}_d(X_d) \leq u_2\} = P\{X_1 > \bar{F}_1^-(1 - u_1), X_2 > \bar{F}_2^-(1 - u_2)\} \\ &= 1 - (1 - u_1) - (1 - u_2) + F(F_1^-(1 - u_1), F_2^-(1 - u_2)) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \end{aligned}$$

注意 X の接合分布関数が C のとき、 $-X = (-X_1, \dots, -X_d)$ の接合分布関数は \hat{C} である。

Archimedean Copula

$\varphi: [0, 1] \rightarrow [0, \infty]$ が単調減少関数で $\varphi(1) = 0$ であるとする。 φ^{-1} をその逆関数とする。ただし定義域を拡大して、

$$\varphi^{-1}(t) = 0, \quad \varphi(0) \leq t \leq \infty,$$

とする。 $(\varphi(0) = \infty$ のとき φ は strict であるという。)

定理 15. $C(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$ が 2 変数接合分布関数となる必要十分条件は φ が凸となることである。

この接合関数を “Archimedean Copula” と呼ぶ。これを d 変数 ($d > 2$) に拡張するためには φ が strict で, completely monotone (Laplace 変換表示可能) でなければならない。生成できるのは正のを持つ接合分布関数に限られる。

協調性順序

接合分布関数の間の自然な半順序関係を

$$C_1 \prec C_2 \iff C_1(u) \leq C_2(u), \quad \forall u \in [0, 1]^k,$$

で定義する。この半順序関係を協調性順序 (concordance order) と呼ぶ

- F が正の協調性をもつということは, F の接合分布 C が

$$C_0 \prec C$$

を満たすことである。これは周辺が等しい多変量確率分布の “正の従属性” の定義として強いものである。

- $d = 2$ のときは逆に

$$C \prec C_0$$

によって C の協調性 (従属性) を表すことができる。

協調性尺度 (2 変数接合分布関数の)

確率ベクトルの協調性を調べる。2つの確率ベクトル $(X_1, Y_1), (X_2, Y_2)$ が連続で共通の周辺分布 F, G をもち, それぞれの同時分布 H_1, H_2 が異なるものとする。したがって両者の接合分布関数 C_1, C_2 が異なる。

$$Q := P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\},$$

とする。これは H_1, H_2 間の協調性を表す尺度である。

$$Q = Q(C_1, C_2) = 4 \int_0^1 \int_0^1 C_2(u, v) dC_1(u, v) - 1, \quad Q(C_1, C_2) = Q(C_2, C_1),$$

であり,

$$\begin{aligned} C_1(u, v) \leq C'_1(u, v), \quad C_2(u, v) \leq C'_2(u, v), \quad \forall u, v \in [0, 1] \\ \implies Q(C_1, C_2) \leq Q(C'_1, C'_2). \end{aligned}$$

H_i そのものの協調性が $Q(C_i, C_i)$ で表せるが, これは次のような意味をもっている。Kendall の順位相関係数 τ は 2 変数確率標本において相関係数 (K. Pearson のモーメント相関係数, 線形相関係数) を順位統計量としたノンパラメトリック統計量である。その “母集団分布版” としては上の Q を用いると次のように定義できる。

$$\tau(X, Y) = \tau_C := 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1,$$

あるいは独立な $(0,1)$ 一様確率変数 U, V を用いると

$$\tau_C = 4E(C(U, V)) - 1 = 4 \int_0^1 t dK_C(t) - 1 = 3 - 4 \int_0^1 K_C(t) dt.$$

ただし $K_C(t)$ は $C(U, V)$ の分布関数であり, 接合分布関数 C の Kendall 分布と呼ばれている。

もしも C が関数 φ で定義された 2 変数 Archimedean copula ならば,

$$K_C(t) = t - \varphi(t)/\varphi'(t),$$

である。

5.3.2 極値接合関数

$H(\mathbf{x})$ が極値分布関数であれば、それからの確率標本の、成分ごとの最大値は、基準化すれば、 $H(\mathbf{x})$ に従う。この意味で H は “最大値安定” である。

$$H^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) = H(\mathbf{x}), \quad H_i^n(a_i x + b_i) = H_i(x),$$

両辺の接合分布関数は、周辺分布関数の基準化に依存しないから、

$$C^n(\mathbf{u}^{1/n}) = C(\mathbf{u}), \quad \text{or} \quad C^r(\mathbf{u}) = C(\mathbf{u}^r), \quad \forall r > 0.$$

最後の条件を満たす “最大値安定” 接合分布関数を極値接合分布関数 (extreme value copula) と呼ぶ。極値接合分布関数は “最大値無限分解可能” である。

極値接合分布関数の例

1. 独立接合分布関数 C_0 , 完全正従属接合分布関数 C_+ は極値接合分布関数である。
2. ロジスティック接合分布関数

$$C_\theta(\mathbf{u}) = \exp\left(-\left(\sum_i (-\log u_i)^\theta\right)^{1/\theta}\right), \quad 1 \leq \theta < \infty$$

3. Galambos 接合分布関数

$$C_\theta(\mathbf{u}) = \prod_{i=1}^d u_i \exp\left(-\left(\sum_i (-\alpha_i \log u_i)^{-\theta}\right)^{-1/\theta}\right), \quad 0 \leq \alpha_i \leq 1, \quad 0 < \theta < \infty.$$

4. $C(\mathbf{u})$ を置換不変 (exchangeable) 接合分布関数とする。

$$C(\mathbf{u}) = \mathbf{u}^{1-\alpha} C(\mathbf{u}^\alpha), \quad 0 \leq \alpha_i \leq 1.$$

5. Pickands 表現 $d = 2$ のとき,

$$C(\mathbf{u}) = \exp\left(\log(u_1 u_2) A\left(\frac{\log u_2}{\log(u_1 u_2)}\right)\right),$$

A は次節で詳述する Pickands 従属性関数である。

5.3.3 極値接合関数と単純極値分布との関係

極値接合関数と、次節で述べる単純多変量極値分布との関係を、予め表の形にまとめておく。各周辺確率変数の確率積分変換によって相互に変換できるから、分布関数の変数変換により、互いに他の表現を導けるが、結果は一般に簡単ではない。Mikosch (2006) は極値接合関数への期待が大きすぎると警告している。Mikosch, T. (2006). Copulas: Tales and facts, *Extremes*, 9, 3-20.

表 4: 極値接合関数と単純極値分布

	多変量極値接合分布関数 multivariate extreme copula	単純多変量極値分布 SMvEV
分布関数の特徴	$C(\mathbf{u}) = C^r(\mathbf{u}^{1/r}), r > 0,$	$G(\mathbf{x}) = G^t(t\mathbf{x}), t > 0,$ or $G(\mathbf{x}) = \exp(-V(\mathbf{x})), \quad V(\mathbf{x}) = tV(t\mathbf{x}), t > 0,$
周辺変数の変換	$U = \exp(-1/X) \leftrightarrow X = -1/\log U$	

5.4 単純多変量極値分布

次に周辺分布として、標準 Fréchet 分布 $F(z) = \exp(-1/z)$, $z > 0$, を選ぶ. その性質について、付録を参照. 次の定理が基本的である. これは Pickands (1981), de Haan and Resnick (1977) を Coles and Tawn (1991) が改めたものである.

$(X_{1n}, \dots, X_{dn})_{n=1}^{\infty}$ が i.i.d. 確率変数列で、すべての 1 変量周辺分布が標準 Fréchet 分布 に従うとする.

$$\mathbf{M}_n^* = (M_{1n}^*, \dots, M_{dn}^*) = \left(\max_{1 \leq i \leq n} X_{1i}/n, \dots, \max_{1 \leq i \leq n} X_{di}/n \right),$$

とする. 吸引係数 $a_{j,n} = n$, $b_{j,n} = 0$, は、Fréchet 分布 が自分自身に収束するときの吸引係数である.

$$P\{\mathbf{M}_n^* \leq \mathbf{x}\} \xrightarrow{d} G(\mathbf{x}),$$

で G の周辺分布は標準 Fréchet 分布 である. 最大値安定性から

$$G^t(t\mathbf{x}) = G(\mathbf{x}), \quad t > 0,$$

である. 標準 Fréchet 分布を周辺分布とする多変量極値分布を単純多変量極値分布 (simple multivariate extreme value distribution) と呼ぶことがある. これを SMvEV と表す.

定理 16 (指数測度とスペクトル測度). 単純多変量極値分布を

$$G(\mathbf{x}) = \exp(-V(\mathbf{x})), \quad \mathbf{x} \in \mathbb{R}_{>0}^d, \quad (20)$$

と表すと,

$$V(x_1, \dots, x_d) = \int_{S_d} \max_{1 \leq j \leq d} (w_j/x_j) dH(\mathbf{w}), \quad (21)$$

である. S_d は $d-1$ 次元単位単体

$$S_d = \{\mathbf{w} : w_j \geq 0, \sum_{j=1}^d w_j = 1\},$$

H は S_d 上の非負測度で,

$$\int_{S_d} w_j dH(\mathbf{w}) = 1, \quad j = 1, \dots, d, \quad (22)$$

を満たすものである. したがって $H(S_d) = d$.

注意

1. V (および後述の μ) を指数測度 (exponent measure), H をスペクトル測度 (spectral measure) と呼ぶ後者は言わば極座標表示で、より一般の座標系で定義できる.
2. $d^{-1}H$ は S_d 上の確率分布である.
3. (22) の制約条件は、たとえば $x_2, \dots, x_d \rightarrow \infty$ とすると $V(x_1, \infty, \dots, \infty) = \int_{S_d} (w_1/x_1) dH(\mathbf{w})$ となるから,

$$G(x_1, \infty, \dots, \infty) = \exp(-1/x_1)$$

から必要となる.

4. V , (21), は $V(x_1/a, \dots, x_d/a) = aV(x_1, \dots, x_d)$ を満たすから, G , (20), は $G^n(n\mathbf{x}) = G(\mathbf{x})$, $n = 2, 3, \dots$ を満たす.
5. (21) 式の積分の意味は次の通りである.

$$w_j/x_j = t, \quad j = 1, \dots, d,$$

を満たす $\mathbf{w} = (w_1, \dots, w_d)$ は, $\sum w_j = t \sum x_j = 1$ の解であるから, w_j に関する制約は領域

$$\left\{ \mathbf{w} : w_j \geq \frac{d}{\sum x_j} x_j, i \neq j \right\}, \quad j = 1, \dots, d,$$

に関する積分となる. たとえば w_1 に対応する領域は $\mathbf{w} = (1, 0, \dots, 0)$ を頂点とする凸多角形 (凸多面体, polyhedron) である. 与えられた \mathbf{x} にたいして \mathbf{x} に依存する領域での積分を行うことになる. 別紙図多変量極値分布参照.

6. 付録 11.2 で超単純多変量極値分布について説明する. これは(21) の積分が和で表わせる場合である.

5.4.1 定理の説明

“指数測度” V の表現は次のようにして求められる. 最大値安定性から,

$$tV(tx_1, \dots, tx_d) = V(x_1, \dots, x_d), \quad t > 0,$$

であるから G を Poisson 過程とみなしたときの平均測度 ($\mathbb{R}_{\geq 0}^d$ 上の正測度) を μ とする. $\mu((t, \infty) \times B), B \subset S_d$, より,

$$\begin{cases} r = x_1 + \dots + x_d, & w_i = x_i/r, \\ i = 1, \dots, d-1; w_d = 1 - (w_1 + \dots + w_{d-1}), \end{cases}$$

とおくと

$$\mu(dr \times d\mathbf{w}) = \frac{dr}{r^2} dH(\mathbf{w}),$$

と表せる.

$$A = (0, x_1] \times \dots \times (0, x_d], \quad A^c = \mathbb{R}_{>0}^d \setminus A,$$

とすると

$$\begin{aligned} \mu(A^c) &= \int_{A^c} \frac{dr}{r^2} dH(\mathbf{w}) = \int_{S_d} \int_{r=\min(x_j/w_j)}^{\infty} \frac{dr}{r^2} dH(\mathbf{w}) \\ &= \int_{S_d} \max_{1 \leq j \leq d} (w_j/x_j) dH(\mathbf{w}). \end{aligned}$$

□

5.4.2 指数測度 (平均測度)

$[0, 1] \times \mathbb{R}_{\geq 0}^d$ の上のポアソン過程 (Z_k, \mathbf{X}_k) , $k = 1, 2, \dots$ の強度関数を $dz \times d\mu$ とし

$$\mu(\mathbf{y} \in \mathbb{R}_{>0}^d : \|\mathbf{y}\| > r; \mathbf{y}/\|\mathbf{y}\| \in B) = \frac{H(B)}{r}, \quad B \subseteq S_d = \{\mathbf{y} \in \mathbb{R}_{>0}^d, \|\mathbf{y}\| = 1\},$$

とする. ただし $H(B)$ は S_d の上の有限測度で

$$\int_{S_d} w_j dH(\mathbf{w}) = 1, \quad j = 1, \dots, d,$$

を満たすもの, $\|\cdot\|$ は任意のノルム, とする. すると “空白確率 (void probability)” の計算から

$$G(\mathbf{x}) = \exp(-\mu((0, \mathbf{x}]^c)).$$

参照. 付録. 多変数正規変動関数.

5.4.3 関連定理

一般に $G_i, i = 1, \dots, k$, が任意の独立な SMvEV に従うならば

$$G = \prod_{i=1}^k G_i^{\alpha_i}, \quad \alpha_i > 0, \quad \sum_{i=1}^k \alpha_i = 1,$$

も SMvEV である。このとき、 G の指数測度は G_i の指数測度の α_i の重みとする混合である。

G_i のスペクトル測度を H_i とすると G の指数測度は

$$\begin{aligned} V(x) &= \int_{S_d} \sum_{i=1}^k \alpha_i \max_{1 \leq j \leq d} \frac{w_j}{x_j} dH_i(w) \\ &= \int_{S_d} \max_{1 \leq j \leq d} \frac{w_j}{x_j} dH(w), \quad H = \sum_{i=1}^k \alpha_i H_i, \end{aligned}$$

である。したがって G のスペクトル測度も G_i のスペクトル測度の α_i の重みとする混合である。これらから次の定理が得られる。

定理 17. $Y = (Y_1, \dots, Y_d)$ が SMvEV である必要十分条件は

$$\max(v_1 Y_1, \dots, v_d Y_d), \quad \forall (v_1, \dots, v_d) \in S_d,$$

が標準 Fréchet 分布に従うことである。

- 一変量の場合の指数測度は累積逆ハザード関数 (cumulative reverse hazard function) である。
- 任意の 1 変数分布関数 F_1 にたいして $F_1^\alpha, \forall \alpha > 0$, も分布関数となるが、 d 変数分布関数 F_d の場合、一般には $\alpha \geq d-1$ を満たすすべての α にたいして F_d^α が分布関数となる。

$F_d^\alpha, \forall \alpha > 0$, が分布関数であるとき、 F_d は最大無限分解可能 (max infinitely divisible, max-id) という。 F_d が max-id であれば、任意の正整数 n にたいして $F_d^{1/n}$ からの大きさ n の確率標本 $(X_{1i}, \dots, X_{di})_{i=1}^n$ を用いると

$$\left(\max_{1 \leq i \leq n} X_{1i}, \dots, \max_{1 \leq i \leq n} X_{di} \right) \sim F_d$$

となる。すべての SMvEV は max-id である。

H を d 変量 ($d = 1, 2, \dots$) の確率分布とする。任意の正整数 n にたいして確率分布 H_n が存在し、 H_n の大きさ n の確率標本 (Z_1, \dots, Z_n) を用いると $\sum_{i=1}^n Z_i \sim H$ となるとき、 H を無限分解可能 (infinite divisible) という。max-id は、これと類似の概念である。

5.4.4 独立性と従属性

SMvEV が完全に従属であるとする、確率分布は直線 $x_1 = \dots = x_d$ 上の 1 次元分布に退化し、分布関数は

$$G(x) = \exp \left(-1 / \left(\bigwedge_{j=1}^d x_j \right) \right), \quad x \in \mathbb{R}_{\geq 0}^d,$$

となる。 H は単位単体の中心 $w_1 = \dots = w_d = 1/d$ に点測度 (質量) d をもつ測度となる。

単位単体上の確率分布はすべて負の従属性を持っている。ここで議論しているのはあくまでも積分(21)を媒介とする $G(x)$ の従属性であり、測度 H の特徴とく違いがある。参考のために、別紙図多変量極値分布に後述のロジスティック分布

$$G(x_1, x_2) = \exp \left(- \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} \right)^{1/2} \right)$$

と独立 Fréchet 分布からの標本の散布図がある。標本の大きさは 2000 である。図の範囲から外に数点が外れている。視覚的に見やすい図ではない。対数軸を選ぶ、つまり周辺分布を Gumbel 分布としたときの散布図も別紙図多変量極値分布にある。

定理 18 (高橋倫也). ある $G \sim \text{SMvEV}$ の成分が対ごとに独立であれば, 独立である.
さらに, ある $y \in \mathbb{R}^d$ が存在して

$$G_{ij}(y_i, y_j) = G_i(y_i)G_j(y_j), \quad 1 \leq i < j \leq d, \quad 0 < G_j(y_j) < 1,$$

であれば G は独立である.

実際, 周辺分布が等しい多変量極値分布 G が独立であるとし, $F \in \text{MDA}(G)$ の周辺分布も等しいとすると, F に関する次の条件は同等である. 最後の条件は次節で述べる裾の漸近独立性である.

$$\lim_{n \rightarrow \infty} F^n(a_n \mathbf{x} + b_n \mathbf{1}) = \prod_{i=1}^d G_1(x_i).$$

$$\lim_{n \rightarrow \infty} F_{k,l}^n(a_n x_k + b_n, a_n x_l + b_n) = G_1(x_k)G_1(x_l), \quad \forall 1 \leq k < l \leq n.$$

$$\lim_{n \rightarrow \infty} nP\{X_{1,k} > a_n x_k + b_n, X_{1,l} > a_n x_l + b_n\} = 0, \quad \forall 1 \leq k < l \leq n, \quad x_k, x_l, G_1(x_k), G_1(x_l) > 0.$$

$$\lim_{t \rightarrow x^*} P\{X_{1,k} > t \mid X_{1,l} > t\} = 0, \quad x^* = \sup\{x : F(x) < 1\}.$$

安定裾従属性関数 $G \sim \text{SMvEV}$, $G(\mathbf{y}) = \exp(-V(\mathbf{y}))$, $V(\mathbf{y}) = \mu((\mathbf{0}, \mathbf{y}]^c)$, にたいして,

$$\begin{aligned} \varphi(\mathbf{y}) &= V(1/y_1, \dots, 1/y_d) = -\log G(1/y_1, \dots, 1/y_d) \\ &= \int_{S_d} \max(y_1 w_1, \dots, y_d w_d) dH(\mathbf{w}), \quad \mathbf{y} \in [0, \infty]. \end{aligned}$$

で定義される関数 φ を安定裾従属性関数 (stable tail dependence function) と呼ぶ. φ は次の性質をもつ.

1. $\varphi(t\mathbf{y}) = t\varphi(\mathbf{y})$
2. $\varphi(\mathbf{e}_j) = 1$, $j = 1, \dots, d$, \mathbf{e}_j : unit vector.
3. $\max(y_1, \dots, y_d) \leq \varphi(\mathbf{y}) \leq \sum_{j=1}^d y_j$,
the lower bound is the comonotone case and the upper bound is the independent case.
4. $\varphi(t\mathbf{y})$ is convex.

したがって φ は SMvEV に自然な正従属性半順序関係を導入する. 次節の Pickands 従属性関数参照.

6 2 変量極値分布

6.1 2 変量単純極値分布, 従属性関数

$d = 2$ の場合の単純多変量極値分布 (20)–(22) は, $S_d = [0, 1]$ となるために, より単純となる.

定理 19 (2 変量単純極値分布). 標準 Fréchet 分布を周辺分布とする 2 変量極値分布関数 $G(x_1, x_2)$ は次のように表せる.

$$G(x_1, x_2) = \exp(-V(x_1, x_2)), \quad x_1 > 0, \quad x_2 > 0, \quad (23)$$

$$V(x_1, x_2) = \int_0^1 \max\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right) dH(w). \quad (24)$$

$H/2$ は $[0, 1]$ 上の分布関数で平均値

$$\frac{1}{2} \int_0^1 w dH(w) = \frac{1}{2}, \quad (25)$$

のものである. たとえば $w = 1/2$ に関して対称であれば十分である. \square

V , (24), は $V(x_1/a, x_2/a) = aV(x_1, x_2)$ を満たすから, G , (23) は

$$G^r(rx_1, rx_2) = G(x_1, x_2), \quad r \in \mathbb{R}_{>0},$$

を満たす. つまり G , (23)–(25), は最大値安定である.

1. $H/2$ が 0, 1 を確率 1/2 でとる分布ならば,

$$V(x_1, x_2) = 1/x_1 + 1/x_2, \quad x_1 > 0, \quad x_2 > 0,$$

これから G は独立な標準 Fréchet 分布となる.

2. $H/2$ が $w = 1/2$ に退化した分布ならば,

$$G(x_1, x_2) = \exp(-\max(1/x_1, 1/x_2)), \quad x_1 > 0, \quad x_2 > 0,$$

となるが, これは $x_1 = x_2$ の上に退化した分布で, $X_1 = Y \sim F(z)$ のときの (X_1, Y) の分布関数である.

$d = 2$ のときは, 次の Pickands 従属性関数 (dependence function) A が見やすい. φ と同等の正従属性半順序関係を導入する. 後述の他の標準化との関係も示す.

関数 A を

$$V\left(\frac{1}{v_1}, \frac{1}{v_2}\right) \equiv (v_1 + v_2)A\left(\frac{v_1}{v_1 + v_2}\right) \quad (26)$$

により定義する. 書き換えると

$$A(t) := \int_0^1 \max(t(1-q), (1-t)q)H(dq), \quad t \in [0, 1] \quad (27)$$

関数 A は

$$\max(t, 1-t) \leq A(t) \leq 1,$$

を満たす凸関数で, 上限は独立, 下限は comonotone の場合に相当する. 凸関数であることは,

$$\begin{aligned} A(t) &= t \int_{[0,t]} (1-q)H(dq) + (1-t) \int_{(t,1]} qH(dq), \\ \int_{(t,1]} qH(dq) &= H((t,1]) - \int_{(t,1]} (1-q)H(dq) = (2 - H([0,t])) - (1 - \int_{[0,t]} (1-q)H(dq)) \\ &= 1 - H([0,t]) + \int_{[0,t]} (1-q)H(dq), \\ \int_{[0,t]} (1-q)H(dq) &= \int_{[0,t]} \int_t^1 du H(dq) = \int_0^1 \int [0, q \wedge t] H(dq) du = \int_0^q \int [0, u] du + (1-q)H([0,q]), \\ A(t) &= 1 - t + \int_0^t H([0,q])dq, \quad t \in [0, 1]. \end{aligned}$$

最後の式は非減少関数の積分であるから凸である.

注意

1. 定義

(26) 式の定義で右辺を $(v_1 + v_2)A(v_1/(v_1 + v_2))$ とすることもある. つまり (27) 式の左辺を $A(1-t)$ とすることになる.

2. 一般化

$A(t)$ は前小節で導入した安定従属性関数の特別な場合である: $A(t) = \varphi(1-t, t)$.

3. 接合分布関数

対応する極値接合分布関数は次のようになる. 前節参照.

$$C(u_1, u_2) = \exp\left(\log(u_1 u_2)A\left(\frac{\log u_2}{\log(u_1 u_2)}\right)\right).$$

例

1. mixed model

$$A(t) = \theta t^2 - \theta t + 1, \quad 0 \leq \theta \leq 1,$$

$$V(x_1, x_2) = \frac{1}{x_1} + \frac{1}{x_2} - \frac{\theta}{x_1 + x_2}.$$

2. asymmetric mixed model

$$A(t) = \phi t^3 + \theta t^2 - (\theta + \phi)t + 1, \quad \theta \geq 0, \theta + 2\phi \leq 1, \theta + 3\phi \geq 0,$$

$$V(x_1, x_2) = \frac{1}{x_1} + \frac{1}{x_2} - \frac{(2\theta + \phi)x_1 + (\theta + \phi)x_2}{(x_1 + x_2)^2}.$$

3. logistic model

$$A(t) = ((1-t)^r + t^r)^{1/r}, \quad r \geq 1,$$

$$V(x_1, x_2) = \left(\frac{1}{x_1^r} + \frac{1}{x_2^r} \right)^{1/r}.$$

Klüppelberg and May (2006) は多項式 Pickands 従属性関数から生成される極値分布が、それほど多様でなく、特に強い従属性が得られないことを示した。

6.2 2変量極値分布の例

Pickands 従属性関数により 2 変量極値分布が完全に特徴づけられるが、必要な条件を満たす $V(x_1, x_2)$ を構成することはそれほど簡単ではない。以下でその例を述べる。

例 1. ロジスティック・モデル

$$G(x_1, x_2; \alpha) = \exp(-(x_1^{-1/\alpha} + x_2^{-1/\alpha})^\alpha), \quad x_1 > 0, \quad x_2 > 0, \quad 0 < \alpha < 1,$$

実際対応する関数

$$h(w) = \frac{1}{2}(\alpha^{-1} - 1)(w(1-w))^{-1-1/\alpha}(w^{-1/\alpha} + (1-w)^{-1/\alpha})^{\alpha-2},$$

$$0 < w < 1,$$

が確率密度関数であることが確かめられ、対称性から平均値が 1/2 となる。

$$G(x_1, x_2; \alpha) \rightarrow \begin{cases} \exp(-(x_1^{-1} + x_2^{-1})), & \alpha \rightarrow 1, \\ \exp(-\min(x_1^{-1}, x_2^{-1})), & \alpha \rightarrow 0, \end{cases}$$

であるから、 α は独立性の程度を示すパラメータとなるので役に立つ。

例 2. 双ロジスティック (bilogistic) モデル Joe, Smith and Weissman (1992)

ロジスティックは対称なものに限られるという制約がある。これを修正して

$$h(w) = \frac{1}{2}(1-\alpha)(1-w)^{-1}w^{-2}(1-u)u^{1-\alpha}\{\alpha(1-u) + \beta u\}^{-1}, \quad 0 < w < 1,$$

ただし $0 < \alpha < 1, 0 < \beta < 1$, で $u = u(w, \alpha, \beta)$ は

$$(1-\alpha)(1-w)(1-u)^\beta - (1-\beta)wu^\alpha = 0$$

の一意解である。 $\alpha = \beta$ のときはロジスティック・モデルに帰着する。

例 3. デイリクレ・モデル, Coles and Tawn (1991).

$$h(w) = \frac{\alpha\beta\Gamma(\alpha+\beta+1)(\alpha w)^{\alpha-1}\{\beta(1-w)\}^{\beta-1}}{2\Gamma(\alpha)\Gamma(\beta)\{\alpha w + \beta(1-w)\}^{\alpha+\beta+1}}, \quad 0 < w < 1, \quad \alpha > 0, \quad \beta > 0.$$

説明

$$u := \frac{\alpha w}{\alpha w + \beta(1-w)},$$

と置くと

$$\frac{du}{dw} = \frac{\alpha\beta}{(\alpha w + \beta(1-w))^2} > 0, \quad \frac{d^2u}{dw^2} = \frac{-2\alpha\beta(\alpha-\beta)}{(\alpha w + \beta(1-w))^3},$$

である。したがって $u = u(w; \alpha, \beta)$ は $(0, 1) \mapsto (0, 1)$ で $\alpha > \beta$ なら凹, $\alpha < \beta$ ならば凸の変換であり

$$U = \frac{\alpha W}{\alpha W + \beta(1-W)} \left(\Leftrightarrow W = \frac{\beta U}{\alpha - (\alpha - \beta)U}, 1 - W = \frac{\alpha(1-U)}{\alpha - (\alpha - \beta)U} \right),$$

はベータ分布 $\text{Be}(\alpha, \beta)$ に従う。

これらの分布, 他の分布について, たとえば Kotz and Nadarajah (2000) を見よ。

6.3 推測

GPD を仮定 データ $(X_{1,n}, \dots, X_{d,n})_{n=1}^N$ と閾値 (u_1, \dots, u_d) が与えられ, 閾値を越えた $\bar{X}_{i,n} - u_i$ が一般パレート GPrt (η_i, γ_i) に従うと仮定する. データより η_i, γ_i および $\tau_i = P\{X_{i,n} > u_i\}$ を推定する.

$$Y_{i,n} = \bar{W}_{\tau_i}((X_{i,n} - u)/\hat{\eta}_i)$$

が近似的に一様分布に従うから,

$$Z_{i,n} = (-\log(1 - \tau_i Y_i))^{-1}$$

が標準 Fréchet 分布に従うとみなす: $P\{Z_{i,n} \leq z\} = \exp(-1/z)/\tau_i$, $z \geq (-\log(1 - \tau_i))^{-1}$.

成分ごとに閾値を越えるものが異なる. 一つの方法はすべての成分で閾値を越えるように, 低く設定することだが, それには限度がある. もう一つの方法は, 閾値以下の観測値は, その順位だけを用いて “Fréchet スコア” に置き換えるという近似をする. つまり $X_{i,n}$ を擬似観測値 $Z_{i,n} = 1/\log N/(R_{i,n} - 1/2)$, ただし $R_{i,n}$ は $X_{i,n}$ の $(X_{i,n})_{n=1}^N$ における順位である.

GEV を仮定

Z が標準 Fréchet 分布に従うとき

$$\bar{Z} = \psi(Z; \xi, \eta, \zeta), \quad \psi(z; \xi, \eta, \zeta) := \eta + \frac{\zeta}{\xi}(z^\xi - 1),$$

は GEV 分布に従う. η, ζ は, それぞれ位置, 尺度パラメータである. したがって (M_1, M_2) が (23) に従うとき

$$(\psi(M_1; \xi_1, \eta_1, \zeta_1), \psi(M_2; \xi_2, \eta_2, \zeta_2))$$

が一般の 2 変量極値分布に従う.

その分布関数は

$$G(x_1, x_2) = \exp\{-V(\bar{x}_1, \bar{x}_2)\}, \quad \bar{x}_1 = \psi(x_1; \xi_1, \eta_1, \zeta_1), \quad \bar{x}_2 = \psi(x_2; \xi_2, \eta_2, \zeta_2),$$

である.

各成分を変換した

$$\psi(X_{n,1}; \xi_1, \eta_1, \zeta_1), \quad \psi(X_{n,2}; \xi_2, \eta_2, \zeta_2),$$

が GEV に従うものとしてパラメータを推定する.

$$X_{n,1}^* = \psi^{-1}(X_{n,1}, \bar{\xi}_1, \bar{\eta}_1, \bar{\zeta}_1), \quad X_{n,2}^* = \psi^{-1}(X_{n,2}, \bar{\xi}_2, \bar{\eta}_2, \bar{\zeta}_2),$$

の極値が (23) に従うとする. その同時確率密度関数は

$$g(x_1, x_2) = \{V_1(x_1, x_2) V_2(x_1, x_2) - V_{12}(x_1, x_2)\} \exp(-V(x_1, x_2)), \quad x_1 > 0, x_2 > 0,$$

である. 対数尤度は

$$l(\theta) = \sum_{n=1}^N \log g(z_{n,1}, z_{n,2}), \quad z_{n,1} = \psi^{-1}(x_{n,1}, \bar{\xi}_1, \bar{\eta}_1, \bar{\zeta}_1), \quad z_{n,2} = \psi^{-1}(x_{n,2}, \bar{\xi}_2, \bar{\eta}_2, \bar{\zeta}_2),$$

となる.

6.4 裾の漸近的独立性・従属性

一般に連続対称 2 変数分布関数 $F(x_1, x_2)$, $F_1(x) = F_2(x) = F(x, \infty)$, をもつ確率変数 (X_1, X_2) において

$$\lambda = \lim_{x \rightarrow x^*} P\{X_2 > x | X_1 > x\}, \quad x^* = \sup\{x : F(x) < 1\},$$

が存在するならば, これは (X_1, X_2) の裾における従属性の自然な尺度である. 対称性により X_1, X_2 を交換しても λ は同じである. F の接続関数 C , 生存接続関数 \hat{C} を用いて書き換えると

$$\lambda = 1 - \lim_{x \rightarrow x^*} \frac{F_1(x) - F(x, x)}{1 - F_1(x)} = 2 - \lim_{u \rightarrow 1} \frac{1 - C(u, u)}{1 - u} = 2 - \lim_{u \rightarrow 0} \frac{\hat{C}(u, u)}{u}.$$

$\lambda = 0$ (あるいは $\lambda > 0$) のとき F または (X_1, X_2) の裾が漸的に独立 (あるいは従属) であるという.

2 変数正規分布では, 相関係数 ρ が $\rho < 1$ であれば漸的に独立である, Sibuya (1960). 裾が漸的に独立であると, 多変量極値分布に基づく推論には困難を生じる. 小規模データからは, 弱い従属性があると想定するために, データの外に補外すると偏りを生じる.

漸的に従属な場合をさらに詳しくするために次の関数 $\lambda(u)$ を導入しておく.

$$\lambda(u) := 2 - \frac{\log C(u, u)}{\log u}, \quad 0 < u < 1, \quad \lambda(u) = \lambda + o(1 - u), \quad u \rightarrow 1.$$

$C(u, u)$ の Fréchet-Hoeffding 限界より,

$$2 - \frac{\log(\max(2u - 1, 0))}{\log u} \leq \lambda(u) \leq 1.$$

他の尺度との関係

$$P\{\max(X_1, X_2) \leq x\} = P\{X_1 \leq x, X_2 \leq x\} = F(x, x) =: F_1^{\theta(x)}(x), \quad 1 \leq \theta \leq 2,$$

とすると $\lambda(u) \equiv 2 - \theta(x(u))$ であり, F が独立な場合は, $\lambda = 0, \theta = 2$, 完全正従属の場合は, $\lambda = 1, \theta = 1$, である. また極値分布にたいしては A を Pickands 従属性関数とすると $\theta = 2 - \lambda = 2A(1/2)$ となる.

joint tail models 漸近的従属性の困難に対処するために Ledford and Tawn (1996, 1997) は joint tail models を提案した. 標準 Fréchet 分布を周辺分布とする 2 変量分布の場合, このモデルでは

$$P\{Z_1 > z_1, Z_2 > z_2\} \sim \frac{\mathcal{L}(z_1, z_2)}{z_1^{c_1} z_2^{c_2}}, \quad z_1, z_2 \rightarrow \infty, \quad c_1, c_2 > 0, \quad c_1 + c_2 \geq 1,$$

ただし \mathcal{L} は 2 変数緩慢変動関数である. coefficient of tail dependence $\eta := 1/(c_1 + c_2) \in (0, 1]$ を用いると, $\eta = 1, \mathcal{L}(r, r) \rightarrow \infty, r \rightarrow \infty$ のときは裾が漸的に従属であるが, $\eta < 1$ ならば裾が漸的に独立であり, その速さを η の推定により知ることができる.

6.5 Sibuya の従属性関数

一般に連続な 2 変数分布関数 F にたいして, G, H をその周辺分布関数とし, 2 変数関数 $\Omega(u, v)$, $0 < u < 1, 0 < v < 1$, を

$$F(x, y) = \Omega(G(x), H(y))G(x)H(y), \quad 0 < G(x) < 1, \quad 0 < H(y) < 1, \quad (28)$$

により定義する. Ω は非負で

$$\max\left(0, \frac{u+v-1}{uv}\right) \leq \Omega(u, v) \leq \min\left(\frac{1}{u}, \frac{1}{v}\right)$$

を満たし, $F(x_1, x_2)$ が独立なときには $\Omega(u, v) \equiv 1$ である. このような Ω を F の従属性関数 (dependence function) と呼ぶ. 不等式の下限は確率分布が減少曲線の上に退化したとき, 上限は同じく増加曲線の上に

退化したときである。確率変数 (X, Y) の従属性関数が Ω であるとき、任意の連続非減少関数 φ, ψ にたいして $(\varphi(X), \psi(Y))$ の従属性関数も Ω である。また Ω が従属性関数であれば任意の分布関数 F, G にたいして $\Omega(F(x), G(y))F(x)G(y)$ は 2 変数分布関数である。

Ω が従属性関数である必要十分条件は $\Omega(u, v)uv, 0 < u < 1, 0 < v < 1$ が接合分布関数 (copula) となることである。従って確率表示の限りでは接合分布関数とほとんど同一に扱うことができる。一つの利点は、等高線または鳥瞰図によるグラフ表示のときに、従属性がより明確になる。

最大値安定性

一般に $F_0^n(x, y)$ の従属性関数を $\Omega_n(G, H), n = 1, 2, \dots$, とすると、

$$\begin{aligned} F_0^n(x_1, x_2) &= \Omega_1^n(G(x), H(y))G^n(x)H^n(y) \\ &= \Omega_n(G^n(x), H^n(y))G^n(x)H^n(y), \quad 0 < G(x) < 1, 0 < H(y) < 1, \end{aligned}$$

より

$$\Omega_n(G^n, H^n) = \Omega_1^n(G, H), \quad n = 1, 2, \dots, G, H \in (0, 1), \quad (29)$$

である。

したがって 2 変量極値分布の従属性関数は最大値安定条件

$$\Omega^r(u^{1/r}, v^{1/r}) = \Omega(u, v), \quad r > 0, \forall (u, v) \in (0, 1)^2.$$

この条件は接合分布関数の最大値安定性と同等である：

特徴付け

以下最大値安定条を解く。 $\omega(\xi, \eta) := \log \Omega(e^\xi, e^\eta)$ とおくと

$$\omega(\xi, \eta) = n\omega\left(\frac{\xi}{n}, \frac{\eta}{n}\right), \quad n = 1, 2, \dots,$$

これは Euler の 1 次同次関数と呼ばれる。

定理 20 (Sibuya 1960). 従属性関数 Ω が最大値安定である必要十分条件は

$$\Omega(u, v) = u^{\chi(\log v / \log u)}, \quad (30)$$

ただし、 $\chi(w), 0 \leq w < \infty$, は凸関数で、

$$\max(-w, -1) \leq \chi(w) \leq 0$$

を満たす。

実際、(30) が最大値安定であることは直ちに分かる。 χ にたいする制約は、 Ω が従属性関数であるための条件である。 $\chi(w)$ として次のような例がある。もしも $\chi(w)$ が条件を満たせば $a\chi(bw), 0 \leq a \leq 1, 0 < b \leq 1$, も条件を満たす。

$$\begin{aligned} \chi_1(w) &= -w/(1+w), & \chi_2(w) &= -w/\sqrt{1+w^2}, \\ \chi_3(w) &= -1+e^{-w}, & \chi_4(w) &= 2/(1+e^w) - 1. \end{aligned}$$

最大値安定接合分布関数は次のように、より陽に表せる。

$$\begin{aligned} F(u, u^\alpha) &= u^{\chi(\alpha)+\alpha+1}, \quad 0 < u < 1, 0 < \alpha, \\ F(v^{1/\alpha}, v) &= v^{\chi(\alpha)+1/\alpha+1}, \quad 0 < v < 1, 0 < \alpha. \end{aligned}$$

7 多変量極値モデル

7.1 種々の標準化 (周辺分布の選択)

多変量極値分布を議論するとき、周辺分布を選択することができる。次のような分布が候補となる。実際これらが諸論文に登場している。これまで議論した一様分布変数 X 、標準 Fréchet 変数 Y からの変換を記してある。

- 一様分布 (copula) $X = \exp(-1/Y)$
- 標準 Fréchet $Y = -1/\log X$
- 指数分布 (最小値) $-\log X = 1/Y$
- 標準極値 Weibull $\log X = -1/Y$
- 標準 Gumbel $Y = \log(-1/\log X) = \log Y$

何を採んでも本質的に同じであり、目的に応じてある形式が美しく直観的に理解しやすいことになる。

一様分布すなわち接合分布関数 (copula) は従属性一般の表現として便利であり、皆が使い出した。一般理論の理解と、金融・保険データの解析のためには標準 Fréchet がよい。特に一般 Pareto との対応が理解しやすい。欠点は裾が重すぎるため視覚化が困難である。

指数分布はもっとも扱いやすいが、この場合にだけ最小値を扱うために混乱を生じやすい。しかし製品の寿命など最小値を扱っているときには当然利用すべきである。ただし分布関数でなく生存関数を用いることにより、最大値について、他の場合と平行した議論となる。極値 Weibull を用いることの利点は原点への収束を扱うことである。そのために解析的な議論をしやすい。欠点としてはすべての成分が負の象限で議論することである。

Gumbel 分布に近いデータを扱っている限り、これからは外れることはないが、陽な形に表せることが少なくなる。Gumbel 分布は実軸上の典型的分布であるから、一般論を展開しやすい。

7.1.1 指数周辺分布

周辺分布が指数分布のときは、分布関数でなく生存関数について議論することにより他の場合と平行した議論ができる。

定理 21. 2 変数生存関数 $S(x_1, x_2)$ が最大値安定である必要十分条件は、

$$S(x, y) = \exp\left(- (x+y)A\left(\frac{y}{x+y}\right)\right), \quad (31)$$

$A(w)$ は Pickands の従属性関数 (26) である。

Tawn (1988) は上記の条件を満たすいくつかの確率分布を導入し推測法を議論した。

Sibuya 従属性関数の χ を用いると、生存関数の最大安定条件は、

$$\begin{aligned} S(x, cx) &= \exp(-x(1 + \chi(c) + c)), & 0 < c, \\ S(cy, y) &= \exp(-y(1 + \chi(c) + c)), & 0 < c. \end{aligned}$$

7.2 単純多変量極値分布の例

2 変量極値分布 $G(x_1, x_2)$ とその成分分布 $H(w)$ との関係で、 H が $w = 0, 1$ を確率 $1/2$ での分布のとき G は独立、 H が $w = 0.5$ に退化した分布のときは G も $x = y$ に退化した分布であった。多変量となると、多様な退化が生じる。つまり $C = \{1, \dots, d\}$ を添字集合、その任意の部分集合を $c \subseteq C$ 、とし単位単体 S_d の、 c 成分以外は 0 の '側面、陵' を

$$S_c := \{w \in S_d : w_k = 0, k \notin c\}$$

で表す。ただしこれは記号法の拡張利用で、 S_c の 0 成分を除いたものが、 c の成分数 (cardinality) を $|c|$ で表すと、これまでの記号の $S_{|c|}$ である。

指数測度 V が微分可能であるとする、 H は S_d の内部だけでなくすべての S_c の上で確率密度をもつ。 $(S_c$ は $S_{|c|}$ と同相であり、両者を混用する。) S_c の上の確率密度関数を h_c または h_{mc} , $m = |c|$ で表すと、

$$\frac{\partial V}{\partial x_{i_1} \dots \partial x_{i_m}} = - \left(\sum_{j=1}^m x_{ij} \right)^{-(m+1)} h_c \left(x_{i_1} / \sum_{j=1}^m x_{ij}, \dots, x_{i_m} / \sum_{j=1}^m x_{ij} \right)$$

となる。定理 16 の説明で導入した集合 $A^c = A^c(\mathbf{z})$ を拡張して $b \subseteq c$ にたいして

$$A_b^c(\mathbf{z}) = \{ \mathbf{0} \leq \mathbf{x} \leq \mathbf{z} : 0 < z_j < \infty, j \in b; z_j = 0, j \notin b \} \subseteq \mathbb{R}^d,$$

と定義する。

$$V(\mathbf{z}) = \mu(\cup_{\emptyset \neq b \subseteq c} A_b^c(\mathbf{z})) = \sum_{\emptyset \neq b \subseteq c} (-1)^{|b|-1} \mu(A_b^c(\mathbf{z})),$$

である。最後の式は包含排除法で導く。(同時生存関数を周辺分布関数で表す公式と類似している。) これに微分作用素

$$D_a := \frac{\partial^{|a|}}{\prod_{i \in a} \partial z_i}, \quad \emptyset \neq a \subseteq c$$

を作用する。 $D_a \mu(A_b^c(\mathbf{z})) = 0$, $b \subset a$, であるから

$$D_a V(\mathbf{z}) = \sum_{a \subseteq b \subseteq c} (-1)^{|b|-1} D_a \mu(A_b^c(\mathbf{z})),$$

この和を $a = b$ と $a \subset b$ に分け、後者に再び包含排除法を適用すると、

$$\begin{aligned} D_a V(\mathbf{z}) &= (-1)^{|a|-1} D_a \mu(A_a^c(\mathbf{z})) + (-1)^{|a|} D_a \mu(\cup_{a \subset b \subseteq c} A_b^c(\mathbf{z})) \\ &= (-1)^{|a|-1} D_a \mu(\mathbf{x} \geq \mathbf{0} : x_j > z_j, j \in a \text{ and } x_j \leq z_j, j \notin a) \end{aligned}$$

ここで $z_j \rightarrow 0$, $j \notin a$, とすると

$$\lim_{z_j \rightarrow 0, j \notin a} D_a V(\mathbf{z}) = (-1)^{|a|-1} D_a \mu(A_a^c(\mathbf{z}))$$

H の S_d 上の密度を h_a とすると

$$\mu(A_a^c(\mathbf{z})) = \int_{z_{j_1}}^{\infty} \dots \int_{z_{j_m}}^{\infty} h_a \left(\frac{x_1}{\sum x_i}, \dots, \frac{x_{m-1}}{\sum x_i} \right) \left(\sum z_i \right)^{-(|a|+1)} dx_1 \dots dx_m,$$

これに D_a を適用すれば

$$\lim_{z_j \rightarrow 0, j \in a} D_a V(\mathbf{z}) = -r^{-(m+1)} h_a \left(\frac{z_{j_1}}{r}, \dots, \frac{z_{j_{m-1}}}{r} \right), \quad r = \sum_{j \in a} z_j.$$

□

ロジスティックモデル

$$\begin{aligned} V(\mathbf{x}) &= \left(\sum_{j=1}^d x_j^{-r} \right)^{1/r}, \quad \mathbf{x} \in \mathbb{R}_{>0}^d, \quad r > 1, \\ h(\mathbf{w}) &= \left\{ \prod_{j=1}^{d-1} (j^r - 1) \right\} \left(\prod_{j=1}^d w_j \right)^{-(r+1)} \left(\sum_{j=1}^d w_j^{-r} \right)^{1/r-d}, \quad \mathbf{w} \in S_d, \quad r > 1. \end{aligned}$$

これは 1 個のパラメータ r しかもたず、対称 (exchangeable) である。

7.2.1 非対称ロジスティックモデル

上記の分布を一般的にして次の分布族が得られる.

$$G(\mathbf{x}) = \exp \left(- \sum_{b \subset c} \left(\sum_{i \in b} \left(\frac{\theta_{ib}}{x_i} \right)^{r_b} \right)^{1/r_b} \right),$$

(ただし b は $\binom{d}{|b|}$ 個の添え字集合のそれぞれで, さらに $|b| = 1, \dots, d$ の範囲を動かす.)

$$r_b \geq 1, \quad \theta_{ib} \geq 0, \quad \sum_{b \in c} \theta_{ib} = 1, \quad i = 1, \dots, d.$$

(ただし $\theta_{ib} = 0, i \notin b$ とする).

この分布のよい特徴はすべての周辺分布にたいするスペクトル測度の密度関数を, 次のように陽に表せることである.

$$h_{jb}(\mathbf{w}) = \prod_{k=1}^{j-1} (kr_b - 1) \left(\prod_{i \in b} \theta_{ib} \right)^{r_b} \left(\prod_{i \in b} w_i \right)^{-(r_b+1)} \left(\sum_{i \in b} \left(\frac{\theta_{ib}}{w_i} \right)^{r_b} \right)^{1/r_b - j}, \quad \mathbf{w} \in S_{jb}.$$

$d = 3$ の場合をより詳しく, 早くと次の通りとなる.

$$\begin{aligned} V(\mathbf{x}) = -\log G(\mathbf{x}) &= \frac{\theta_{11}}{x_1} + \frac{\theta_{12}}{x_2} + \frac{\theta_{13}}{x_3} + \left(\left(\frac{\theta_{(21)1}}{x_1} \right)^{r_{(12)}} + \left(\frac{\theta_{(12)2}}{x_2} \right)^{r_{(12)}} \right)^{1/r_{(12)}} \\ &+ \left(\left(\frac{\theta_{(31)1}}{x_1} \right)^{r_{(13)}} + \left(\frac{\theta_{(13)3}}{x_3} \right)^{r_{(13)}} \right)^{1/r_{(13)}} + \left(\left(\frac{\theta_{(32)2}}{x_2} \right)^{r_{(23)}} + \left(\frac{\theta_{(23)3}}{x_3} \right)^{r_{(23)}} \right)^{1/r_{(23)}} \\ &+ \left(\left(\frac{\theta_{31}}{x_1} \right)^{r_3} + \left(\frac{\theta_{32}}{x_2} \right)^{r_3} + \left(\frac{\theta_{33}}{x_3} \right)^{r_3} \right)^{1/r_3}, \quad \mathbf{x} \in \mathbb{R}_{>0}^3, \quad r_{(12)}, r_{(13)}, r_{(23)} \geq 1, \quad r_3 \geq 1, \quad \theta_{ij} \geq 0, \\ &\theta_{11} + \theta_{(21)1} + \theta_{(31)1} + \theta_{31} = 1, \quad \theta_{12} + \theta_{(12)2} + \theta_{(32)2} + \theta_{32} = 1, \\ &\text{and } \theta_{13} + \theta_{(13)3} + \theta_{(23)3} + \theta_{33} = 1. \end{aligned}$$

パラメータの数は 13 個である. $r_{(12)} = r_{(13)} = r_{(23)} = r_3 = 1$ ならば $G(\mathbf{x})$ は独立である. $r_{(12)}, r_{(13)}, r_{(23)}, r_3$ を (他のパラメータは固定して) 増加すると $G(\mathbf{x})$ はより従属となる. 2 変量周辺分布の例を 1 つあげる.

$$\begin{aligned} V(x_1, x_2, \infty) &= -\log G(x_1, x_2, \infty) \\ &= \frac{\theta_{11} + \theta_{(31)1}}{x_1} + \frac{\theta_{12} + \theta_{(32)2}}{x_2} \\ &+ \left(\left(\frac{\theta_{(21)1}}{x_1} \right)^{r_{(12)}} + \left(\frac{\theta_{(12)2}}{x_2} \right)^{r_{(12)}} \right)^{1/r_{(12)}} + \left(\left(\frac{\theta_{31}}{x_1} \right)^{r_3} + \left(\frac{\theta_{32}}{x_2} \right)^{r_3} \right)^{1/r_3}. \end{aligned}$$

7.2.2 h の別表現

制約条件

$$\int w_j dH(\mathbf{w}) = 1, \quad j = 1, \dots, d, \quad (32)$$

の代わりに, 周辺分布の尺度母数が (m_1, \dots, m_d) の $\mathbf{X}^* \sim H^*(\mathbf{v})$ を用いる. つまり

$$\int v_j dH^*(\mathbf{v}) = m_j, \quad j = 1, \dots, d,$$

で $Y_j = X_j/m_j, j = 1, \dots, d$, とすると (Y_1, \dots, Y_d) の周辺分布は確率分布で(32)を満たす. \mathbf{X}^* の分布関数

$$G^*(\mathbf{x}) = \exp \left(- \int_S \bigvee_{j=1}^d \frac{v_j}{x_j} dH^*(\mathbf{v}) \right)$$

の変換より Y の分布関数は

$$G(y) = \exp\left(-\int_S \bigvee_{j=1}^d \frac{w_j}{m_j y_j} dH^*(w)\right) = \exp\left(-\int_S \bigvee_{j=1}^d \frac{w_j}{y_j} dH(w)\right)$$

H^* が S の内部での密度 $h^*(w)$ をもつとすれば, H の密度 $h(w)$ は

$$\begin{aligned} h(w) &= \frac{J(w)}{m \cdot w} h^*\left(\frac{m_1 w_1}{m \cdot w}, \dots, \frac{m_d w_d}{m \cdot w}\right), \quad J(w) = \frac{\partial(u_1, \dots, u_d)}{\partial(w_1, \dots, w_d)}, \\ &= (m \cdot w)^{-(d+1)} \left(\prod_{j=1}^d m_j\right) h^*\left(\frac{m_1 w_1}{m \cdot w}, \dots, \frac{m_d w_d}{m \cdot w}\right), \end{aligned}$$

となる. ($J(w)$ の計算は簡単ではない.)

Dirichlet モデル

S_d 上の Dirichlet 分布で, 周辺分布の期待値が 1 となるように尺度変換すると次の分布が得られる. この例は Pickands の表現の拡張から得られる.

$$h(w) = \prod_{j=1}^d \frac{\alpha_j}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha \cdot 1)}{(\alpha \cdot w)^{d+1}} \prod_{j=1}^d \left(\frac{\alpha_j w_j}{\alpha \cdot w}\right)^{\alpha_j - 1}, \quad w \in S_d, \quad \alpha = (\alpha_1, \dots, \alpha_d), \quad \alpha_j > 0.$$

7.3 スペクトル表現

定理 16 のスペクトル測度 H をより広くし, それにより極値理論の適用分野を広くすることができる.

定理 22 (de Haan, 1984). $F(x_1, \dots, x_p)$ が SMVEV である必要十分条件は, 可測非負関数 $f_t(s), s \in [0, 1], t \in \{1, \dots, p\}$ で,

$$\int_0^1 f_t(s) ds = 1, \quad t \in T := \{1, \dots, p\},$$

を満たすものが存在して

$$F(x_1, \dots, x_p) = \exp\left(-\int_0^1 \max_{i \in T} (f_i(s)/x_i) ds\right) \quad (33)$$

と表せることである.

この定理は $T = \mathbb{R}$ の場合まで拡張できる.

(33) の表現は次のようなポアソン過程により生成できる.

$(X_k, W_k)_{k=1}^{\infty}$ を $\mathbb{R}_{>0} \times [0, 1]$ 上の, 強度 $(dx/x^2)dw$ のポアソン過程とする. 非負可測関数 $f_t(w), w \in [0, 1], t \in \{1, \dots, p\}$ で,

$$\int_0^1 f_t(w) dw < \infty, \quad t \in T,$$

を満たすとき

$$Y_t := \sup_{k \geq 1} f_t(W_k) \cdot X_k, \quad t \in T,$$

の分布関数は, 周辺尺度を標準化すれば, (33) となる.

$C[0, 1]$ 上の極値

たとえば長い海岸線があり, そのパラメータ t による表示を $L = \{(x(t), y(t)); t \in [0, 1]\}$ とする. 点 $(x(t), y(t))$ での高波を $Z(t), t \in [0, 1]$, とする. これが堤防の高さ (閾値) $w(t)$ をどこかで超えると (地表が海面下の場合など) 氾濫の危険がある.

台風来襲などのイベント $k = 1, \dots, n$ における高度 $Z_k(t), t \in [0, 1]$, が k に関して独立であるとし、地点ごとの最高波

$$Z_{(n)}(t) := \max_{1 \leq k \leq n} Z_k(t), \quad t \in [0, 1]$$

を調べる. $Z_k(t), t \in [0, 1], k = 1, \dots, n$ は $[0, 1]$ 上の連続関数のクラス $C[0, 1]$ を値とする確率過程であり, $Z_{(n)}(t), t \in [0, 1], n = 2, 3, \dots$ はその極値統計量である. $t \in \{1, \dots, d\}$ の多変量極値統計量の連続版である.

$Z(t)$ にたいして適当な関数列 $a_n(t) > 0, b_n(t) \in \mathbb{R}, t \in [0, 1]$, が存在して

$$\lim_{t \rightarrow \infty} \frac{Z_{(n)}(t) - b_n(t)}{a_n(t)} = Y(t), \quad t \in [0, 1],$$

が退化していない確率過程であれば, $Y(t)$ は最大値安定過程である. たとえば, 与えられた $w(t), t \in [0, 1]$ にたいする事象

$$\bigcup_{t \in [0, 1]} \{Z_{(n)}(t) > w(t)\}$$

の確率近似, 予測することができる.

これらの議論が de Hann and Rootzén (1993), de Haan and Lin (2001) などによって行われ, 応用研究が始まっている.

円周上の極値

$C[0, 1]$ の変形として円周 $C[0, 2\pi]$ 上の極値統計量を Coles and Walshaw (1994) が求めた. $f_t(s), s \in [0, 2\pi], t \in \mathbf{T} = \{\theta_1, \dots, \theta_p\}, \theta_i \in [0, 2\pi]$ として von Mises 円周分布

$$f(s; \theta_i, \zeta) = \frac{1}{2\pi I_0(\zeta)} \exp(\zeta \cos(s - \theta_i))$$

により p 方向多変量極値分布が得られ, さらに $T = [0, 2\pi]$ として全方向の極値分布が得られる.

7.4 ノンパラメトリック推定

定理 4 をさらに補うと, 次の 4 条件は同等である, Takahashi (1994a) および Fougères (2003) 参照.

1. $(\mathbf{X}_i)_{i=1}^n$ i.i.d. $\sim F \in MDA(G), G \sim \text{SMvED}$
- 2.

$$\lim_{t \rightarrow \infty} \frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = \lim_{t \rightarrow \infty} \frac{-\log F(t\mathbf{x})}{-\log F(t\mathbf{1})} = \frac{-\log G(\mathbf{x})}{-\log G(\mathbf{1})} = \frac{\mu((0, \mathbf{x}]^c)}{\mu((0, \mathbf{1}]^c)}$$

- 3.

$$\lim_{t \rightarrow \infty} tP \left\{ \|\mathbf{X}_i\| > t; \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|} \in B \right\} = H(B), \quad B \in S_d.$$

4. $(n^{-1}\mathbf{X}_i)_{i=1}^\infty \xrightarrow{d} \text{Poisson process with mean measure } \mu(\cdot)$.

したがって 3. は

$$\lim_{t \rightarrow \infty} P\{\mathbf{X}_i/\|\mathbf{X}_i\| \in B \mid \|\mathbf{X}_i\| > t\} = \frac{H(B)}{H(S_d)},$$

と同等であり, これからノンパラメトリック推定量

$$\widehat{S}_n(B) = \frac{1}{k_n} \sum_i I \left[\|\mathbf{X}_i\| > \|\mathbf{X}_{(k_n)}\|, \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|} \in B \right]$$

が得られる。ただし $k_n = \#\{i : \|\mathbf{X}_i\| > t\}$.

同様に

$$\begin{aligned} A_n &:= (0, nu_1] \times \cdots \times (0, nu_d] \subseteq \mathbb{R}_{>0}^d \\ P(A_n^c) &= 1 - F(n\mathbf{u}) \approx -\frac{1}{n} \log G(\mathbf{u}) \\ &= \frac{1}{n} \mu((0, \mathbf{u}]^c) = \frac{1}{n} \int_{S_d} \max_j \frac{w_i}{u_i} dH(\mathbf{w}) \end{aligned}$$

から閾値を超える確率の推定

$$\widehat{P(A_n^c)} = \frac{1}{nk_n} \sum_{j=1}^n \left(\max_j \frac{X_{ij}}{u_j \|\mathbf{X}_i\|} \right) I[\|\mathbf{X}_i\| > \|\mathbf{X}\|_{[k_n]}]$$

を得る.

核関数を用いて、密度関数を推定することも可能である.

8 移動最大値過程 (強定常時系列)

8.1 強定常時系列

$t_1, \dots, t_n \in \mathbb{Z}, k, n \in \mathbb{Z}_{>0}$ にたいして,

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

であるとし, (X_t) の周辺分布関数を F とする. $(\tilde{X}_t)_{t \in \mathbb{N}}$ を同じ分布関数 F に従う i.i.d. 系列とし,

$$M_n = \max(X_1, \dots, X_n), \quad \tilde{M}_n = \max(\tilde{X}_1, \dots, \tilde{X}_n),$$

とする.

任意の $i < j < i' < j'$ と, ある係数列 $(a_n > 0)_n, (b_n)_n$ にたいして, $\max(X_i, \dots, X_j), \max(X_{i'}, \dots, X_{j'})$ が $i' - j \rightarrow \infty$ のときに漸近的に独立ならば

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{(\tilde{M}_n - b_n)/a_n \leq x\} &= H(x), \\ \iff \lim_{n \rightarrow \infty} P\{(M_n - b_n)/a_n \leq x\} &= H^\theta(x), \quad \theta \in (0, 1]. \end{aligned}$$

強定常時系列の例 (1). 定常ガウス時系列.

(2). $X_t = Z_t + cZ_{t-1}$, Z_0, Z_1, \dots i.i.d. $\text{Var}(Z_t) < \infty$.

強定常時系列では, 極大値の塊りが生じる (clustering). これを十分に長い区間に分けて区間ごとの最大値を採ると, これらが近似的に独立となる. そのために M_n は \tilde{M}_n と同じ H に同じ係数で収束する. $u_n := a_n x + b_n$ とすると,

$$P\{\tilde{M}_n \leq u_n\} = F^n(x), \quad P\{M_n \leq u_n\} \approx F^{n\theta}(x),$$

となるから, 有効なサンプル・サイズが θ 倍に減少する. これは区間内の利用可能なデータが平均 $(1/\theta)$ 個あるのが無視されるためである. パラメータ θ を **極値指数** (extremal index) と呼ぶが, tail index γ を extreme value index とも呼ぶのと紛らわしい. Nandagopalan (1994) は多変量の場合の extremal index を議論した.

参照. 吉原健一 (2004) 弱従属性をもつデータに基づく極値統計の最近の話題, 統計数理, 52, 25-43.
別紙図強定常時系列参照.

8.2 極値時系列 (M3/M4 process)

標準 Fréchet 変数の i.i.d. 2 重系列, および非負係数の 2 重系列を導入し, 新しい添え字についてさらに最大値をとることにより, 多様な重み付けができる. この強定常時系列を **maxima of moving maxima**, M3 と呼ぶ. (Deheuvels, 1983)

$$X_t := \max_{\ell \geq 1} \max_k a_{\ell,k} Z_{\ell,t-k}, \quad a_{\ell,k} \geq 0, \quad \sum_{\ell \geq 1} \sum_k a_{\ell,k} = 1,$$

これはさらに多変数時系列 $(\mathbf{X}_t)_{t \in \mathbb{Z}}$, $\mathbf{X}_t = (X_{1t}, \dots, X_{Dt})$ に拡張できる. 変数はそのまま, 係数を 3 重数列とし, 各成分が M3 系列である. これを **multivariate maxima of moving maxima**, M4 と呼ぶ. 従って M3 系列についての議論をせず, 始めから M4 系列を議論する.

$$X_{t,d} := \max_{\ell \geq 1} \max_k a_{\ell,k,d} Z_{\ell,t-k}, \quad a_{\ell,k} \geq 0, \quad \sum_{\ell \geq 1} \sum_k a_{\ell,k,d} = 1, \quad 1 \leq d \leq D. \quad (34)$$

実際には添え字の範囲を $1 \leq \ell \leq L_d$, $-K_{1,\ell,d} \leq k \leq K_{2,\ell,d}$ に限定しその他の $a_{i,k,d}$ の値は 0 とする. また各成分が標準に従うように $\sum_{1 \leq \ell \leq L_d} \sum_{-K_{1,\ell,d} \leq k \leq K_{2,\ell,d}} a_{i,k,d} = 1, \forall d$ を仮定する. 以下ではこの制限を前提として, 添え字の範囲を明記しないこともある. 多変数強定常時系列の収束と extremal index について, 1 変数強定常時系列と同様なことが言えるが, extremal index は定数ではなく, 分布関数の引数に依存する. このような M4 にたいして

$$\begin{aligned} P\{X_{t,d} \leq x_{t,d}, 1 \leq t \leq T; 1 \leq d \leq D\} &= \exp \left(- \sum_{\ell=1}^{\infty} \sum_{s=-\infty}^{\infty} \max_{1-s \leq k \leq T-s} \max_{1 \leq d \leq D} \frac{a_{\ell,k,d}}{x_{s+k,d}} \right), \\ P\{X_{t,d} \leq x_{t,d}, X_{t+1,d} \leq x_{t+1,d}\} &= \exp \left(- \sum_{\ell} \sum_s \max \left\{ \frac{a_{\ell,1-s,d}}{x_{t,d}}, \frac{a_{\ell,2-s,d}}{x_{t+1,d}} \right\} \right), \\ P\{X_{t,d} \leq x_{t,d}, X_{t,d'} \leq x_{t,d'}\} &= \exp \left(- \sum_{\ell} \sum_s \max \left\{ \frac{a_{\ell,1-s,d}}{x_{1,d}}, \frac{a_{\ell,1-s,d'}}{x_{1,d'}} \right\} \right). \end{aligned}$$

また $X_{t,d}, X_{t,d'}$ の間の漸近 (裾) 従属性指数, および $\{X_{t,d}\}$ の $k = \max_{\ell} (K_{1,\ell,d} + K_{2,\ell,d})$ にたいする k 遅れ (k -lag) 漸近従属性指数は

$$\begin{aligned} \lambda_{d,d'} &= 2 - \sum_{\ell=1}^{\max(L_d, L_{d'})} \sum_{s=1-\max(K_{2,\ell,d}, K_{2,\ell,d'})}^{1+\max(K_{1,\ell,d}, K_{1,\ell,d'})} \max\{a_{\ell,1-s,d}, a_{\ell,1-s,d'}\}, \\ \lambda_d(k) &= 2 - \sum_{\ell} \sum_{s=1-K_{2,\ell,d}}^{1+k+K_{1,\ell,d}} \max\{a_{\ell,1-s,d}, a_{\ell,1+k-s,d}\}. \end{aligned}$$

また extremal index は

$$\theta(\tau_1, \dots, \tau_D) = \frac{\sum_{\ell} \max_k \max_d a_{\ell,k,d} \tau_d}{\sum_{\ell} \sum_k \max_d a_{\ell,k,d} \tau_d},$$

により定まる.

M3/M4 process の signature

モデル M4 (34) において, ある Z (たとえば $Z_{\ell,k}$) の値が非常に大きいと, それを含む $X_{t,d} = a_{\ell,t-k,d} Z_{\ell,k}$ が k に近いいくつかの t にたいして大きくなり $X_{t,d}$ の断片が定数列 $a_{\ell,t-k,d}, 1 \leq \ell \leq L_d$ のいくつかに比例する. これを痕跡パターン (signature pattern) と呼ぶ. この事実を利用して定数列を推定する. 次の等式に注意する.

$$P\{X_{t,d} \leq u, X_{t+1,d} \leq u+x\} = P\{X_{t,d} \leq 1, X_{t+1,d} \leq (u+x)/u\}^{1/u}.$$

あるいは

$$P\{X_{t,d} \leq 1, X_{t+1,d} \leq x\} = \exp\left(-\sum_{\ell} \sum_s \max(a_{\ell,1-s,d}, a_{\ell,2-s,d}/x)\right),$$

$$P\{X_{t,d} \leq 1, X_{t,d'} \leq x\} = \exp\left(-\sum_{\ell} \sum_s \max(a_{\ell,1-s,d}, a_{\ell,1-s,d'}/x)\right).$$

そこで

$$b_d(x) := -\log(P\{X_{t,d} \leq 1, X_{t+1,d} \leq x\}), \quad b_{d,d'}(x) := -\log(P\{X_{t,d} \leq 1, X_{t,d'} \leq x\}),$$

$$b_d(x) = \sum_{\ell=1}^{L_d} (a_{\ell,K_2,\ell,d} + \max(a_{\ell,K_2,\ell,d-1,d}, a_{\ell,K_2,\ell,d}/x) + \max(a_{\ell,K_2,\ell,d-2,d}, a_{\ell,K_2,\ell,d-1,d}/x) +$$

$$\dots + \max(a_{\ell,-K_1,\ell,d}, a_{\ell,-K_1,\ell,d+1,d}/x) + a_{\ell,-K_1,\ell,d}/x), \quad 1 \leq d \leq D.$$

$$b_{d,d'}(x) = \sum_{\ell=1}^{\max(L_d, L_{d'})} \sum_{s=1-\max(K_2,\ell,d, K_2,\ell,d')}^{1+\max(K_1,\ell,d, K_1,\ell,d')} \max\{a_{\ell,1-s,d}, a_{\ell,1-s,d'}/x\}.$$

$q_d(x) := xb_d(x)$, $q_{d,d'}(x) := xb_{d,d'}(x)$ は折れ線関数であり, 折れ線の曲がり角 $x_{t,d}$ は $a_{\ell,j,d}/a_{\ell,j',d}$, $a_{\ell,k,d}/a_{\ell,k,d'}$, である. 定数列の間に比例関係がない限り, これらの曲がり角における $b_d(x)$, $b_{d,d'}(x)$ の値から $\{a_{\ell,t,d}\}$ が一意に定まる.

曲がり角を見つけるためには

$$\zeta_{t,d} := X_{t,d}/X_{t+1,d}, \quad E(\zeta_{t,d}) = a_{\ell,K_2,\ell,d}/a_{\ell,K_2,\ell,d-1,d}$$

の事実を用いる. ℓ は $Z_{\ell,k}$ の値が非常に大きいことからきまる.

M3/M4 process の推定法

Zhang and Smith (2007) で提案されている $x_{t,d}$ 決定の手続きは次の通りである.

1. $X_{t+j-1,d}/X_{t+j,d}$, $1 \leq j \leq K$, を L_d 個の大きなクラスターのグループに分ける.
2. 各クラスター内で, 比 $X_{t+j-1,d}/X_{t+j,d}$, $1 \leq j \leq K$, が同時に大きくなるものには同じグループ番号を付ける.
3. 各クラスター内で比の平均を求め曲がり角 $x_{t,d}$ とみなす.
4. $x'_{t,d'}$ を決定するするには, 第 2 段を異なる数列間で行う.
5. $x_{t,d}$, $x'_{t,d'}$ が定まれば, $\hat{b}_d(x)$, $\hat{b}_{d,d'}(x)$ から $a_{\ell,k,d}$ を推定する.

8.3 A.I/D.M4 process

M4 過程ではすべての変数が正の従属性をもっている. 証券市場での上場銘柄の株価, 大洋での風速, 波高, 水位などは正の従属性をもって動いている (co-moving). しかし漸近的に, 部分的に独立に変化する量も予測されている. たおえば風速, 波高, 水位も大洋ではなく離れた沖合い地点 (off-shore) の間では独立となる.

Hefferman, Tawn and Ahang (2007) は次の D 変数 asymptotic (in)dependent M4 process を導入した. $\{U_{t,d}; t \in \mathbb{Z}, 1 \leq d \leq D\}$ を独立正値確率変数の 2 重列, $\{W_{\ell,t,d}; \ell \in \mathbb{Z}_{>0}, t \in \mathbb{Z}\}$ を iid 一般極値 rv の 2 重列, として

$$Y_{t,d} = \max(U_{t,d}^{1/\alpha}, \max_{\ell} \max_k W_{\ell,t-k}/a_{\ell,k,d}), \quad t \in \mathbb{Z}, 1 \leq d \leq D,$$

と定義する. このモデルで係数を変えることにより, ほとんど独立, 正の従属性をもつ, 漸近的に従属な M4 過程が得られる. M4 過程との違いは, 標準分布を一般極値分布に代えること, 別の独立正値 rv の二重列を導入することである.

8.4 MCM4

これまでの議論では大きなピーク (ジャンプ) があつたとき, その影響が間もなく消えるというモデルを扱っていた. これと異なる別のアプローチは, 連続状態空間に関して Markov 連鎖を仮定するモデルである. 定常性を仮定し, 閾値を導入し, 閾値を越える期間のクラスターを Poisson 過程とみなすことにより状態変数は一般 Pareto 分布に従うとし, ピーク間の独立性を保障する. クラスター内の諸統計量について, Harris chain の閾値と期間を増加するときの極限定理を利用して, 情報の損失を防ぎ, クラスター間の独立性を検定により確認し, 必要なパラメータを推定する. たとえば Smith, Tawn and Coles (1997) 参照.

Zhang and Huang (2006) は離散状態空間の Markov 連鎖を導入して, M4 と Markov 連鎖モデルを結合し, 3つの株価指数 Down Jones industrial average (DJI), NASDAQ index, S & P 500 index からなるポートフォリオの分析に適用している. 対数日収益が負, 零, 正の期間があるとし, 状態が $\{-1, 0, 1\}$ の r 重 Markov 連鎖 $\xi_{t,d}$ と, $\xi_{t,d} = \mp$ によって異なる係数行列 A^-, A^+ をもつ M4 過程 $Y_{t,d}^-, Y_{t,d}^+$ を想定し, $\xi_{t,d} = 0$ のときは $Y_{t,d} = 0$ としている. データの前処理は M4 の場合と同様である.

9 応用

1. 保育園のデザート戦争

Sibuya (1960) の動機となった問題を漫画化すると次の通りである. ある保育園で n 人の幼児が昼食を取る. 食べ終わると続いてデザートを食べる. 早い子が両方食べ終わったとき, 遅い子が昼食も終わっていないと残っているデザートを巡って騒動が起こる. 幼児 k の昼食を終える時間, デザートを終える時間をそれぞれ X_k, Y_k とする. 騒動が生じる確率は

$$P\{\min_{1 \leq k \leq n} (X_k + Y_k) < \max_{1 \leq k \leq n} X_k\}$$

X_k, Y_k が独立であると仮定しても, $(X_k + Y_k), X_k$ は独立でない. つまり $(\max X_k, \max(-(X_k + Y_k)))$ の同時分布が必要である.

幸い当時使用していた日本最初の商業用電気機械的計算機 FACOM 128 の能力が低かったため, 同時分布の漸近的独立性を予想し, 周辺分布の極値分布近似の差の分布による, 上記確率分布の近似を計画した.

第一段階が終わり, 表題に I と番号を付けた論文を投稿したところで, 計画が挫折した. 論文の続きは書かずに終わった.

2 変量極値分布の研究は, 統計学の辺境で, まったく独立に, ほとんど同時に始まった. Geffroy (1958/1959), Tiago de Oliveira (1958), Sibuya (1960).

2. 諸応用

Kotz and Nadarajah (2000), には Gumbel の初期の応用を始め, いくつかの紹介がある. パラメトリックモデルとして双ロジスティック分布が代表的であり, Coles and Tawn (1991) はイギリス東海岸 Immingham, Lowestoft, Sheerness, における高波 (surge) の毎時測定値に, Joe, Smith and Weissman (1992) は酸性雨中の硫酸塩, 硝酸塩濃度に双ロジスティック分布を当てはめている.

3. 気候変動

気候変動について, 気象統計学者が注目しているのは平均気温ではなく, 最低気温の上昇と最大降水量の増加である.

Smith, et al. (1997) は Markov chain 極値モデルを提案した. これは M3 と類似した強定常過程の一つである. 水準超過点過程で, あい続く超過のクラスターと, その中の諸統計量の汎関数についての議論から, あい続く k 時点の k 変量極値分布に従うというモデルである. $k = 2$ の場合をカナダの 1 地点における長期データに適用している.

Smith (2004) はアメリカにおける 4 つの離れた気象台におけるデータを詳しく調べ, 大洋海温変動を補助変数として選んだ. この回帰モデルを Historical Climatological Network の 187 観測所における 1951-1997 年の日降水量 (98% 閾値) に適用し, 年最大日降水量の傾向を推定した. 地域によ

て傾向が異なるので、1 次傾向係数の推定値の空間正規過程を想定した。結果としてアメリカ全州を 4 地域に分割し、その 3 地域では増加、1 地域では減少を見出している。

4. 自然災害

de Haan and de Ronde (1998) は、オランダの防波堤 Pettemer zeedijk を波が越える危険確率を推定した。沖合い観測点における潮位、波高、波長について、過去 13 年間に生じた 828 回の暴風雨の際の最高値がデータである。沖合いの状況、防波堤の状況、可能な危険との間の関係は工学的知識で定まる。これから潮位・波高平面、あるいは潮位・波高・波長空間における危険領域が定まる。これらの暴風雨による事故はなく、危険領域はデータの散布図からかなり離れている。SMvEV を推定し、その補外によって確率を推定する。

神田、西嶋 (2004) は台風時の風速を解析し、スペクトル表現におけるスペクトル関数 $f_t(s), s \in [0, 1], t \in T = \{1, \dots, p\}$, を $T = \mathbb{R}^2$ にまで拡張する必要を表明している。

5. ファイナンス

Smith (2003) は Pfizer, General Electric, Citibank の 1982–2001 年間の daily return を解析した。ある社の第 t 日の daily return を y_t とし、

$$y_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad iid,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad t = 2, 3, \dots$$

という GARCH (1, 1) モデルを適用し、その推定値により $y_t/\hat{\sigma}_t$ という標準化を行う。

標準化データで閾値 $u = 1.2$ を越すものにたいして一般パレート分布を当てはめる。当てはめた分布を変換して標準 Fréchet 分布からの標本を構成する。ここまでは 3 社独立に計算を行う。

その結果は、3 社間の従属性があるだけでなく、前回との弱い従属性が見られた。そこで離散時間最大値定常過程のモデルである Deheuvels (1983) の 'Maxima of Moving Maxima (M3) process' を多変量過程に拡張した M4 process を当てはめ、その妥当性を検討している。

10 多変量極値分布の新定義

多変量データの解析において成分ごとの最大値にどれだけ意味があるか、実際家は疑いをもつだろう。空間統計学では標本点の凸包を作り、その頂点を極値とみなし、その割合を調べるような試みもあったが、見るべき成果はないようである。

これまで述べた 1980, 1990 年代に確立した理論の前提となっているのは、周辺分布を標準 Fréchet 分布としたとき、同時生存関数が多変量正則変動をする、という条件であった。多変量極値分布の例が豊富となり、応用範囲が広がるにつれて、この条件が強過ぎると感じられ始めた。

このような流れの中で Heffernan and Tawn (2004) の画期的なアプローチが提案され、制約がずっと弱い極値極限分布が提案された。この論文に対する Royal Statistical Society での諸議論も示すように、新しいアプローチの意味するところ、推測の方法、応用の可能性などについて、考えるべきことは多い。この論文が刺激となって、しばらくは新方向の模索が行われるのであろう。ここでは Heffernan and Tawn (2004) の簡単な紹介に止める。

条件付極値

多変量確率変数 $\mathbf{Y} = (Y_1, \dots, Y_d)$ のすべての周辺分布が標準 Gumbel 分布に従うとする。通常の指定と異なるのは、後で示すように分布範囲を $\mathbb{R}_{>0}$ でなく \mathbb{R} としたためである。 \mathbf{Y}_{-i} により Y_i を除く $d-1$ 変量確率変数を表わし、条件付分布 $P\{\mathbf{Y}_{-i} \leq \mathbf{y}_{-i} | Y_i = y_i\}$ の構造を議論する。簡単のため、ベクトルに関する代数演算により、成分ごとの演算を表す。通常の極値理論と類似した、次の条件付分布に関する収束条件が本質的である。

すべての $i = 1, \dots, d$, にたいして、吸引 (標準化) ベクトル関数の系列 $\mathbf{a}_{1i}(y_i), \mathbf{b}_{1i}(y_i), (\mathbf{b}_{1i}(y_i) > \mathbf{0})$ であって、共に $\mathbb{R} \rightarrow \mathbb{R}^{d-1}$ が存在し、任意に固定した \mathbf{z}_{1i} と y_i の値の任意の系列 $(\lim_{i \rightarrow \infty} y_i = \infty)$ にた

いして

$$\lim_{y_i \rightarrow \infty} \Pr\{Y_{-i} \leq a_{|i}(y_i) + b_{|i}(y_i)z_{|i} | Y_i = y_i\} = G_{|i}(z_{|i}), \quad (35)$$

が成り立ち $G_{|i}$ のすべての周辺分布が退化していないと仮定する。つまり標準化した確率変数

$$Z_{|i} := \frac{Y_{-i} - a_{|i}(y_i)}{b_{|i}(y_i)}$$

の条件 $Y_i = y_i$ の下での極限 $y_i \rightarrow \infty$ の分布である。

条件 (35) の下では, $Y_i > u_i$, $u_i \rightarrow \infty$, のとき $Y_i - u_i$ と $Z_{|i}$ は漸近的に独立で, 極限での周辺分布は, それぞれ指数分布, $G_{|i}(z_{|i})$ となる。実際 Y_i の周辺密度関数 (Gumbel) を f_{Y_i} とすると,

$$\begin{aligned} & P\{Z_{|i} \leq z_{|i}, Y_i - u_i = y | Y_i > u_i\} \\ &= P\{Y_{-i} \leq a_{|i}(u_i + y) + b_{|i}(u_i + y)z_{|i} | Y_i = u_i + y\} \frac{f_{Y_i}(u_i + y)}{P\{Y_i > u_i\}} \\ &\rightarrow G_{|i}(z_{|i}) \exp(-y), \quad u_i \rightarrow \infty, \end{aligned}$$

である。

$G_{|i}$ の周辺分布を考える。 $a_{|i}(y_i)$, $b_{|i}(y_i)$ の Y_j ($j \neq i$) に対応する成分を $a_{j|i}(y_i)$, $b_{j|i}(y_i)$ とし

$$Z_{j|i} = \frac{Y_j - a_{j|i}(y_i)}{b_{j|i}(y_i)},$$

の $Y_i = y_i$ が与えられた条件の下での極限 $y_i \rightarrow \infty$ の分布を $G_{j|i}$ とすると, これは $G_{|i}$ の周辺分布である。もしも

$$G_{|i; z_{|i}} = \prod_{j \neq i} G_{j|i}(z_{j|i})$$

のとき Y_{-i} は Y_i が与えられた条件の下で漸近的独立であるという。

吸引関数 $a_{|i}(y_i)$, $b_{|i}(y_i)$ と対応する極限分布は通常の場合のように一意ではない。吸引関数と極限分布が (35) のとき, 任意の定数ベクトル \mathbf{A} , \mathbf{B} ($\mathbf{B} > \mathbf{0}$) にたいして, 吸引関数

$$\begin{aligned} a_{|i}^*(y_i) &= a_{|i}(y_i) + \mathbf{A}b_{|i}(y_i) \\ b_{|i}^*(y_i) &= \mathbf{B}b_{|i}(y_i) \end{aligned}$$

により極限分布 $G_{|i}(\mathbf{B}z_{|i} + \mathbf{A})$ が得られる。ただし ∞ に正の確率がない限り, 極限分布は定数 \mathbf{A} , \mathbf{B} が自由であることを除いては一意である。

添字 i を固定したとき, 吸引関数の選択は $d-1$ 個の成分吸引関数の選択に帰着する。 $Y_i = y_i$ を与えた条件の下での Y_j ($j \neq i$) の分布関数を $F_{j|i}(y_{j|i})$ とし, そのハザード関数を

$$h_{j|i} = \frac{f_{j|i}(y_{j|i})}{1 - F_{j|i}(y_{j|i})}$$

とする。

定理 23. 多変量確率変数 \mathbf{Y} が絶対連続な同時分布を持ち, 吸引関数 $a_{|i}(y_i)$, $b_{|i}(y_i) > 0$ にたいして Y_{-i} が (35) を満たすとすると, 吸引関数の j 成分 ($j \neq i$) は

$$\lim_{y_i \rightarrow \infty} F_{j|i}(a_{j|i}(y_i)|y_i) = p_{j|i},$$

ただし $p_{j|i} \in (0, 1)$, で

$$b_{j|i}(y_i) = (h_{j|i}(a_{j|i}(y_i)|y_i))^{-1},$$

を満たす。

系 $a_{j|i}(y_i), b_{j|i}(y_i) > 0$ が定理の条件を満たし, ある定数 $s_{j|i} < \infty$ が存在して

$$\lim_{y_i \rightarrow \infty} \left(\frac{a_{j|i}(y)}{b_{j|i}(y)} \right) = s_{j|i}$$

ならば, $a_{j|i}(y_i) = 0$ にたいして定理が成り立つ. さらに $t_{j|i} > 0$ と y_i によらない任意の定数 $t_{j|i} > 0$ にたいして $b_{j|i}(y_i) = t_{j|i} k_{j|i}(y_i)$ ならば, 定理 23 が $b_{j|i}(y)$ の代わりに $k_{j|i}(y_i)$ を用いて成り立つ.

より具体的に極限分布 $G_{j|i}$ は次のように多様であることが, 理論的諸分布から具体的に示すことができる.

Y が (1) 漸近的に従属の場合 (2) 漸近的独立で正の関連性をもつ場合 (3) ほとんど独立の場合 (4) 負の従属性を持つ場合; がある. それぞれの場合に (a) $G_{j|i}$ が条件付漸近独立でない場合があり, (3), (4) の場合には (b) $G_{j|i}$ が条件付漸近独立となり得る. (1), (2) の場合は Y が完全従属となって $G_{j|i}$ が存在しないことがある.

(3a)(3b) の場合に $G_{j|i}$ として正規分布, Weibull 分布, Gumbel 分布が現われる例がある.

吸引関数は多くの場合に次の形をしている.

$$a_{j|i}(y) = a_{j|i} y + I[a_{j|i} = 0, b_{j|i} < 0](c_{j|i} - d_{j|i} \log(y)), \quad b_{j|i}(y) = y^{b_{j|i}}.$$

これを仮定すれば定数ベクトルの組 $\theta_{j|i} = (a_{j|i}, b_{j|i}, c_{j|i}, d_{j|i})$, (ただし $a_{j|i} = 0, b_{j|i} < 0$ でなければ $c_{j|i} = d_{j|i} = 0$ とする) が未知パラメータである.

モデルの構成と推定

前小節の極限定理に基づいて条件付従属モデルを次のように構成する. Y の周辺分布が標準 Gumbel 分布であるとし, $Y_i = y_i, y_i \rightarrow \infty$, のときに Y_{-i} の条件付極限分布が得られたのであるから, 十分に大きな閾値 u_{Y_i} にたいして

$$\begin{aligned} P\{Y_{-i} < a_{j|i}(y_i) + b_{j|i}(y_i)z_{j|i} | Y_i = y_i\} \\ = P\{Z_{j|i} < z_{j|i} | Y_i = y_i\} = G_{j|i}(z_{j|i}), \quad \forall y_i > u_{Y_i}, \end{aligned}$$

が成り立つと仮定する. $Z_{j|i}$ は標準化した条件付確率変数である. 言い換えれば

$$Y_{-i} = a_{j|i}(y_i) + b_{j|i}(y_i)Z_{j|i}, \quad Y_i = y_i > u_{Y_i}, \quad (36)$$

を仮定する.

推測には多くの障害がある. $G_{j|i}$ の分布型を指定することが困難である. $G_{j|i}$ は互いに関連しているが, そのために生ずる制約条件が簡単ではない. そしてモデルは簡単でなければならない. これらを考慮して Heffernan and Tawn (2004) が提案している推定法は次の通りである. 観測データは確率標本 X_1, \dots, X_n であるとする.

まず X_i の適当な閾値 u_{X_i} を選び, 成分ごとに一般パレート分布を当てはめる. 推定値に基づいて X_1, \dots, X_n を周辺分布が標準 Gumbel 分布に従う変数 Y_1, \dots, Y_n に変換する. $Y_i > u_{Y_i}$ を仮定し, これを満たす観測値の数を $n_{u_{Y_i}}$ とする.

$G_{j|i}$ についての仮定を少なくするために $Y_{-i} | Y_i = y_i$ が 2 次モーメントを持つと仮定すると

$$\begin{aligned} \mu_{j|i}(y) &= a_{j|i}(y) + \mu_{j|i} b_{j|i}(y), \\ \sigma_{j|i}(y) &= \sigma_{j|i} b_{j|i}(y), \end{aligned}$$

となる. 未知パラメータは $\theta_{j|i}$ と $\lambda_{j|i} = (\mu_{j|i}, \sigma_{j|i})$ である. これを解くのに, $Z_{j|i}$ が独立正規分布に従うという仮定を行って

$$Q_{j|i}(\theta_{j|i}, \lambda_{j|i}) = - \sum_{j \neq i} \sum_{k=1}^{n_{u_{Y_i}}} \left(\log(\sigma_{j|i}(y_{i,i,k})) + \frac{1}{2} \left(\frac{y_{j|i,k} - \mu_{j|i}(y_{i,i,k})}{\sigma_{j|i}(y_{i,i,k})} \right)^2 \right)$$

を最大にするようにパラメータを定める. Y_{-i} は独立ではないが, 個別にパラメータを推定することでませる.

以上かなり単純化した推定法であるが, ある程度の裏付けが各段階でなされている.

11 付録

11.1 多変量正則変動関数

正則変動確率変数

実確率変数 X が両裾で、正則変動をして、しかも対称に近い場合を考える。詳しくは

$$\begin{aligned} \bar{F}(x) &= P\{X > x\} = pL(x), \quad L \in R_{-\alpha}, \\ F(-x) &= P\{X \leq -x\} = qL(x), \quad p + q = 1, \end{aligned}$$

とする。実際 t 分布、対数ガンマ分布はこのような裾を持つ。パレート分布は $p = 1, q = 0$ の場合である。上の裾の条件は

$$\left. \begin{aligned} \lim_{x \rightarrow \infty} \frac{P\{X > tx\}}{P\{|X| > x\}} &= pt^{-\alpha}, \\ \lim_{x \rightarrow \infty} \frac{P\{X \leq tx\}}{P\{|X| > x\}} &= qt^{-\alpha}, \end{aligned} \right\} \quad (37)$$

に等しい。この極限から $\bar{\mathbb{R}} \setminus \{0\}$ 上の測度 μ を

$$\mu(dt) = \alpha dt(pt^{-\alpha-1}I[t \in (0, \infty]] + qt^{-\alpha-1}I[t \in [-\infty, 0)])$$

が定義できて、極限(37) が

$$pt^{-\alpha} = \mu(t, \mu], \quad qt^{-\alpha} = \mu[-\infty, -t],$$

と表わされる。これから(37) が

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{P\{x^{-1}X \in (t, \infty]\}}{P\{|X| > x\}} &= pt^{-\alpha} = \mu(t, \infty], \\ \lim_{x \rightarrow \infty} \frac{P\{x^{-1}X \in [-\infty, -t]\}}{P\{|X| > x\}} &= qt^{-\alpha} = \mu[-\infty, -t], \end{aligned}$$

と表わされ、さらに一般化すると

$$\lim_{x \rightarrow \infty} \frac{P\{X \in xA\}}{P\{|X| > x\}} = \mu(A),$$

ただし A は 0 を含まぬ $\mu(\partial A) = 0$ を満たす任意の Borel 集合である。このような測度の収束を vague convergence という。またこのような A にたいして

$$\mu(tA) = \lim_{x \rightarrow \infty} \frac{P\{X = xtA\}}{P\{|X| > tx\}} \frac{P\{|X| > tx\}}{P\{|X| > x\}} = t^{-\alpha} \mu(A), \quad \forall t > 0,$$

が満たされる。2点集合 $\{-1, 1\}$ の上の「極座標」 $\theta \in \{-1, 1\}$ の確率分布

$$P\{\theta = 1\} = p, \quad P\{\theta = -1\} = q,$$

を用いると、任意の $S \subset \{-1, 1\}, \forall t > 0$ にたいして両端を含めた条件付確率が求まる。

$$\lim_{x \rightarrow \infty} \frac{P\{|X| > tx, X/|X| \in S\}}{P\{|X| > x\}} = t^{-\alpha} P\{\theta \in S\}.$$

多変量正則変動関数

$\mathbf{X} \in \mathbb{R}_{\geq 0}^d$ の生存関数が正則変動であることを次のように定義する. 超平面 $S^{d-1} = \{\mathbf{x} : |\mathbf{x}| = |\mathbf{x}_1| + \dots + |\mathbf{x}_d| = 1\}$ 上のある確率変数 Θ が存在して

$$\lim_{x \rightarrow \infty} \frac{P\{|\mathbf{X}| > tx, \mathbf{X}/|\mathbf{X}| \in S\}}{P\{|\mathbf{X}| > x\}} = t^{-\alpha} P\{\Theta \in S\} \quad (38)$$

であるとき \mathbf{X} の分布が指数 $-\alpha$ の正則変動であると言う. $P\{\Theta \in S\}$ は \mathbf{X} のスペクトル測度と呼ばれる. 上式が示すように $|\mathbf{X}|$ と $\mathbf{X}/|\mathbf{X}|$ は漸近的に独立である.

(38) の右辺を集合

$$A(t, S) = \{\mathbf{x} : |\mathbf{x}| > t, \mathbf{x}/|\mathbf{x}| \in S\}$$

の測度 μ とみなすと, (38) は

$$\lim_{x \rightarrow \infty} \frac{P\{x^{-1}\mathbf{X} \in A(t, S)\}}{P\{|\mathbf{X}| > x\}} = \mu(A(t, S)) = t^{-\alpha} \mu(A(1, S)),$$

となる. $\mu(A(t, S))$ から $\mathbb{R}^d \setminus \{0\}$ 上の Radon 測度 (コンパクト集合上の有限測度) μ が定まり, (38) は 0 を含まぬ $\mu(\partial A) = 0$ を満たす任意の集合 $A \subset \mathbb{R}_{\geq 0}^d$ に関する vague convergence

$$\lim_{x \rightarrow \infty} \frac{P\{x^{-1}\mathbf{X} \in A\}}{P\{|\mathbf{X}| > x\}} = \mu(A)$$

と同等であり, μ は

$$\mu(tA) = t^{-\alpha} \mu(A)$$

を満たす.

これを多変量極値理論に適用すれば定理 16 が証明できる.

11.2 超単純多変量極値分布

11.2.1 2 変量独立 Fréchet 分布

標準フレシェ (Fréchet) 分布の分布関数, 確率密度関数は

$$G(x) = \exp(-1/x), \quad g(x) = x^{-2} \exp(-1/x), \quad x \in \mathcal{R}_+,$$

である.

Y が標準指数分布に従えば, $1/Y$ が標準 Fréchet 分布に従う.

この分布関数は尺度パラメータを除いて

$$G^t(tx) = G(x), \quad \forall t \in \mathcal{R}_+,$$

によって特徴付けられる.

実際, $-\log G(x) = V(x)$ とすれば $tV(tx) = V(x)$. これを t で微分すれば

$$\begin{aligned} V(tx) + txV'(tx) &= 0 \Leftrightarrow \frac{tV'(tx)}{V(tx)} = -\frac{1}{x} \\ \Leftrightarrow \frac{d}{dx} \log V(x) &= -\frac{d}{dx} \log x + c \Leftrightarrow V(x) = c/x, \quad c > 0. \end{aligned}$$

注意 負の標準 Fréchet 分布 (すなわち負の標準 Weibull 分布) は 0 が分布の上限であるが, Gumbel 分布の MDA に属する.

独立 2 変量標準 Fréchet 分布 独立な 2 変量標準 Fréchet 分布 (多変量極値分布の一つ) の諸性質は "2 変量正規分布中心の相関" を教育されたものにとっては, 新鮮な印象を与える. (X_1, X_2) が独立な 2 変量標準 Fréchet 分布に従うとき

$$V = \max(X_1, X_2), \quad W = X_1/(X_1 + X_2),$$

の同時分布関数を求める. $[W = X_1^{-1}/(X_1^{-1} + X_2^{-1})$ が一様分布に従うことは良く知られている.]

$$V = X_2 \iff X_2 > X_1 \iff W < 1/2,$$

$$P\{W \leq w < 1/2\} = P\{X_1 \leq \frac{w}{1-w} X_2\} = \exp\left(-\frac{1-w}{wx_2}\right)$$

$$F(v, w) = P\{W \leq w, V \leq v\} = \int_0^v \exp\left(-\frac{1-w}{wx_2}\right) e^{-\frac{1}{x_2^2}} \frac{dx_2}{x_2^2}$$

$$= we^{-\frac{1}{wv}} = 1 - F(v, 1-w), \quad w < 1/2, \quad 0 < v < \infty.$$

$$F(v) = 2F(v, 1/2) = e^{-2/v}, \quad F(w) = w, \quad f(v, w) = f(v, 1-w) = \frac{1}{w^2 v^3} e^{-\frac{1}{wv}},$$

$$f(v|w) = f(v|1-w) = \frac{1}{wv^2(wv+1)}, \quad f(w|v) = f(1-w|v) = \frac{1}{2vw^2} e^{-\frac{1}{v}(\frac{1}{wv}-2)}.$$

$f(w|v)$ は標準 Fréchet 分布の尺度を変え, 上側を打ち切ったものとなっている. 別紙図 2 変量独立 Fréchet 分布参照.

積, 最大値, 最小値の演算

実数空間における演算 \vee, \wedge の演算を調べる. 実数 a, b にたいして $a \vee b := \max(a, b)$, $a \wedge b := \min(a, b)$ とする. \cdot は積で省略する. c, x も実数とすると,

$$a \vee b = b \vee a, \quad a \wedge b = b \wedge a,$$

$$(a \vee b) \vee c = a \vee (b \vee c) = a \vee b \vee c, \quad (a \wedge b) \wedge c = a \wedge (b \wedge c) = a \wedge b \wedge c,$$

$$x > 0 \text{ のとき } (ax) \vee (bx) = (a \vee b)x, \quad (ax) \wedge (bx) = (a \wedge b)x.$$

ベクトルの最大値, 最小値を成分ごとの最大値, 最小値で定義する. たとえば 2 次元ベクトルにたいして,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \mathbf{a} \vee \mathbf{b} = \begin{bmatrix} a_1 \vee b_1 \\ a_2 \vee b_2 \end{bmatrix}, \quad \mathbf{a} \wedge \mathbf{b} = \begin{bmatrix} a_1 \wedge b_1 \\ a_2 \wedge b_2 \end{bmatrix},$$

$$\mathbf{a} \vee \mathbf{b} = \mathbf{a}, \quad \text{and} \quad \mathbf{a} \wedge \mathbf{b} = \mathbf{b}, \quad \text{if} \quad \mathbf{a} \geq \mathbf{b}.$$

ただしベクトルの大小比較は, 全成分に関する比較である.

m 次元の n 個のベクトルにたいしては

$$\mathbf{a}_1 \vee \cdots \vee \mathbf{a}_n = \bigvee_{j=1}^n \mathbf{a}_j, \quad \max_{1 \leq j \leq n} a_{i,j}, \quad i = 1, \dots, m$$

と表せるが, 最後の表現が分かりやすいであろう. 最小値についても同様である. $\bigvee_{j=1}^n \mathbf{a}_j$, $\mathbf{a}_j \in \mathbb{R}^m$, は高々 m 個のベクトルの最大値として表せる. ベクトルの最大値をベクトルの和と対比すると, 次の拡張が考えられる.

$m \times n$ 行列 $A = (a_{i,j})$ によるベクトルの線形変換 $\mathbf{y} = A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$ に対応して,

$$y_i = \max_{1 \leq j \leq n} a_{i,j} x_j, \quad i = 1, \dots, m$$

という変換を考える. ただし今後は A を正値行列に限る.

$a_{i,j} > 0$ とすると, 連立線形不等式

$$\max_j a_{i,j} x_j \leq y_i, \quad i = 1, \dots, m$$

の解は

$$x_j \leq \min_{1 \leq i \leq m} y_i / a_{i,j}, \quad i = 1, \dots, n,$$

である. 実際

$$\begin{aligned} \bigcap_{i=1}^n [\max_j a_{i,j} x_j \leq y_i] &= \bigcap_{i=1}^n \bigcap_{j=1}^m [a_{i,j} x_j \leq y_i] \\ &= \bigcap_{j=1}^m \bigcap_{i=1}^n [x_j \leq y_i / a_{i,j}] = \bigcap_{j=1}^m [x_j \leq \min_i y_i / a_{i,j}]. \end{aligned}$$

演算 “+ , \vee ” の代数学は max-plus algebra と呼ばれ, 微分方程式の超離散化で用いられている. 例えば広田良吾, 藤原大輔 (2003) 差分と離散化, 共立出版, 参照. 計画数学, 計算数学などの分野では tropical algebra と呼ばれ, 最近急速に話題が広がっている.

11.2.2 超単純多変量極値分布

Z_1, \dots, Z_n を標準 Fréchet 分布に従う独立確率変数数列, $A = (a_{i,j}), a_{i,j} > 0$ を $\sum_{j=1}^n a_{i,j} = 1, i = 1, \dots, m$ を満たす $m \times n$ 正要素行列とし,

$$X = (X_1, \dots, X_m), \quad X_i = \max_{1 \leq j \leq n} a_{i,j} Z_j, \quad i = 1, \dots, m$$

とする. あるいは

$$X = \bigvee_{j=1}^n Z_j \mathbf{a}_j, \quad A = [\mathbf{a}_1, \dots, \mathbf{a}_n].$$

X の分布関数は m 変量極値分布に従う. その分布関数は

$$\begin{aligned} F(\mathbf{x}) = F(x_1, \dots, x_m) &= P\{\max_j a_{1,j} Z_j \leq x_1, \dots, \max_j a_{m,j} Z_j \leq x_m\} \\ &= \prod_{j=1}^n P\{Z_j \leq \min(x_1/a_{1,j}, \dots, x_m/a_{m,j})\} \\ &= \prod_{j=1}^n \exp(-1/\min(x_1/a_{1,j}, \dots, x_m/a_{m,j})) \\ &= \exp(-\sum_{j=1}^n \max(a_{1,j}/x_1, \dots, a_{m,j}/x_m)). \end{aligned} \tag{39}$$

と表せ, これが最大値安定であることが確かめられる:

$$F^r(rx_1, \dots, rx_m) = F(x_1, \dots, x_m), \quad \forall r > 0.$$

$\sum_{j=1}^n a_{i,j} = 1$ の条件から, たとえば X_1 の周辺分布関数が,

$$F(x_1, \infty, \dots, \infty) = \exp(-\sum_{j=1}^n \max(a_{1,j}/x_1, 0, \dots, 0)) = \exp(-1/x_1).$$

つまり標準 Fréchet 分布となり, F は本文で述べる単純 m 変量極値分布である. F を超単純多変量極値分布と呼ぶことにする. 単純多変量極値分布の中でもっとも単純な場合である. $F(\mathbf{x})$ のいくつかの成分を ∞ とした結果は同じ型の分布関数になるから, すべての低次元周辺分布が超単純多変量極値分布である.

(39) の最終式における各 j にたいする \max でどの項が採られるかは, $\mathbf{x}/|\mathbf{x}|, |\mathbf{x}| = \sum_{j=1}^m x_j$, と $a_j/|a_j|$ の関係で定まる. m 次元単位単体を $\Delta_m = \{w \in \mathbb{R}_{\geq 0}^m : |w| = 1\}$ とすると, Δ_m の $\binom{m}{2}$ 本の辺および, $a_j/|a_j|$ と Δ_m の頂点とを結ぶ m 本の線分で作る m 個の多面体のどれに $\mathbf{x}/|\mathbf{x}|$ が入るかで定まる. たとえば $x_1 \ll x_i (i \neq 1)$ のときには面 $w_1 = 0$ の近くにあり, $F(\mathbf{x}) = \exp(-1/x_1)$ となる.

超単純多変量極値分布の性質

特に $m = n = 2$ のときの $X = (X_1, X_2)$ の分布を理解するために, $W_X = X_1/(X_1 + X_2)$ の分布関数を求める. $n = 2$ のときの $W_Z = Z_1/(Z_1 + Z_2)$ は $(0, 1)$ 一様分布に従う. 一般性を失うことなく (成分番号の交換により) $a_{1,1}/a_{1,2} < a_{2,1}/a_{2,2}$ と仮定する. これは $0 < q_1 < q_2 < 1, q_i = a_{1,i}/(a_{1,i} + a_{2,i})$ と同等である.

$$\begin{aligned} W_X = q_1 &\iff Z_1 a_1 > Z_2 a_2 \iff W_Z a_1 > (1 - W_Z) a_2, \\ &\iff W_Z > \max(a_{1,2}, a_{2,2}) = a_{1,2}, \end{aligned}$$

これから,

$$P\{W_X = q_1\} = 1 - a_{1,2} = a_{1,1}.$$

同様の計算で $P\{W_X = q_2\} = a_{2,2}$ である.

$\{q_1 < W_X < q_2\} \iff \{X_1 = a_{1,2}Z_2 \text{ \& } X_2 = a_{2,1}Z_1\}$ の確率は

$$1 - (a_{1,1} + a_{2,2}) = a_{1,2} - a_{2,2} = a_{2,1} - a_{1,1}.$$

さらにこのとき,

$$\begin{aligned} W_X \leq w &\iff \frac{X_1}{X_2} \leq \frac{w}{1-w} \iff \frac{Z_2}{Z_1} \leq \frac{w}{1-w} \frac{a_{2,1}}{a_{1,2}} \\ &\iff W_Z \geq \frac{(1-w)a_{1,2}}{(1-w)a_{1,2} + wa_{2,1}}. \end{aligned}$$

あるいは $W_X = X/|X|$ の分布関数を $H(w)$ とすれば,

$$H(w) = \begin{cases} 0, & 0 < w < q_1, \\ wa_{2,1}/((1-w)a_{1,2} + wa_{2,1}), & q_1 \leq w < q_2, \\ 1, & q_2 \leq w < 1. \end{cases}$$

つまり W_X は 2 点 q_1, q_2 で離散確率, 区間 (q_1, q_2) で連続確率密度をもつ.

$m = 2, n = 3$ の場合を考える. $n = 2$ の場合と同様に $q_i = a_{1,i}/(a_{1,i} + a_{2,i}), i = 1, 2, 3, q_1 < q_2 < q_3$ を仮定する. $W_X = q_i, i = 1, 2, 3$, が止の確率を持ち, $q_i < W_X < q_{i+1}, i = 1, 2$, が連続確率密度をもつ. $q_1 < W_X < q_2$ となるのは $X_1 = a_{1,2}Z_2 \text{ \& } X_2 = a_{2,1}Z_1$ の場合だけでなく, $X_1 = a_{1,3}Z_3 \text{ \& } X_2 = a_{3,1}Z_1$ の場合もある. $q_2 < W_X < q_3$ についても同様である. W_X の分布関数は次のようになる.

$$H(w) = \begin{cases} 0, & 0 \leq w < q_1, \\ a_{1,1}, & w = q_1, \\ 1 - \frac{(a_{1,2} + a_{1,3})(1-w)}{(a_{1,2} + a_{1,3})(1-w) + a_{2,1}w}, & q_1 \leq w < q_2, \\ 1 - \frac{a_{1,3}a_{2,2}}{(a_{1,2} + a_{1,3})a_{2,2} + a_{1,2}a_{2,1}}, & w = q_2, \\ \frac{a_{2,1}w}{(a_{2,1} + a_{2,2})w + a_{1,3}(1-w)} - \frac{a_{1,2}a_{2,1}}{(a_{2,1} + a_{2,2})a_{1,2} + a_{1,3}a_{2,2}}, & q_2 \leq w < q_3, \\ 1, & q_3 \leq w \leq 1. \end{cases}$$

$H(w)$ および $h(w) = d/dH(w)$ の図を別紙図超単純図に示す. 別紙図超単純にシミュレーションの結果がある.

$m = 2, n \geq 3$ のとき W_X の離散確率 (point mass) は

$$P\{W_x = q_j\} = \frac{1}{A_{1,j}/a_{1,j} + A_{2,j}/a_{2,j} + 1}, \quad q_j = a_{1,j}/(a_{1,j} + a_{2,j}), \quad A_{1,j} = \sum_{k=j+1}^n a_{1,k}, \quad A_{2,j} = \sum_{k=1}^{j-1} a_{2,k},$$

である. その特別な場合として

$$P\{W_x = q_1\} = \frac{1}{A_{1,1}/a_{1,1} + 1} = a_{1,1}, \quad P\{W_x = q_n\} = \frac{1}{A_{2,n}/a_{2,n} + 1} = a_{2,n}.$$

$$A = \begin{bmatrix} 1, & 2, & \dots & n \\ n, & n-1, & \dots & 1 \end{bmatrix}$$

とするときの $P\{W_x = q_j\}$ およびその和のリストは表の通りである.

表 5: 超単純多変量 離散確率例

3	0.1667	0.2500	0.1667								0.583
4	0.1000	0.1714	0.1714	0.1000							0.543
5	0.0667	0.1212	0.1429	0.1212	0.0667						0.519
6	0.0476	0.0893	0.1143	0.1143	0.0893	0.0476					0.502
7	0.0357	0.0682	0.0915	0.1000	0.0915	0.0682	0.0357				0.491
8	0.0278	0.0536	0.0741	0.0855	0.0855	0.0741	0.0536	0.0278			0.482
9	0.0222	0.0432	0.0609	0.0727	0.0769	0.0727	0.0609	0.0432	0.0222	0.0000	0.475
10	0.0182	0.0356	0.0507	0.0621	0.0682	0.0682	0.0621	0.0507	0.0356	0.0182	0.470

一般の (m, n) にたいする W_X の分布は複雑である. $\{a_1/|a_1|, \dots, a_n/|a_n|\}$ の凸包の境界および内部に分布し, これらの点で離散確率, これらの点を結ぶ線分の上で 1 次元確率密度, など 1 次元から m 次元の確率密度関数を含んでいる. A の各行は離散確率とみなせる. これを一般の確率分布とすることにより, 単純多変量極値分布 (21) が得られる. 逆に任意の単純多変量極値分布を, n を増やすことにより, 超単純多変量極値分布で近似できる.

Pickands の従属関数

$m = 2$ の場合, 超単純多変量極値分布の指数部

$$V(x) = \sum_{j=1}^n \max\left(\frac{a_{1,j}}{x_1}, \frac{a_{2,j}}{x_2}\right), \quad x \in \mathbb{R}_{\geq 0}^m \setminus \mathbf{0},$$

の変数を逆数に変え, 次のように変形する. 再び一般性を失うことなく,

$$\frac{a_{1,1}}{a_{2,1}} < \dots < \frac{a_{1,n}}{a_{2,n}} \iff q_1 < \dots < q_n, \quad q_j = \frac{a_{1,j}}{a_{1,j} + a_{2,j}}$$

を仮定する. $t = 1/x$ とし,

$$|x|V(x) = |t|^{-1}V(1/t) = |t|^{-1} \sum_{j=1}^n \max(t_1 a_{1,j}, t_2 a_{2,j}),$$

の変数を, さらに $u = t_2/(t_1 + t_2) = x_2^{-1}/(x_1^{-1} + x_2^{-1})$ に変え, 記号を濫用して

$$A(u) = \sum_{j=1}^n \max((1-u)a_{1,j}, ua_{2,j})$$

とする.

$$(1-u)a_{1,j} \geq ua_{2,j} \iff u \leq q_j = \frac{a_{1,j}}{a_{1,j} + a_{2,j}}$$

であるから

$$A(u) = \sum_{u \leq q_j} (1-u)a_{1,j} + \sum_{u > q_j} ua_{2,j}$$

となる. u が 1 区画動くとき, 項が $(1-u)a_{1,j}$ から $ua_{2,j}$ に変わるので折れ線の勾配が増加し $A(u)$ は凸関数である. $n=3$ のとき

$$A(u) = \begin{cases} 1-u, & 0 \leq u \leq q_1, \\ ua_{2,1} + (1-u)(a_{1,2} + a_{1,3}), & q_1 \leq u < q_2, \\ q_2(1-a_{2,3}) + (1-q_2)a_{1,3}, & u = q_2, \\ u(a_{2,1} + a_{2,2}) + (1-u)a_{1,3}, & q_2 < u < q_3, \\ u, & q_3 \leq u \leq 1. \end{cases}$$

$m=2$ の場合には, 任意の Pickands 従属関数を, 超単純多変量極値分布の Pickands 従属関数で近似できる.

別紙図超単純図参照.

11.2.3 多変量指数分布

指数分布に関して, Fréchet 分布と双対な結果が得られる. Z_1, Z_2, \dots を標準指数分布 (標準 Weibull 分布) に従う独立変数列, $a_{i,j} > 0$ とし, $Y_i = \min_{1 \leq j \leq n} Z_j/a_{i,j}$, $i=1, \dots, m$ とすると, (Y_1, \dots, Y_m) の生存関数は多変量指数分布 (超単純多変量最小極値分布) に従う. 実際

$$\begin{aligned} S(y_1, \dots, y_m) &= P\{\min_j Z_j/a_{1,j} \geq y_1, \dots, \min_j Z_j/a_{m,j} \geq y_m\} \\ &= \prod_{j=1}^n P\{Z_j \geq y_1 a_{1,j}, \dots, Z_j \geq y_m a_{m,j}\} = \prod_{j=1}^n P\{Z_j \geq \max_{1 \leq i \leq m} y_i a_{i,j}\} \\ &= \exp\left(-\sum_{j=1}^n \max_{1 \leq i \leq m} a_{i,j} y_i\right), \end{aligned}$$

$\sum_{j=1}^n a_{i,j} = 1$, $i=1, \dots, m$ であると, たとえば X_1 の周辺生存関数が,

$$S(x_1, 0, \dots, 0) = \exp\left(-\sum_{j=1}^n \max(a_{1,j} y_1, 0, \dots, 0)\right) = \exp(-y_1).$$

標準指数分布となる. 同様に (Y_1, \dots, Y_m) の任意の周辺分布は多変量指数分布に従う.

参考文献

- [1] Akahira, M., and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency*, Lecture Notes in Statistics 7, Springer.
- [2] Athreya, K.B. and Fukuchi, J. (1994). Asymptotic theory for bootstrapping the extremes, *Commun. Statist. Theory Meth.*, 22, 15-30.
- [3] Balkema, G. and Embrechts, P. (2007). *High Risk Scenarios and Extremes: A geometric approach*, Zurich Lectures in Advanced Mathematics, European Mathematical Society, Zürich, Switzerland.

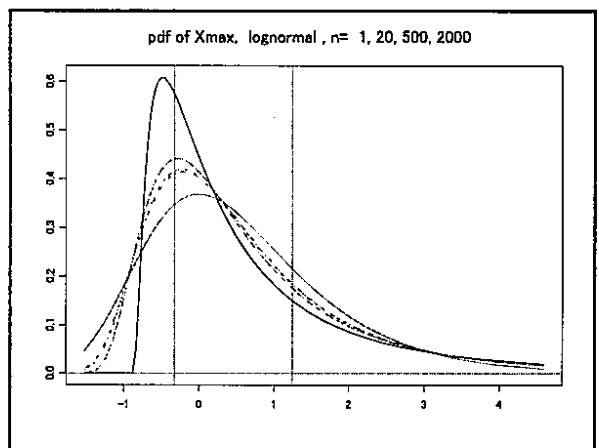
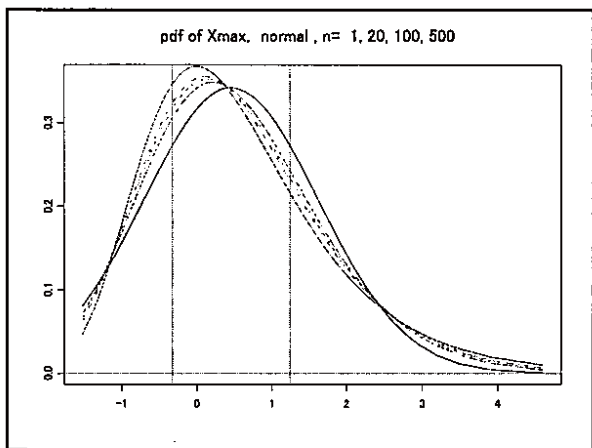
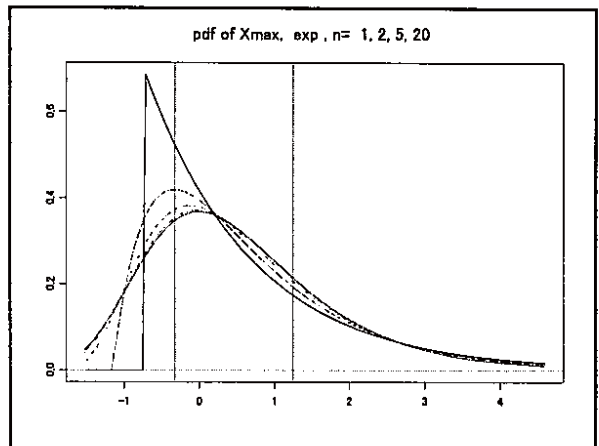
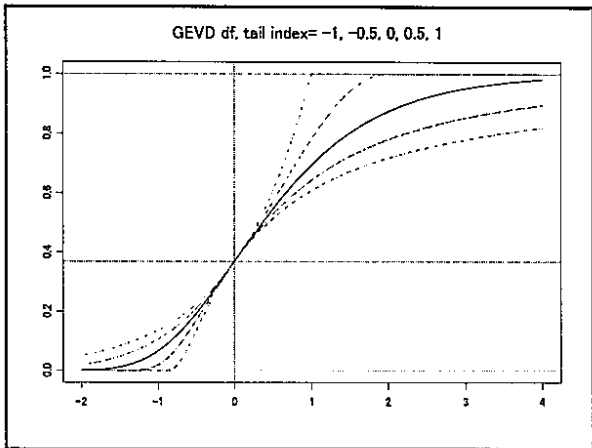
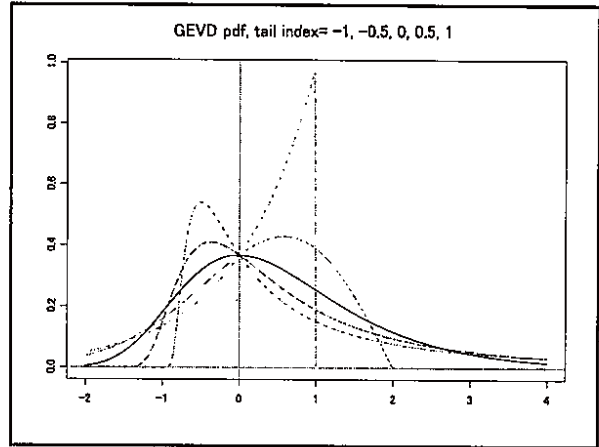
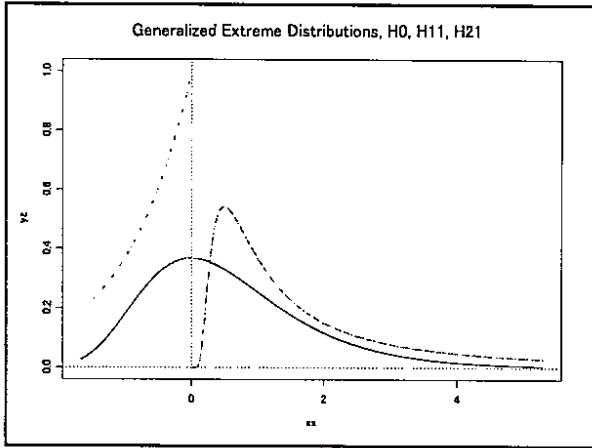
- [4] Balkema, A. A. and de Haan, L. (1974). Residual life time at great age, *Ann. Probab.* **2**, 792–804.
- [5] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Exteremes*. Wiley, Chichester, UK.
- [6] Bertail, P., Politis, D.N. and Romano, J.P. (1999). On subsampling estimators with unknown rate of convergence, *Journal of the American Statistical Society*, **94** (446), 569–579.
- [7] Bortot, P., Coles, S.G. and Sisson, S.A. (2007). Inference for stereological extremes, *Journal of the American Statistical Society*, **102** (477), 84–92.
- [8] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Springer, New York, NY.
- [9] Coles, S. (2001). *An Introduction to Statical Modeling of Extreme Values*. Springer, London, UK.
- [10] Coles, S. (2003). *The use and misuse of extreme value models in practice*, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, 79–100.
- [11] Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events, *Journal of the Royal Statistical Society*, **B 53**, 377–392.
- [12] Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data, *Appl. Statist.*, **45**, 463–478.
- [13] Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process, *Journal of the Royal Statistical Society*, **B 58**, 329–347.
- [14] Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds, *Appl. Statist.*, **43**, 139–157.
- [15] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, 3rd. ed., Wiley, New York, NY.
- [16] Davis, R.A., and Resnick, S.I. (1993). Prediction of stationary max-stable processes, *Annals of Appl. Probab.*, **3**, 497–525.
- [17] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with Discussion), *J. R. Statist. Soc.*, **B52**, 393–442.
- [18] de Haan, L. (1984). A spectral representation for max-stable processes, *Ann. Probab.*, **12**, 1194–1204.
- [19] de Haan, L. (1985). Extremes in higher dimensions: the model and some statistics, *Bulletin of the International Statistical Institute*, **51**, Amsterdam, 185–192.
- [20] de Haan, L. (1990). Fighting the arch-enemy with mathematics, *Statistica Neerlandica*, **44**, 45–68.
- [21] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory, An Introduction*, Springer, New York, NY.
- [22] de Haan, L. and Lin, T. (2001). On convergence towards an extreme value distributions in $C[0, 1]$, *Ann. Probab.*, **29**, 467–483.
- [23] de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes, *Z. Wahrscheinlichkeitstheorie verw. Gebite*, **40**, 317–337.
- [24] de Haan, L. and de Ronde, J. (1998). Sea and wind: Multivariate extremes at work, *Extremes*, **1**, 7–45.
- [25] de Haan, L. and Rootzén, H. (1993). On the estimation of high quantiles, *J. Statist. Planning and Inferences*, **35**. 1–13.
- [26] Deheuvels, P. (1983). Point processes and multivariate extreme values, *J. Multiv. Anal.*, **13**, 257–272.

- [27] Embrechts, P. (2003). Extremes in economics and the economics of extremes, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, 169–183.
- [28] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, Germany.
- [29] Falk, M., Hüsler, J. and Reiss, R.-D. (2004). *Laws of Small Numbers: Extremes and Rare Events*, 2nd. ed., Birkhäuser.
- [30] Finkenstädt, B. and Rootzén, H. Eds. (2003). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman and Hall/CRC, Boca Raton, FL. (available on line at <http://www.stat.unc.edu/postscript/rs/semstatrls.pdf>)
- [31] Fougères, A. (2003). Multivariate extremes, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, 373–388.
- [32] Geffroy, J. (1958/1959). Contribution à l'étude de la théorie des valeurs extrêmes, *Publ. Inst. Statist. Univ. Paris*, 7/8, 37-185.
- [33] Geluk, J.L., and de Haan, L. (1980). *Regular Variation, Extension and Tauberian Theorems*, CWI Tracts 40, Centre for Mathematics and Computer Science.
- [34] Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University, Springer, 河田竜夫, 岩井重久, 加瀬滋男訳「極値統計学」生産技術センター新社 (再刊版).
- [35] Heffernan, J.E., and Tawn, J. A. (2001). Extreme values analysis of a large designed experiment: A case study in bulk carrier safety, *Extreme*, 4, 359–378.
- [36] Heffernan, J. E., and Tawn, J. A., (2004). A conditional approach for multivariate extreme values (with discussions), *J. R. Statist. Soc. B*, 66, 497–546.
- [37] Heffernan, J. E., Tawn, J. A., and Zhang, Z. (2007) Asymptotic (in)dependent multi-variate maxima of moving maxima processes, *Extremes*, 10, 57–82.
- [38] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, 3, 1163–1174.
- [39] Hosking, J.R.M. (1990). L -moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society*, B 52, 105–124.
- [40] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, 29, 339–349.
- [41] Itoh, Y. and Mahmoud, H. M. (2005). Age statistics in the Moran population model, *Statist. Probab. Letters*, 74, 21-30.
- [42] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London, UK.
- [43] Joe, H., Smith, R. L., and Weissman, I. (1992). Bivariate threshold methods for extremes, *Journal of the Royal Statistical Society*, B 54, 171–183.
- [44] Klüppelberg, C. (2003). Risk management with extreme value theory, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, 101–168.
- [45] Kolev, N., dos Anjos, U., and Mendes, B.V.M.(2003). Copulas: A review and recent developments, *Statistical Models*, 22, 617–660.
- [46] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions*, Imperial College Press.
- [47] Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series*, Springer, New York, NY.
- [48] McNeil, J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management*, Princeton University Press.

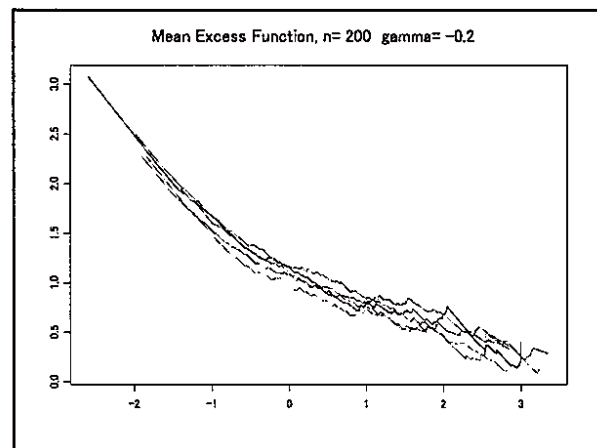
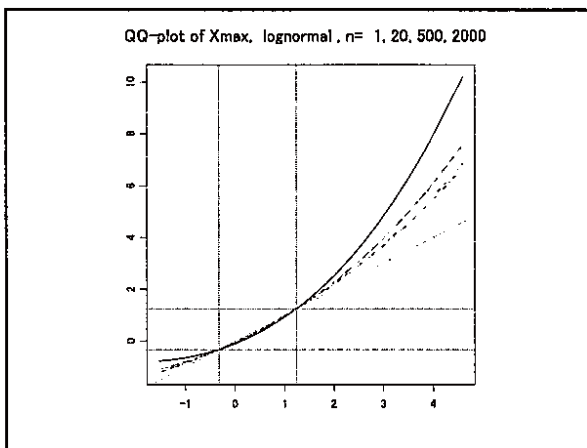
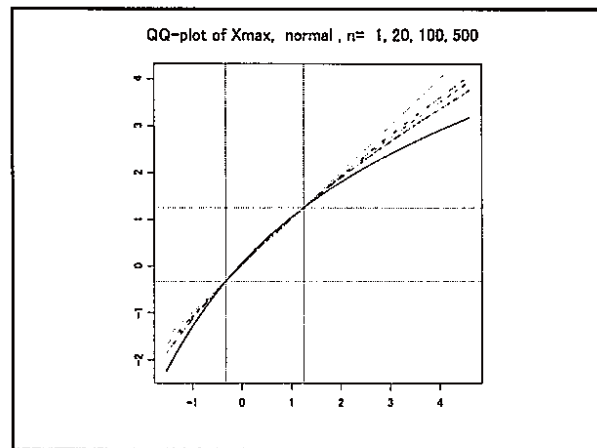
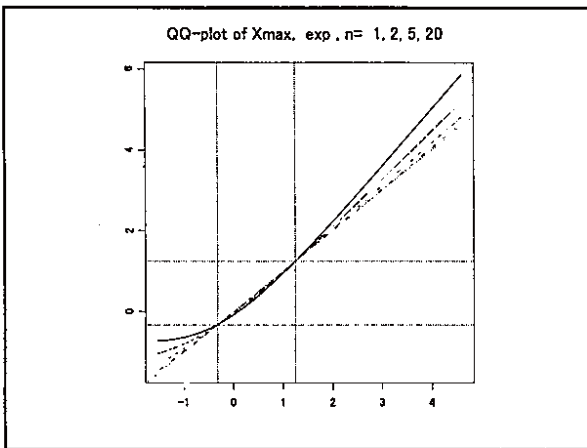
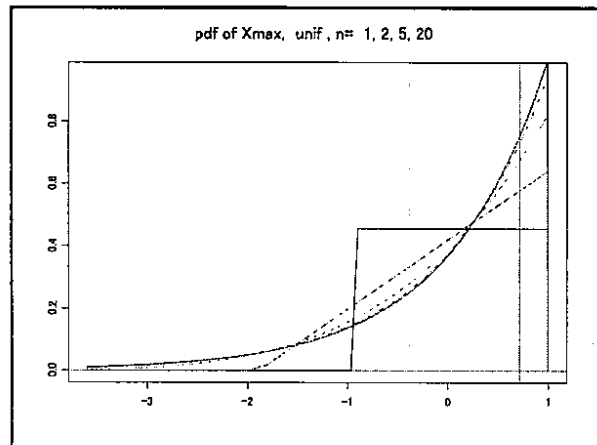
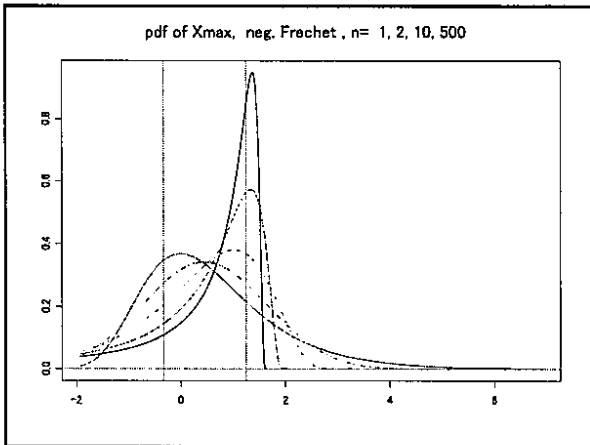
- [49] Mikosch, T. (2003). Modeling dependence and tails of financial time series, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, 185–286.
- [50] Nandagopalan, S. (1983) On the multivariate extremal index, *J. of Research, National Inst. of Standards and Technology*, **99**, 543–550.
- [51] Nelsen, R. B. (2006) *An Introduction to Copula*, 2nd ed., Springer;
- [52] Onoyama, T., Sibuya, M. and Tanaka, H. (1984). Limit distribution of the minimum distance between independent and identically distributed d -dimensional random variables, Tiago de Oliveira, J. Ed., *Statistical Extremes and Applications*, D. Reidel, Dordrecht, 549–562.
- [53] Pickands, J. (1975). The two-dimensional poisson process and extremal processes. *Journal of Applied Probability*, **8**, 547–560.
- [54] Pickands, J. (1981). Multivariate extreme value distributions, *Bulletin of the International Statistical Institute*, **53**, Buenos Aires, 859–878.
- [55] Prescott, P. and Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme value distribution, *Biometrika*, **67**, 723–724.
- [56] Reiss, R. -D. (1993). *A Course on Point Processes*, Springer, New York, NY.
- [57] Resnick, S. (2003). Modeling data networks, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman and Hall/CRC, 287–371.
- [58] Sibuya, M. (1960). Bivariate extreme statistics, *Ann. Inst. Statist. Math.*, **11**, 195–210.
- [59] Sibuya, M. (1999). Wicksell's corpuscle problem, in Kotz, S., Read, C. B. and Banks, D. L., Eds., *Encyclopedia of Statistical Sciences*, Update Volume 3, Wiley, New York, N.Y., 809.
- [60] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.
- [61] Smith, R. L. (2003). Statistics of extremes, with applications in environment, insurance, and finance, in Finkenstädt and Rootzén, Eds., *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman and Hall/CRC, 1–78.
- [62] Smith, R.L. (2004). Statistics of extremes, with applications in the environment, insurance and finance, in Finkenstadt and Rootzén, Eds, *Extreme Value in Finance, Telecommunications and the Environment*, 1–78.
- [63] Smith, R.L., Tawn, J.A., and Coles, S.G. (1997). Markov chain models for threshold exceedances, *Biometrika*, **84**, 249–268.
- [64] Smith, R.L., and Weissman, I. (1996) Characterization and estimation on the multivariate extremal index, Manuscript, UNC.
- [65] Stephenson, A., and Gilleland, E. (2006). Software for the analysis of extreme events: The current state and future directions, *Extremes*, **8**, 67–109.
- [66] Strawderman, R.L. and Zelterman, D. (1998). A semiparametric bootstrap for simulating extreme, in Balakrishnan, N. and Rao, C.R., Eds. *Handbook of Statistics*, **16**, Elsevier Science B. V., 441–462.
- [67] Tajvidi, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions, *Extremes*, **6**, 111–123.
- [68] Takahashi, R. (1987). Some properties of multivariate extreme value distributions and multivariate tail equivalence, *Ann. Inst. Statist. Math.*, **A**, **39**, 637–647.
- [69] Takahashi, R. (1994a). Domains of attraction of multivariate extreme value distributions, *J. Res. National Inst. Standards Tech.*, **99**, 551–554.

- [70] Takahashi, R. (1994b). Asymptotic independence and perfect dependence of vector components of multivariate extreme statistics, *Statist. Probab. Lett.*, **19**, 19–26.
- [71] Takahashi, R. and Sibuya, M. (2002). Metal fatigue, Wickwell transform and extreme values, *Applied Stochastic Models in Business and Industry*, **18**, 301–302.
- [72] Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation, *Biometrika*, **75**, 397–415.
- [73] Tiago de Oliveira, J. (1958). Extremal distributions, *Rev. Fac. Sci. Lisboa*, Ser. A, **7**, 215–227.
- [74] von Mises, R. (1936). La distribution de la plus grande de n valeurs, *Rev. Math. Union Interbalcanique*, **1**, 141–160.
- [75] Zhang, Z. (2002) Multivariate extremes, max-stable process estimation and dynamic financial modeling, Doctoral thesis, Department of Statistics, University of North Carolina.
- [76] Zhang, Z. (2007) Quotient correlation: a sample based alternative to Pearson's correlation, *Ann. Statist.*, to appear.
- [77] Zhang, Z. (2008) On approximating max-stable processes and constructing extremal copula functions, submitted.
- [78] Zhang, Z, and Huang, R. (2006) Extremal financial risk models and portfolio evaluation The behavior of multivariate maxima of moving maxima processes, *Computational Stat. & Data Analysis*, **51**, 2313–2338.
- [79] Zhang, Z, and Smith, R. (2004) The behavior of multivariate maxima of moving maxima processes, *J. Appl. Probab.*, **41**, 1113–1123.
- [80] Zhang, Z, and Smith, R. (2007) On the estimation and application of max-stable processes, Univ. Wisconsin, Madison,
- [81] 岡健一 (2007). サブサンプリングによる極値解析：金融リスク管理への応用, *日本統計学会誌*, **37-1**, 179–200.
- [82] 神田順, 西嶋一欽 (2004). 多変量極値分布を用いた多地点強風および地震危険度解析, *統計数理*, **52**, 151–173.
- [83] 渋谷政昭・華山宣胤 (2004). 年齢時代区分データによる超高年齢寿命分布の推測, *統計数理*, **52-1**, 117–134.
- [84] 志村隆彰 (2007). 正則変動性と確率論, Center for International Research on the Japanese Economy, Research Report Series, CIRJE-R-6, 4–31, 東京大学, 経済学研究科.
- [85] 高橋倫也, 渋谷政昭 (2004). 上位 r 個の観測値に基づく確率点の推定, *統計数理*, **52-1**, 93–116.
- [86] 塚原英敦 (2003). 接合分布関数とその応用—統計的従属性と1次元周辺分布を所与とした多変量モデリング, *応用統計学*, **32**, 77–87.
- [87] 吉原健一 (2004). 弱従属性をもつデータに基づく極値統計の最近の話題, *統計数理*, **52**, 25–43.

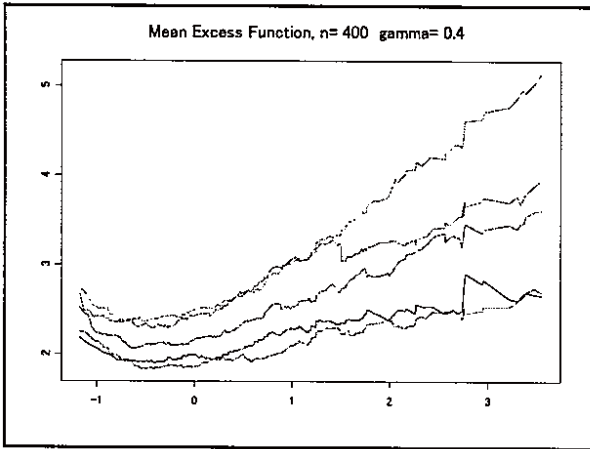
一般極值分布



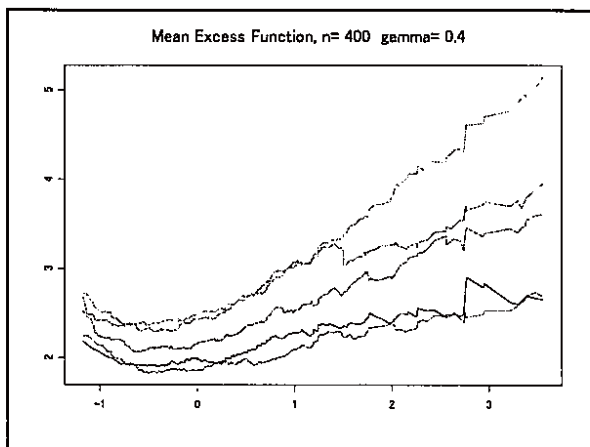
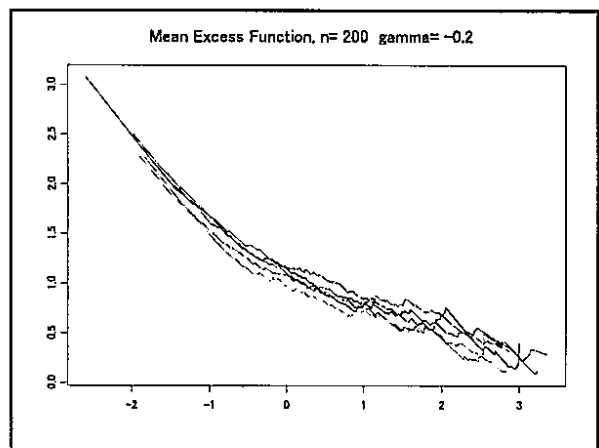
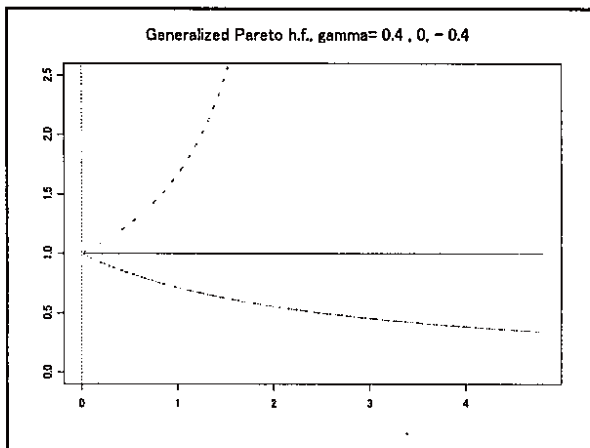
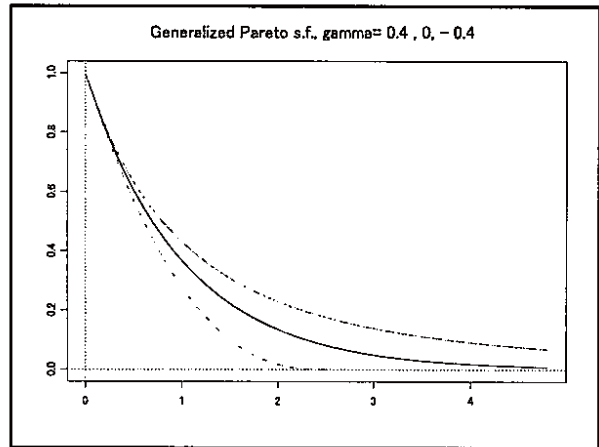
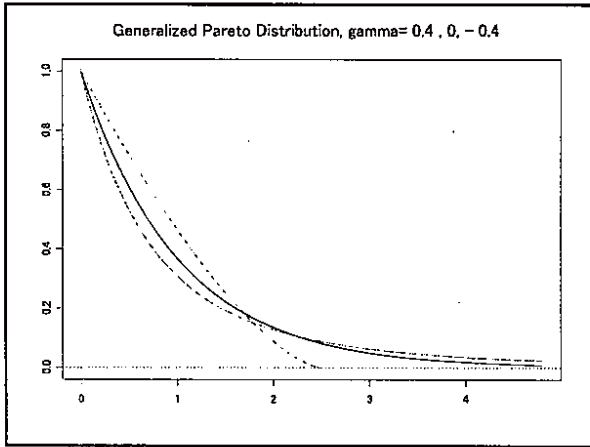
一般極值分布



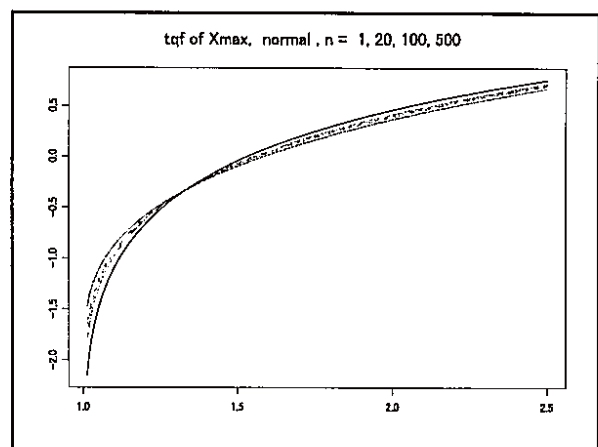
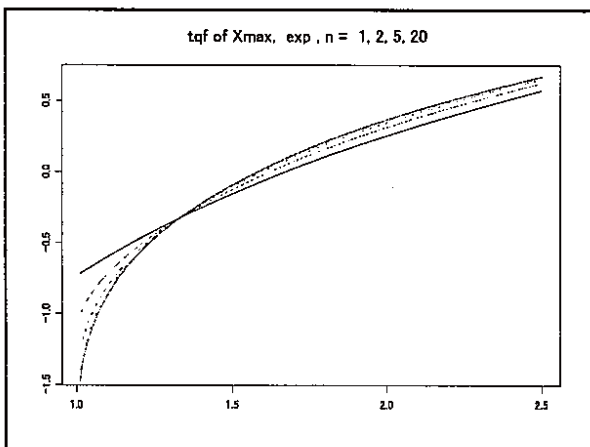
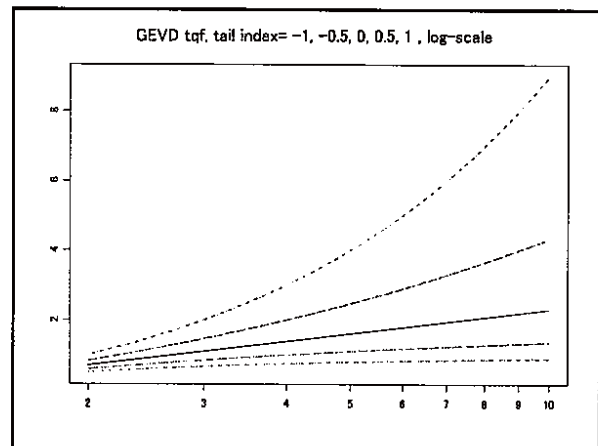
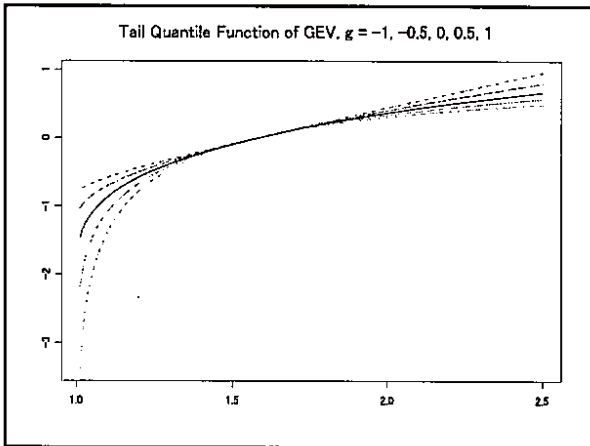
一般極値分布



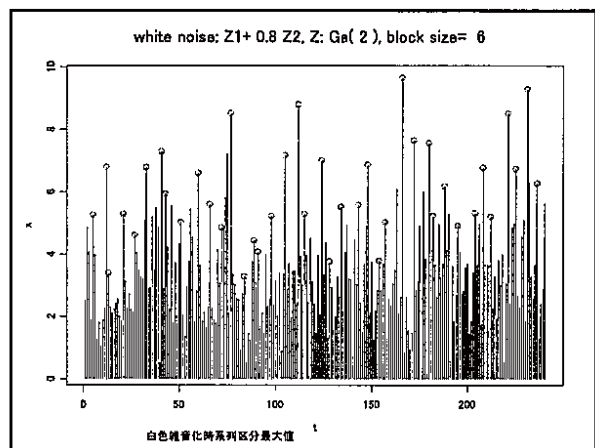
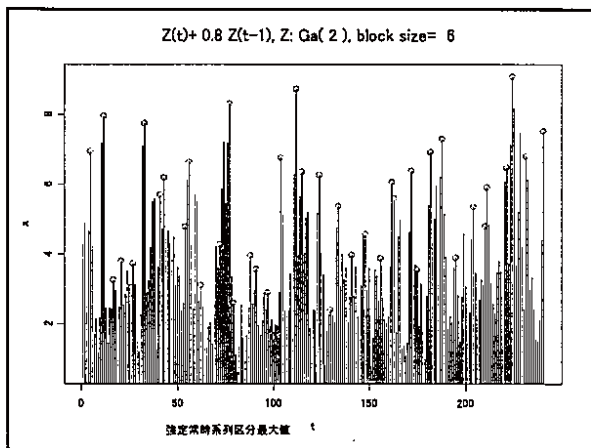
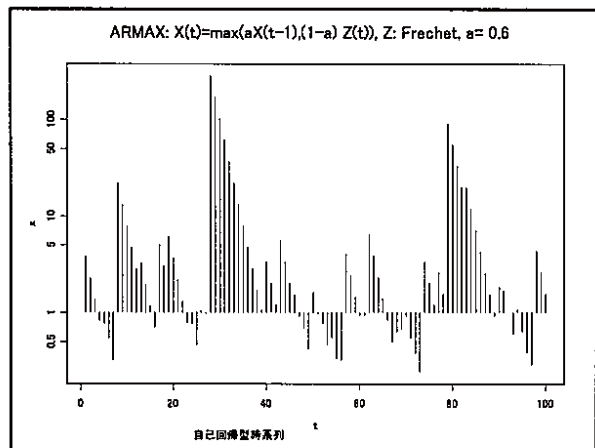
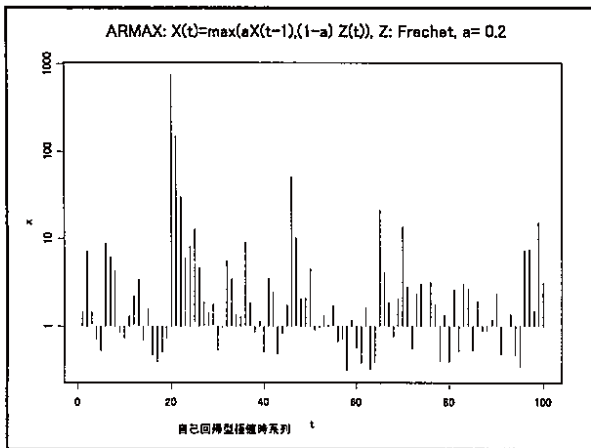
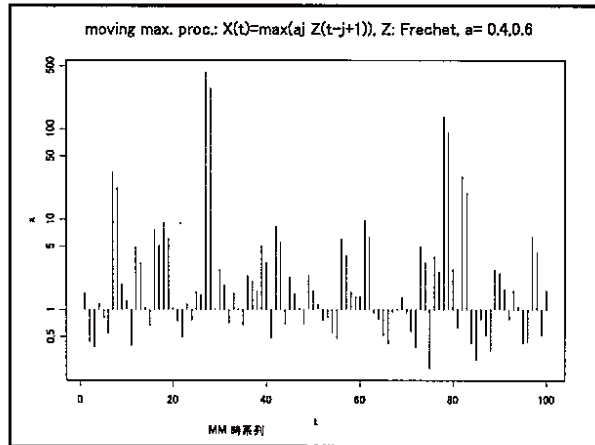
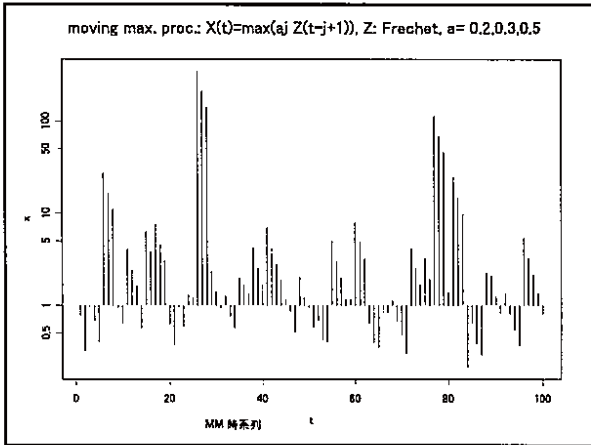
一般 Pareto 分布



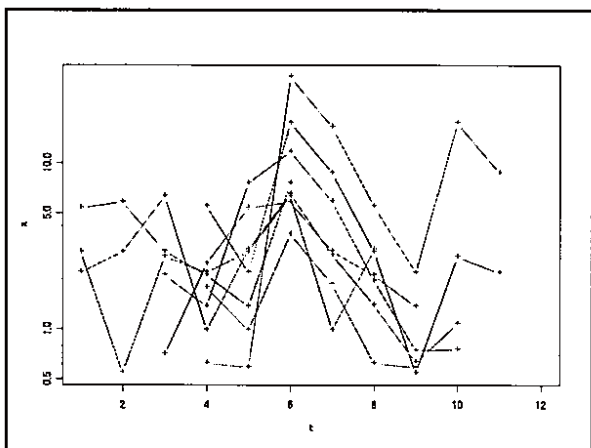
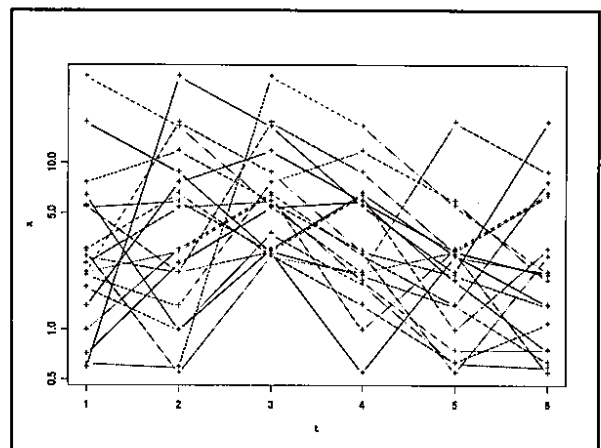
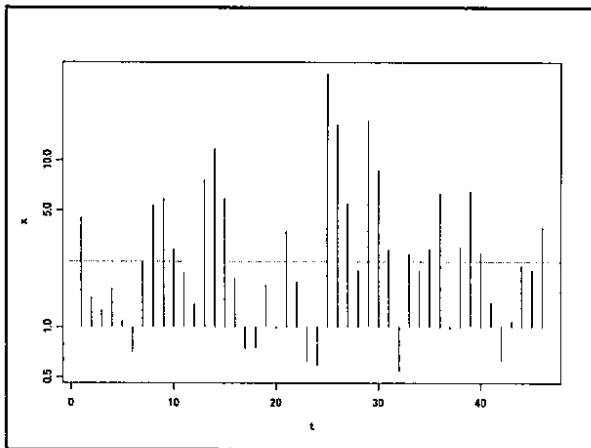
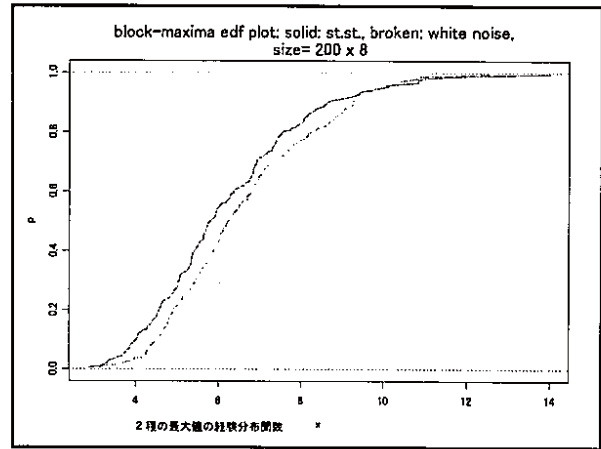
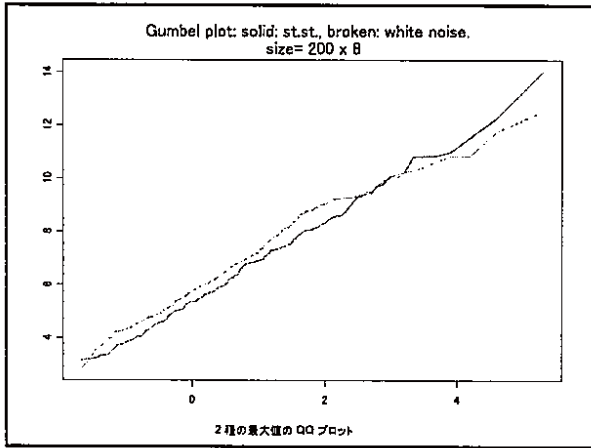
GEVD 裾確率関数, 最大値関数の収束



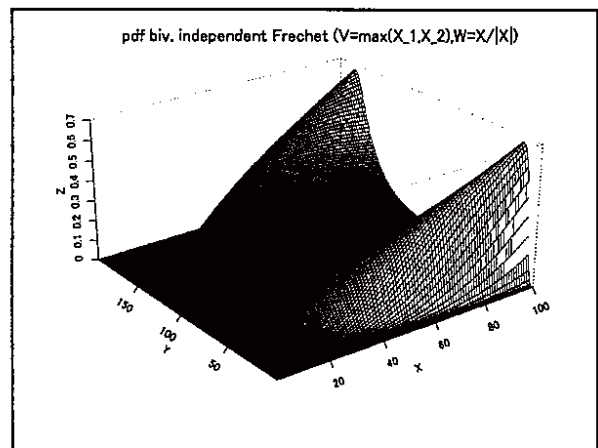
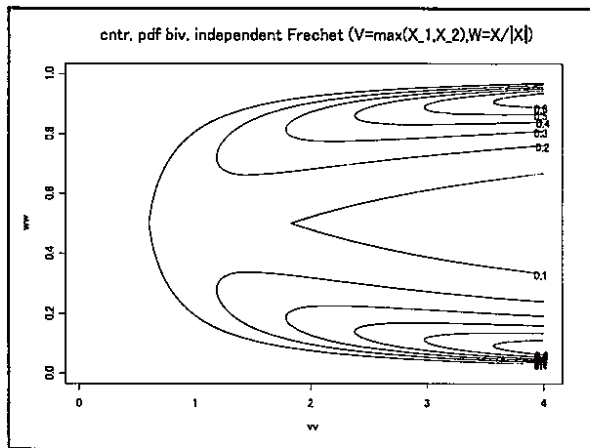
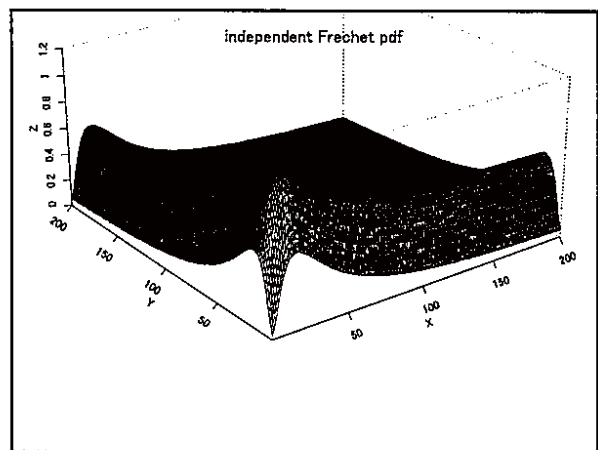
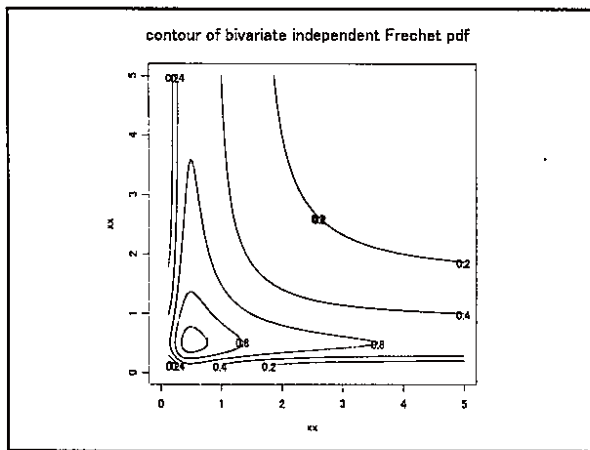
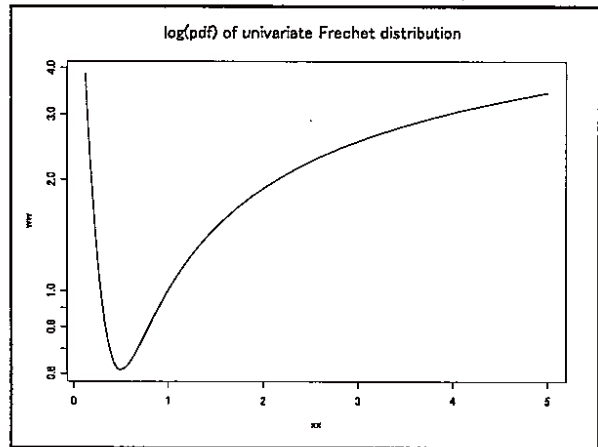
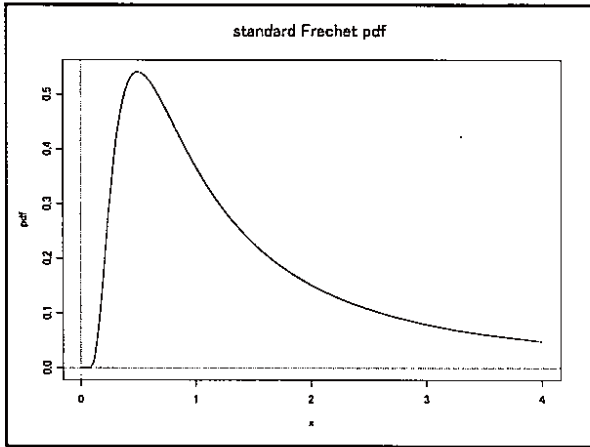
強定常時系列, MM時系列



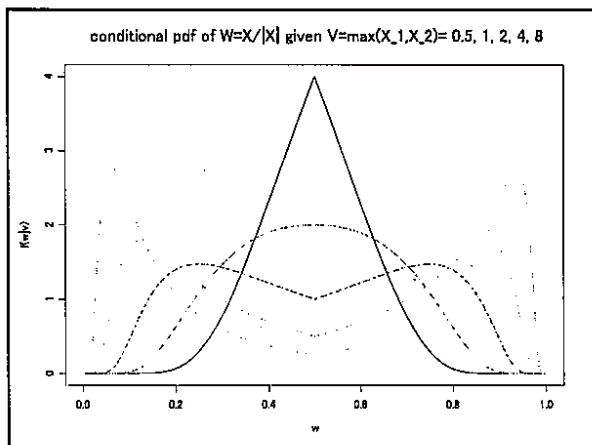
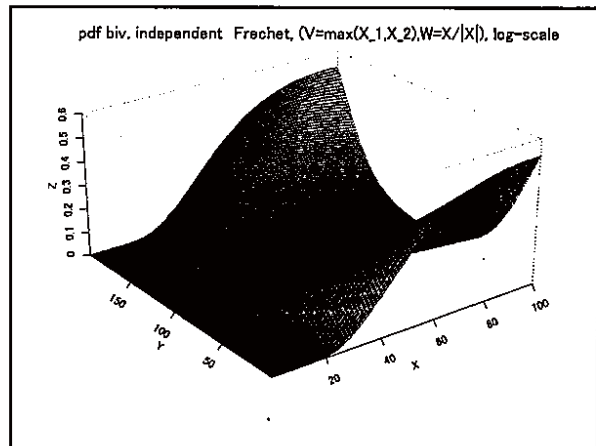
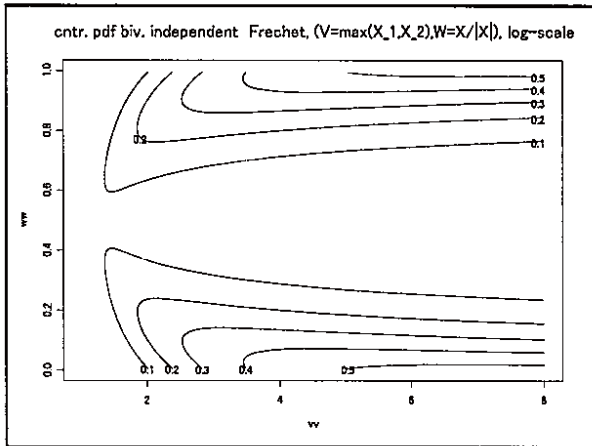
強定常時系列, MM時系列



2 变量独立 Frechet 分布



2 变量独立 Frechet 分布



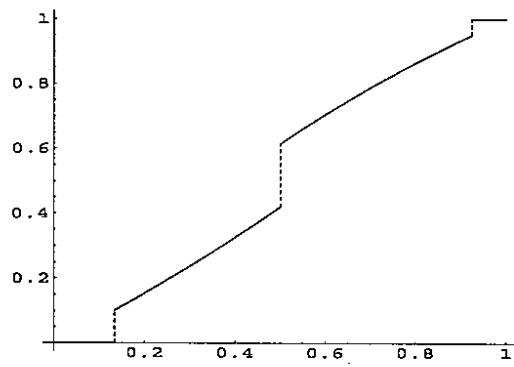


図 1: 超単純 2 変量極値スペクトル分布関数 A

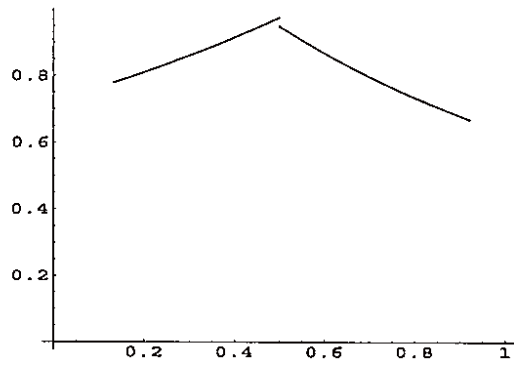


図 2: 超単純 2 変量極値スペクトル密度関数 A

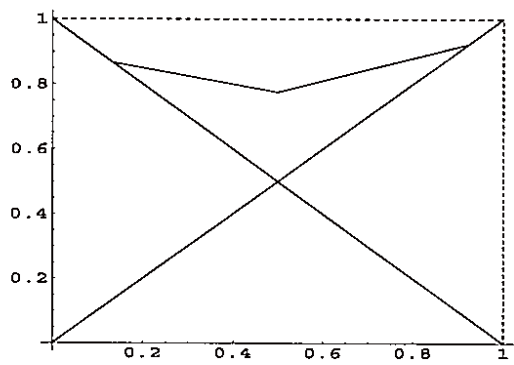


図 3: 超単純 2 変量極値 Pickands 従属関数 A

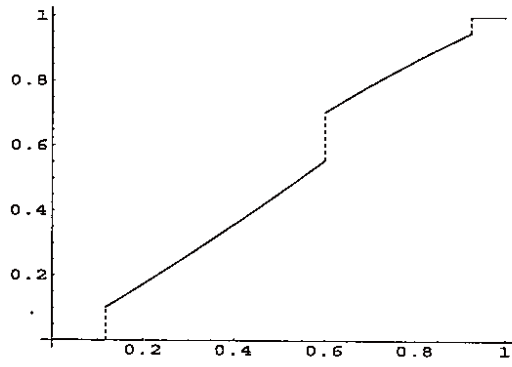


図 4: 超単純 2 変量極値スペクトル分布関数 B

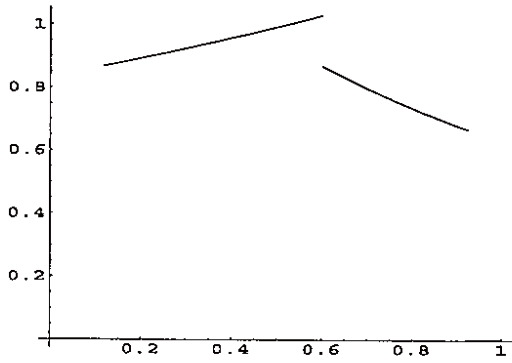


図 5: 超単純 2 変量極値スペクトル密度関数 B

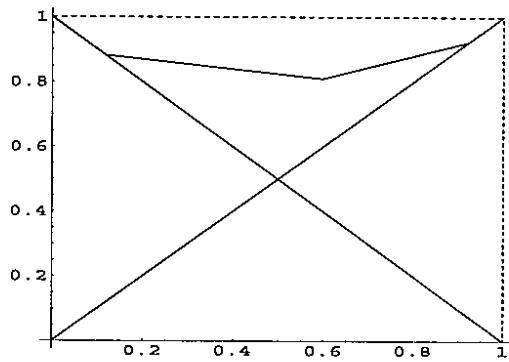
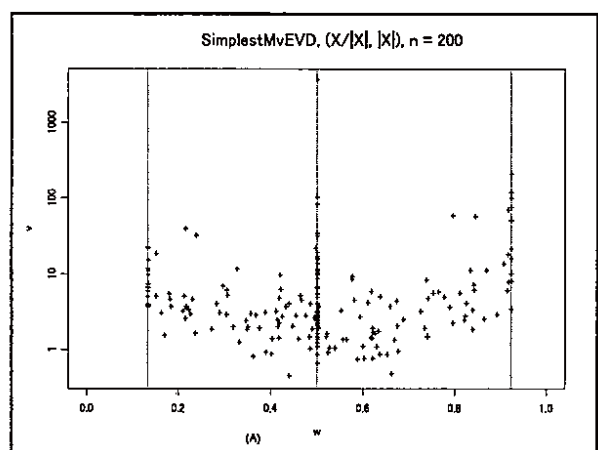
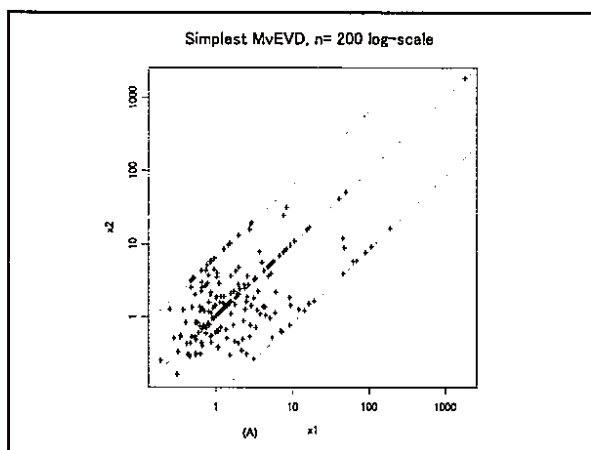
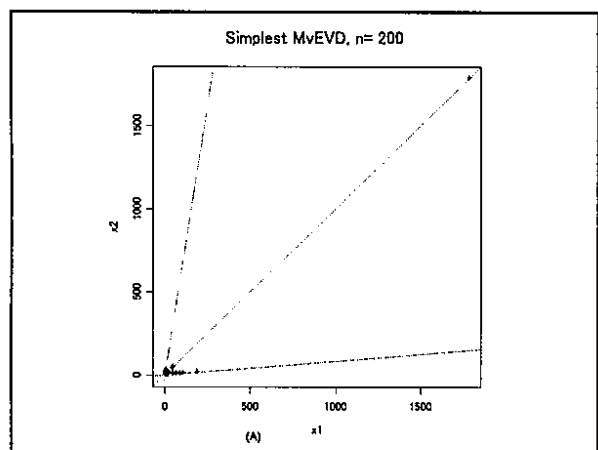
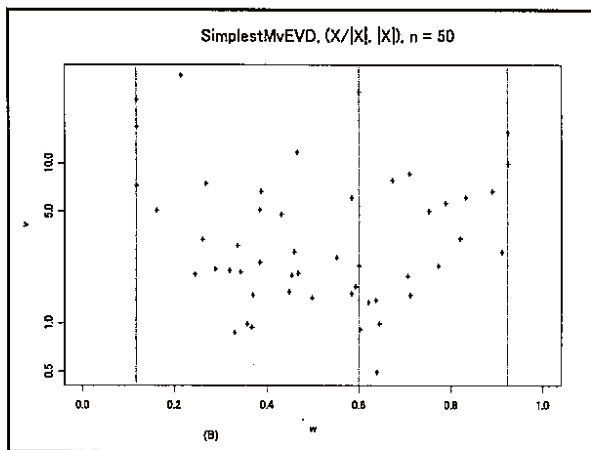
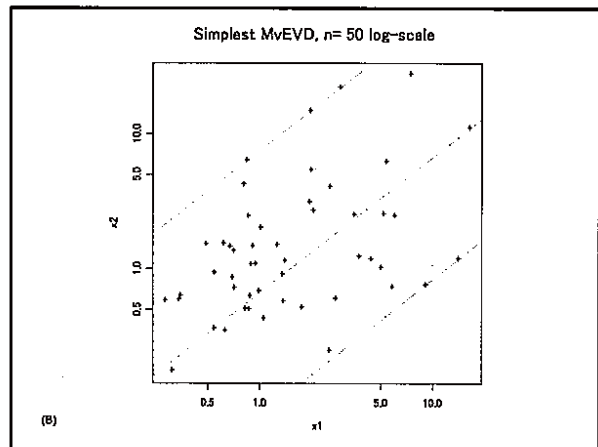
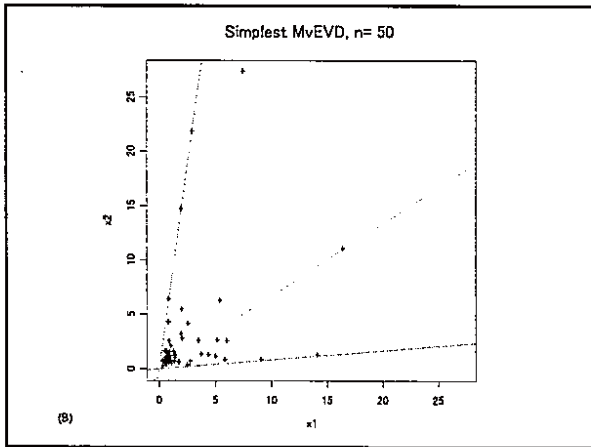
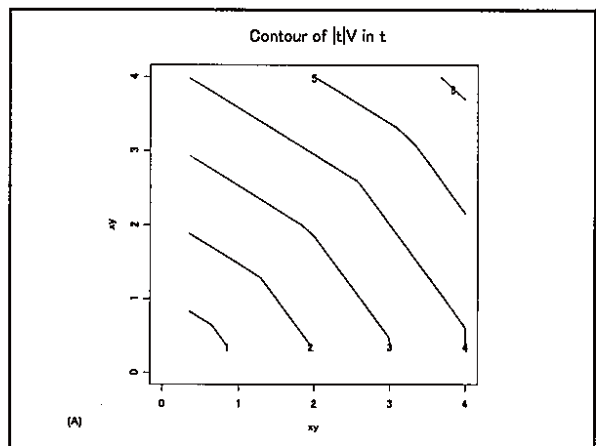
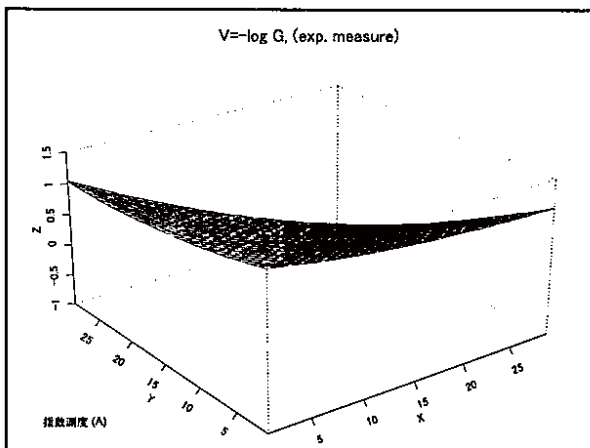
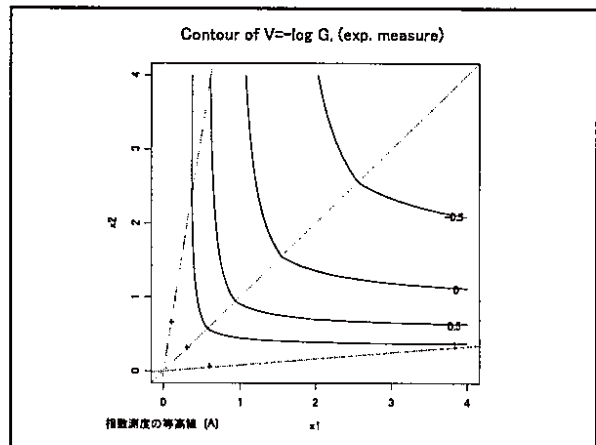
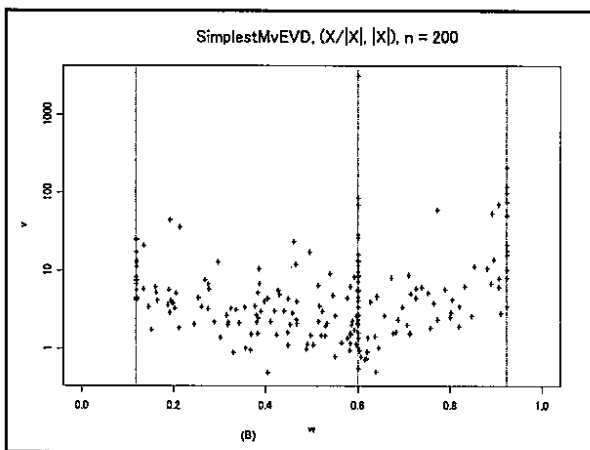
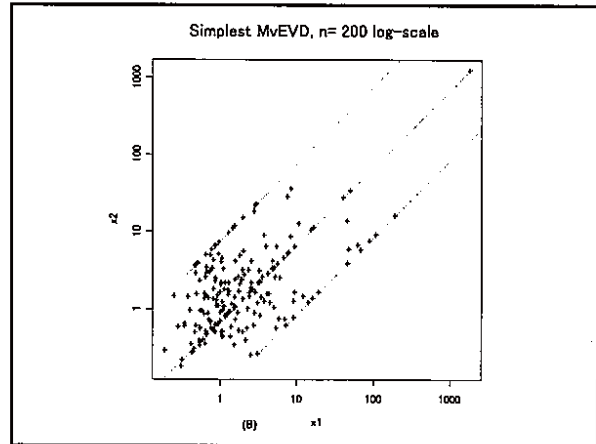
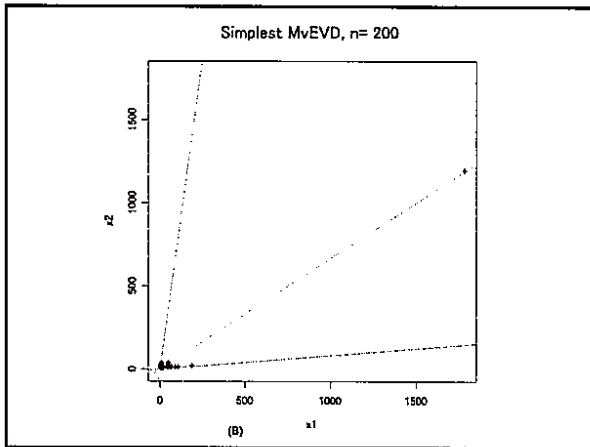


図 6: 超単純 2 変量極値 Pickands 従属関数 B

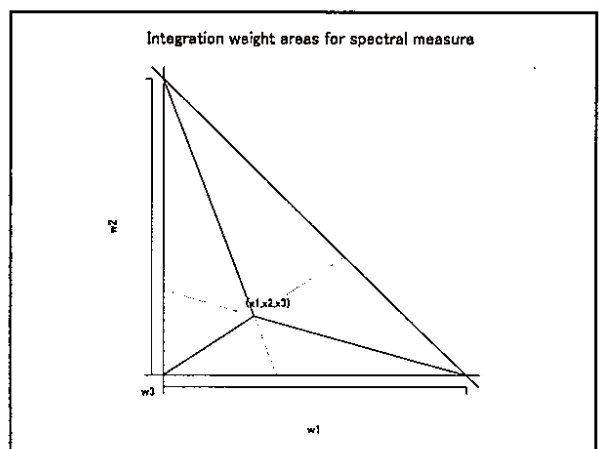
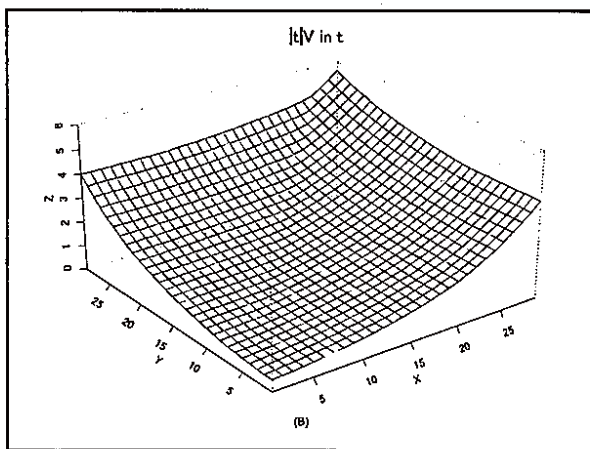
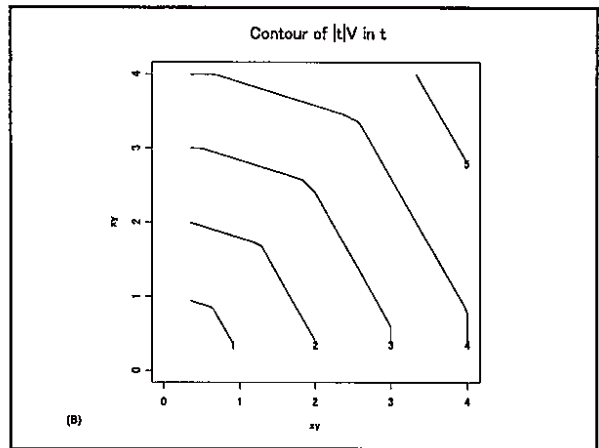
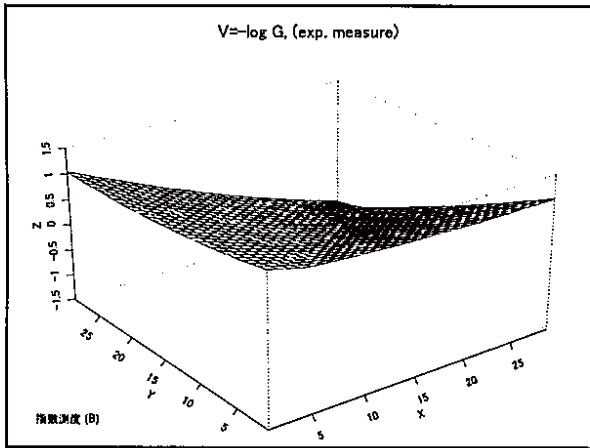
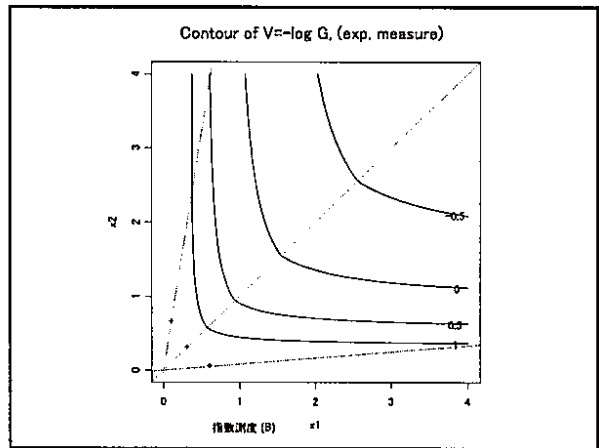
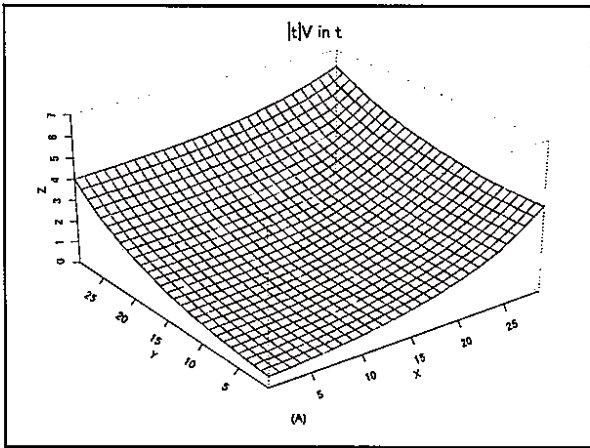
超単純2変量極値分布



超単純2変量極値分布



超単純2変量極値分布



マルチンゲール理論の基礎

西山陽一

(統計数理研究所)

nisiyama@ism.ac.jp

2007年9月

本稿は、2007年5月に東京大学経済学部において2回にわたって行った講義のノートに若干の加筆をしたものである。内容は、測度論的確率論の知識を前提とした上で、条件付き期待値の復習から入り、マルチンゲールの定義、予測可能性、有界変動性、Doob-Meyer 分解定理、二次変分の解説をし、確率積分と伊藤の公式の紹介をしたものである。伊藤の公式の応用として、マルチンゲール中心極限定理のひとつの証明を与え、その応用として拡散過程の母数推定（最尤推定量の漸近正規性）を扱っている。

マルチンゲール理論は本来、通年の講義で扱うべき広く深い理論である。それを2回（計3時間）で解説したわけであるから、証明はほとんど省略せざるをえなかった。同じ理由で、本稿も「レジュメ」と呼ぶべきものとなり、数学的にはまったく不完全なものとなった。しかしながら、マルチンゲール理論を詳細に扱った良書は既にたくさん出版されている。そういった中で、本稿は同理論のエキスのみを抽出したパンフレットのようなものを取って目指した。より深い研究のための道しるべとなれば幸いである。

最後に、同講義は研究会「保険と金融の統計学」の一環として行ったものであり、このような機会を与えてくださった国友直人教授（東京大学）と川崎能典准教授（統計数理研究所）に感謝したい。

1 条件付き期待値

(Ω, \mathcal{F}, P) を確率空間とする。 \mathcal{F} の部分 σ -加法族 \mathcal{G} 、すなわち

$$\{\phi, \Omega\} \subset \mathcal{G} \subset \mathcal{F}$$

が与えられたとする。

定理 1 X が \mathcal{F} -可測な実数値可積分確率変数であるとき、ある \mathcal{G} -可測な実数値可積分確率変数 $E[X|\mathcal{G}]$ であって

$$\int_{\mathcal{G}} E[X|\mathcal{G}] dP = \int_{\mathcal{G}} X dP \quad \forall \mathcal{G} \in \mathcal{G}$$

をみたすものが一意的に存在する。この $E[X|\mathcal{G}]$ のことを \mathcal{G} が与えられたもとでの X の条件付き期待値とよぶ。

特に $\mathcal{G} = \{\phi, \Omega\}$ であるとき $E[X|\mathcal{G}] = E[X]$ である。

特に $\mathcal{G} = \mathcal{F}$ であるとき $E[X|\mathcal{G}] = X$ である。

一般には X をよりシンプルな確率変数に近似したものが $E[X|G]$ であると解釈すればよい。 G が貧弱な σ -加法族であるときには $E[X|G]$ は階段関数に近いようなシンプルな形をしており、 G が豊富な σ -加法族であるときには $E[X|G]$ は情報をたくさんもった複雑な形をしている。

2 確率基、適合過程、マルチンゲール

(Ω, \mathcal{F}, P) を確率空間とする。フィルトレーション $(\mathcal{F}_t)_{t \in [0, \infty)}$ とは \mathcal{F} の部分 σ -加法族が時間パラメータ $t \in [0, \infty)$ によって系列になっているものであって、右連続かつ増加するものであるとする。ただし「右連続」は本稿の範囲内ではあまり気にする必要はない。また「増加」とは

$$\mathcal{F}_s \subset \mathcal{F}_t \text{ for } s \leq t$$

を意味するものとする。このとき、四つ組

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, \infty)}, P)$$

を確率基とよぶ。

実数値確率過程 $(X_t)_{t \in [0, \infty)}$ が $(\mathcal{F}_t)_{t \in [0, \infty)}$ に適合しているとは、各 $t \in [0, \infty)$ に対し X_t が \mathcal{F}_t -可測であることをいう。この性質は、 X_t が時刻 t までの情報のみによって定まる確率変数であることを意味する。

マルチンゲールを定義するためには必ず確率基を用意し、適合過程の概念を思い出さなければならない。

定義 2 右連続かつ左極限をもつ適合過程 $(X_t)_{t \in [0, \infty)}$ に対し

$$(1) \quad X_s = E[X_t | \mathcal{F}_s] \quad \text{a.s. for } s \leq t$$

$$(2) \quad X_s \leq E[X_t | \mathcal{F}_s] \quad \text{a.s. for } s \leq t$$

$$(3) \quad X_s \geq E[X_t | \mathcal{F}_s] \quad \text{a.s. for } s \leq t$$

を考える。(1) のときマルチンゲールといい、(2) のとき劣マルチンゲールといい、(3) のとき優マルチンゲールという。

劣マルチンゲールと優マルチンゲールは、なんとなく用語が逆のような気がするかもしれないが、なぜかこうよばれている。

面白い例は後に述べることにして、とりあえず簡単な例を挙げる。 $t \mapsto M_t$ がマルチンゲールであるとき、

$$X_t = t + M_t \quad \text{は劣マルチンゲールであり、}$$

$$X_t = -t + M_t \quad \text{は優マルチンゲールであるが、}$$

$$X_t = \sin t + M_t \quad \text{はいずれでもない。}$$

しかし、最後のものも含めて、いま挙げた例は全て半マルチンゲールである。半マルチンゲールの正確な定義は後に述べるが、先走って簡単に言うと

$$X_t = \text{有界変動適合過程} + \text{マルチンゲール}$$

の形に表されるもののことである。

では最も重要な例を2つ挙げる。一つ目はウィナー過程である。

定義 3 $t \mapsto W_t$ が以下の2条件を満たすとき、標準ウィナー過程であるという：

- (1) 確率1で、 $W_0 = 0$ および $t \mapsto W_t$ が連続である。
- (2) 任意の $0 = t_0 < t_1 < \dots < t_n$ (n は任意の自然数) に対して $\{W_{t_i} - W_{t_{i-1}}\}_{i=1}^n$ は互いに独立であり、それぞれ $N(0, t_i - t_{i-1})$ に従う。

命題 4 $t \mapsto W_t$ が標準ウィナー過程であるとき、

$$X_t = W_t$$

はマルチンゲールであり、

$$X_t = W_t^2 - t$$

もマルチンゲールである。

注意 この命題を示すだけなら、確率解析の難しい道具を使う必要は全くなく、いまの時点で証明することもできる。しかしながら、本稿の目標はマルチンゲール理論の見通しの良い統一的な扱いを紹介することである。よって、上記の命題は、もっと話を一般化して覚えやすい公式の形にして述べる。

二つ目の例はポアソン過程である。

定義 5 $t \mapsto N_t$ が以下の2条件を満たすとき、強度 λ のポアソン過程であるという：

- (1) 任意の $0 = t_0 < t_1 < \dots < t_n$ (n は任意の自然数) に対して $\{N_{t_i} - N_{t_{i-1}}\}_{i=1}^n$ は互いに独立である。
- (2) $s < t$ に対して $N_t - N_s$ はパラメータ $\lambda(t - s)$ のポアソン分布に従う。

命題 6 $t \mapsto N_t$ が強度 λ のポアソン過程であるとき、

$$X_t = N_t - \lambda t$$

はマルチンゲールであり、

$$X_t = (N_t - \lambda t)^2 - \lambda t$$

もマルチンゲールである。

注意 やはり、この命題の覚え方は後で述べる。

3 局所●●って何？

マルチンゲール理論を勉強されたことのある方なら、「停止時刻」に基づいた「局所●●」の概念に出会って戸惑われたことがあるかもしれない。なぜこのような概念が必要なのかを説明しよう。とりあえずその概念の定義を述べる。

定義 7 停止時刻とは、写像 $T: \Omega \rightarrow [0, \infty]$ であって $\{T \leq t\} \in \mathcal{F}_t$ が任意の $t \geq 0$ に対して成り立つようなもののことである。

定義 8 \mathcal{C} が確率過程の族であるとき、局所化された族を次のように定義し \mathcal{C}_{loc} と表す：
 $X \in \mathcal{C}_{loc}$ であるのは、停止時刻の増大列 (T_n) であって確率 1 で $T_n \rightarrow \infty$ ($n \rightarrow \infty$) が成り立ちかつ $t \mapsto X_{t \wedge T_n}$ が \mathcal{C} に属するものが存在するときという。

例で説明しよう。適合増加過程 $t \mapsto A_t$ が可積分であるとは $EA_\infty < \infty$ が成り立つときにいう。この定義によると、ポアソン過程は可積分ではない： $EN_\infty = \infty$ 。そこで局所可積分という概念を導入したい。ひとつの自然な定義は

$$EN_t < \infty \quad \forall t > 0$$

であると思われるかもしれない。確かに、ポアソン過程の場合 $EN_t = \lambda t < \infty$ である。しかし、強度 λ が定数でなく確率過程 $\lambda_t(\omega)$ であるような場合（つまり、一般の計数過程である場合）には $EN_t = E \int_0^t \lambda_s ds$ が有限であるかどうかは明らかではない。ところが「局所●●」の定義を上述のように与えておけば、停止時刻 T_n を

$$T_n = \inf\{t \geq 0 : N_t \geq n\}$$

にとっておくだけで容易に局所可積分であることがチェックできる。実際、

$$E(N_{\infty \wedge T_n}) = E(N_{T_n}) \leq n < \infty$$

つまり局所●●とは、「各 t に対し X_t が●●を満たす」という概念を緩めて、「停止時刻の増大列 T_n に対し X_{T_n} が●●を満たす」という概念にまで一般化させたものである。

4 $\varphi(M)$ は劣マルチンゲール

次の命題は後に重要になる。特に $\varphi(x) = x^2$ の形で繰り返し用いることになる。

命題 9 M はマルチンゲールであるとし、 $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ は下に凸であるとする。可積分性の条件のもとで、 $\varphi(M)$ は劣マルチンゲールになる。

証明：条件付き期待値に対するイェンセンの不等式より、全ての $s \leq t$ に対し

$$\varphi(E[M_t | \mathcal{F}_s]) \leq E[\varphi(M_t) | \mathcal{F}_s]$$

が成り立つ。 M がマルチンゲールであることから左辺の $\varphi(\cdot)$ の中身は確率 1 で M_s に等しい。すなわち、

$$\varphi(M_s) \leq E[\varphi(M_t) | \mathcal{F}_s]$$

であり、 $\varphi(M)$ は劣マルチンゲールである。□

5 予測可能性と有界変動性

マルチンゲール理論を展開するにあたって最も基本的な概念は「予測可能 (predictable) 性」と「有界変動性」である。これなくしては、Doob-Meyer 分解定理も語れない。

5.1 予測可能性

定義 10 予測可能 σ -加法族とは、 $\Omega \times [0, \infty)$ 上の σ -加法族 \mathcal{P} であって、全ての左連続適合過程によって生成されるもののことである。

丁寧に言えば次のようになる。まず左連続適合過程 $t \mapsto X_t$ をひとつとってきて、これを $\Omega \times [0, \infty)$ 上で定義された実数値関数 $(\omega, t) \mapsto X_t(\omega)$ と見なす。これに対し、 $\Omega \times [0, \infty)$ の部分集合族 \mathcal{A}_X を

$$\mathcal{A}_X = (\{(\omega, t) : X_t(\omega) \in B\} : B \in \mathbf{B}(\mathbb{R}))$$

によって定義する。この操作を全ての左連続適合過程 X について行い、それを含む最小の σ -加法族を \mathcal{P} とおく：

$$\mathcal{P} = \sigma(\mathcal{A}_X : X \text{ は左連続適合過程})$$

お分かりのように、予測可能 σ -加法族 \mathcal{P} の定義の中には適合性の概念を要請している。つまり、フィルトレーション $(\mathcal{F}_t)_{t \in [0, \infty)}$ が与えられることによってはじめて定まるものである。

定義 11 確率過程 $t \mapsto X_t$ が予測可能であるとは、それを $\Omega \times [0, \infty)$ 上で定義された実数値関数と見たときに \mathcal{P} -可測であるときにいう。

実用上は次の2つの事実を覚えておけば十分である：

- (1) 左連続適合過程は予測可能である。よって連続適合過程はもちろん予測可能である。
- (2) ランダムでない (ω に依存しない) 確率過程は予測可能である。

5.2 有界変動性

有界変動性の概念は、ランダム性とは関係なく、単なる実数値関数に対して定義されるものである。

定義 12 関数 $F : [0, T] \rightarrow \mathbb{R}$ が有界変動であるとは、それが2つの単調増加関数 F_+, F_- の差として表されるときにいう： $F(t) = F_+(t) - F_-(t)$

ここで有界変動関数に関するルベーグ・スティルチェス積分 $\int_0^T \varphi(t) dF(t)$ の定義を思い出しておこう。

有界変動関数 F が与えられたとき、まずそれを上述のように $F = F_+ - F_-$ と分解する。そして F_+ に対して $\int_0^T \varphi(t) dF_+(t)$ を次のように定義する：分点 $0 = t_0 < t_1 < \dots < t_n = T$

をとり、 $\xi_i \in (t_{i-1}, t_i]$ をとり、

$$S_n = \sum_{i=1}^n \varphi(\xi_i)[F_+(t_i) - F_+(t_{i-1})]$$

とおく。そして分点を細かくとって $n \rightarrow \infty$ としたときの極限をとる：

$$\int_0^T \varphi(t) dF_+(t) := \lim_{n \rightarrow \infty} S_n$$

同様にして $\int_0^T \varphi(t) dF_-(t)$ を定義し、最終的に

$$\int_0^T \varphi(t) dF(t) := \int_0^T \varphi(t) dF_+(t) - \int_0^T \varphi(t) dF_-(t)$$

と定義する。厳密には、極限が分点の取り方に依らないことを示さなければならない。

さて、ランダムな世界に話を戻そう。確率過程 $t \mapsto X_t$ が有界変動であるとは、全ての $\omega \in \Omega$ に対して $t \mapsto X_t(\omega)$ が有界変動関数であるときにいう。この場合、さらにもう一つ適合過程 $t \mapsto H_t$ が与えられたとき、確率積分 $H \bullet X_t$ を次のように定義する：

$$H \bullet X_t(\omega) := \int_0^t H_s(\omega) dX_s(\omega) \quad \forall \omega \in \Omega$$

注意：連続関数は有界変動であるとは限らない。実際、ウィナー過程 $t \mapsto W_t(\omega)$ は確率 1 で有界変動ではない。よって

$$H \bullet W_t(\omega) = \int_0^t H_s(\omega) dW_s(\omega)$$

というような安直な定義はできない。だから後に述べるような確率積分を定義する必要があるのである。

なお、予測可能性と有界変動性の重要性を先走って説明しておく。後に述べる Doob-Meyer 分解定理の系として「予測可能かつ有界変動なマルチンゲールは 0 に限られる」ということが出てくる。これ自体、美しい結果である。すなわち、予測可能であるが有界変動でないマルチンゲールの例はいくらでも存在する。例えばウィナー過程がそうである。一方、予測可能ではないが有界変動であるマルチンゲールの例も存在する。例えば計数過程の残差マルチンゲールがそうである。ところが、予測可能性と有界変動性の両方を併せもつマルチンゲールは 0 という自明な例しか存在しないわけである。

6 さまざまなマルチンゲールとクラス D

抽象的な説明になるが、マルチンゲールのさまざまなバリエーションと「クラス D」について結果のみまとめておく。クラス D は次の節で Doob-Meyer 分解定理を述べるために必要な概念である。

マルチンゲール X が一様可積分であるとは、

$$\lim_{K \rightarrow \infty} \sup_{t \in [0, \infty)} E|X_t| 1_{\{|X_t| > K\}} = 0$$

であるときにいい、その全体を \mathcal{M} と書く。 \mathcal{M}_{loc} の元のことを局所マルチンゲールという。

マルチンゲール X が二乗可積分であるとは、

$$\sup_{t \in [0, \infty)} E|X_t|^2 < \infty$$

であるときにいい、その全体を \mathcal{H}^2 と書く。 \mathcal{H}_{loc}^2 の元のことを局所二乗可積分マルチンゲールという。

確率過程 X がクラス D に属するとは

$$\lim_{K \rightarrow \infty} \sup_T E|X_T| 1_{\{|X_T| > K\}} = 0$$

が成り立つときにいう。ただし \sup_T は全ての有限停止時刻 T でとる。

これらの間には次のような関係がある。

定理 13 (a) マルチンゲールは局所マルチンゲールである。つまり $\mathcal{M} \subset \{\text{マルチンゲール}\} \subset \mathcal{M}_{loc}$

(b) 一様可積分マルチンゲールはクラス D に属する。つまり $\mathcal{M} \subset \text{クラス D}$

(c) 局所マルチンゲールが一様可積分であるための必要十分条件はそれがクラス D に属することである。つまり $\mathcal{M} = \mathcal{M}_{loc} \cap \text{クラス D}$

7 Doob の不等式と Doob-Meyer 分解定理

我々の最終目標は伊藤の公式であるが、そのためには二次変分を定義する必要がある。二次変分の一意的存在を示すためには Doob-Meyer 分解定理を用いる。Doob-Meyer 分解定理の仮定をチェックするために、Doob の不等式が有用となる。

7.1 Doob の不等式

定理 14 (Doob の不等式) X が二乗可積分マルチンゲールであるとき

$$E \left(\sup_{t \in [0, \infty)} X_t^2 \right) \leq 4 \sup_{t \in [0, \infty)} E(X_t^2) = 4E(X_\infty^2)$$

Doob の不等式で著しいのは、左辺において $\sup_{t \in [0, \infty)}$ が積分の中に入っている点である。その量が右辺のように $\sup_{t \in [0, \infty)}$ が積分の外にあるという扱いやすいもので押さえられているのが便利なのである。

7.2 Doob-Meyer 分解定理

まず単調増加過程はもちろん有界変動であることを思い出しておこう。

定理 15 (Doob-Meyer 分解定理) クラス \mathcal{D} に属する劣マルチンゲール X に対し、単調増加可積分予測可能過程 A であって $A_0 = 0$ かつ $X - A$ が一様可積分マルチンゲールとなるようなものが一意的に存在する。

ラフに言えば、 X がよい可積分性をもつ劣マルチンゲールであるとき、

$$X = A + M$$

と分解できる。ただし A は単調増加 (よって有界変動) かつ予測可能で、 M はよい可積分性をもつマルチンゲール。もっとラフに言えば

$$\text{劣マルチンゲール} = \text{単調増加トレンド} + \text{マルチンゲール}$$

ということである。

なお、トレンド + マルチンゲールへの分解自体は一意的ではない。Doob-Meyer 分解定理のポイントは、トレンドの部分が予測可能かつ有界変動であると仮定した瞬間、分解が一意的になる、という点である。そのことは、次の系をみるとはっきりする。

系 16 局所マルチンゲールであって予測可能かつ有界変動であるものは、0 に限られる。

この結果については、5 節の末尾で述べた説明を思い出して欲しい。

例 17 $t \mapsto W_t$ が標準ウィナー過程であるとき $t \mapsto W_t^2 - t$ はマルチンゲールとなると先に予告した。イェンセンの不等式から $t \mapsto W_t^2$ は劣マルチンゲールである。よって Doob-Meyer 分解 $W^2 = A + M$ が可能である訳であるが、その $t \mapsto A_t$ がこの場合ランダムでない関数 $A_t = t$ となるのである。

さて、次の系は単調増加過程に対する予測可能カンペンセイターの存在を保証するものである。

系 18 X は単調増加局所可積分過程であるとする。このとき、予測可能単調増加局所可積分過程 X^p であって $X - X^p$ が局所マルチンゲールとなるようなものが一意的に存在する。

この X^p のことを X の予測可能カンペンセイターとよぶ。

例 19 $t \mapsto N_t$ は整数値単調増加局所可積分過程であって $N_0 = 0$ かつ $N_t - N_{t-} \leq 1$ であるものであるとする。このとき、予測可能単調増加局所可積分過程 $t \mapsto A_t$ であって $N - A$ が局所マルチンゲールとなるようなものが一意的に存在する。このような N を計数過程とよび、それに対応する A を予測可能カンペンセイターとよぶ。この A がルベグ測度に関して絶対連続であるとき、すなわち

$$A_t(\omega) = \int_0^t \lambda_s(\omega) ds$$

と書けるとき、 $t \mapsto \lambda_t$ のことを強度とよぶ。強度は増加のトレンドの微係数である。

8 二次変分

いままで Doob の不等式と Doob-Meyer 分解定理を見てきたが、それらを組み合わせることによって、局所二乗可積分マルチンゲールの二次変分の一意的な存在が証明できる。

定理 20 M が局所二乗可積分マルチンゲールであるとき、単調増加（よって有界変動）予測可能過程 $t \mapsto \langle M, M \rangle_t$ であって $M^2 - \langle M, M \rangle$ が局所マルチンゲールとなるものが一意的に存在する。

証明の概略：まずイェンセンの不等式から $t \mapsto M_t^2$ が劣マルチンゲールであることを示し（命題 9）、また Doob の不等式からそれがクラス D に属することを示す。そうすると Doob-Meyer 分解定理が適用できて $\langle M, M \rangle$ の一意的な存在がいえる。□

定理 21 M, N が局所二乗可積分マルチンゲールであるとき、有界変動予測可能過程 $t \mapsto \langle M, N \rangle_t$ であって $MN - \langle M, N \rangle$ が局所マルチンゲールとなるものが一意的に存在する。

証明の概略：

$$\langle M, N \rangle = \frac{1}{4}(\langle M + N, M + N \rangle - \langle M - N, M - N \rangle)$$

とおけばよい。右辺は前述の定理によって定義できている。□

二次変分は次のように解釈するのがよい。

マルチンゲール M が与えられたとき、 $\langle M \rangle = \langle M, M \rangle$ は条件付き分散であって、予測可能かつ有界変動であるものである。

マルチンゲール M, N が与えられたとき、 $\langle M, N \rangle$ は条件付き共分散であって、予測可能かつ有界変動であるものである。

最初の方で挙げた命題 4, 6 は、次のように覚えておくのがよい。

例 22 (a) W が標準ウィナー過程であるとき

$$\langle W, W \rangle_t = t$$

(b1) N が強度 λ のポアソン過程であるとき、 $M_t = N_t - \lambda t$ は局所二乗可積分マルチンゲールであって

$$\langle M, M \rangle_t = \lambda t$$

(b2) より一般に、 N が強度 $t \mapsto \lambda_t(\omega)$ の計数過程であるとき、 $M_t = N_t - \int_0^t \lambda_s ds$ は局所二乗可積分マルチンゲールであって

$$\langle M, M \rangle_t(\omega) = \int_0^t \lambda_s(\omega) ds$$

9 確率積分

いよいよ確率積分の説明に入る。使いこなせるようになるのが本稿の主目標であるから、(1) まず「初等確率過程」に対する確率積分の定義を行い、(2) 次に一般の予測可能過程に対する確率積分の性質を取って先にリストアップし、(3) 最後に一般の予測可能過程に対する確率積分の構成についての少し詳しい説明をする。

9.1 初等確率過程に対する確率積分の定義

まず

$$H_t(\omega) = \sum_{i=1}^k Y_{i-1}(\omega) 1_{(t_{i-1}, t_i]}(t)$$

の形をした確率過程を考える。ただし $0 = t_0 < t_1 < \dots < t_k$ は $[0, \infty)$ の分点で、各 Y_i は \mathcal{F}_{t_i} 可測で有界なものである。このような確率過程を初等確率過程とよび、その全体を \mathcal{L}_0 と表すことにする。初等確率過程は予測可能過程の特別な場合である。

二乗可積分マルチンゲール M が与えられたときに、各 $t \in [0, \infty)$ に対し $t_l \leq t < t_{l+1}$ となる l を選び、

$$I(H)_t(\omega) = \sum_{i=1}^l Y_{i-1}(\omega)(M_{t_i} - M_{t_{i-1}}) + Y_l(\omega)(M_t - M_{t_l})$$

と定義する。これに対し次が成り立つ。

補題 23 $I(H)$ は二乗可積分マルチンゲールであり、

$$\langle I(H), I(H) \rangle_t = \int_0^t H_s^2 d\langle M, M \rangle_s$$

この事実により、 $I(H)$ のことを H の M に関する確率積分とよび、

$$H \bullet M_t = \int_0^t H_s dM_s := I(H)_t(\omega)$$

と書く。

確率積分は H が予測可能過程で M が局所マルチンゲールである状況まで一般化できる。

9.2 確率積分の性質

以下 H, K は適当な可積分性をもつ予測可能過程で、 M, N は二乗可積分マルチンゲールであるとする。このとき、次の諸性質が成り立つ。

定理 24 (1) $t \mapsto H \bullet M_t$ は二乗可積分マルチンゲール

(2) $(aH + bK) \bullet M = a(H \bullet M) + b(K \bullet M)$

(3) M が有界変動ならば $H \bullet M$ も有界変動

(4) $\Delta(H \bullet M) = H \Delta M$

(5) $K \bullet (H \bullet M) = (KH) \bullet M$

(6) $\langle H \bullet M, K \bullet N \rangle = (HK) \bullet \langle M, N \rangle$

最後の公式 (6) は特に重要であるからもうひとつの記法でも書いておこう：

$$\left\langle \int_0^\cdot H_s dM_s, \int_0^\cdot K_s dN_s \right\rangle_t = \int_0^t H_s K_s d\langle M, N \rangle_s$$

ここは急所であるから、もう一度繰り返しておこう。 X_t が二乗可積分マルチンゲールであるとき、その二次変分とは $X_t^2 - \langle X \rangle_t$ がマルチンゲールとなるような予測可能単調増加可積分過程 $t \mapsto \langle X \rangle_t$ のことであった。その計算例を2つ示そう。

例 25 (ウィナー過程) W を標準ウィナー過程とし、確率積分

$$X_t^{(i)} = \int_0^t H_s^{(i)} dW_s, \quad i = 1, 2$$

を考える。 $M = W$ とみなす。例 22 (a) より

$$\langle M \rangle_t = \int_0^t ds \quad \text{よって} \quad d\langle M \rangle_s = ds$$

であるから、 $X^{(1)}$ と $X^{(2)}$ の二次変分は

$$\begin{aligned} \langle X^{(1)}, X^{(2)} \rangle_t &= \int_0^t H_s^{(1)} H_s^{(2)} d\langle M \rangle_s \\ &= \int_0^t H_s^{(1)} H_s^{(2)} ds \end{aligned}$$

例 26 (計数過程) 確率積分

$$X_t^{(i)} = \int_0^t H_s^{(i)} (dN_s - \lambda_s ds), \quad i = 1, 2$$

を考える。 $M_t = N_t - \int_0^t \lambda_s ds$ とみなす。例 22 (b2) より

$$\langle M \rangle_t = \int_0^t \lambda_s ds \quad \text{よって} \quad d\langle M \rangle_s = \lambda_s ds$$

であるから、 $X^{(1)}$ と $X^{(2)}$ の二次変分は

$$\begin{aligned} \langle X^{(1)}, X^{(2)} \rangle_t &= \int_0^t H_s^{(1)} H_s^{(2)} d\langle M \rangle_s \\ &= \int_0^t H_s^{(1)} H_s^{(2)} \lambda_s ds \end{aligned}$$

9.3 確率積分の構成 (少し詳しく)

以下、簡単のため $(M_t)_{t \in [0, \infty)}$ ではなく $(M_t)_{t \in [0, T]}$ を考える。 $\langle M, M \rangle$ のことを簡単に $\langle M \rangle$ と書く。3つの補題を用意する。

補題 27 \mathcal{H}^2 は $\|M\|_T = \sqrt{EM_T^2}$ をノルムとするヒルベルト空間である。

次に

$$\mathcal{L}_2(\langle M \rangle) = \left\{ H : H \text{ は予測可能で } E \int_0^T H_s^2 d\langle M \rangle_s < \infty \right\}$$

とおく。

補題 28 $\mathcal{L}_2(\langle M \rangle)$ は

$$\|H\|_T = \sqrt{E \int_0^T H_s^2 d\langle M \rangle_s}$$

をノルムとするヒルベルト空間である。

補題 29 \mathcal{L}_0 は $\mathcal{L}_2(\langle M \rangle)$ 内で稠密である。

以上を使って $H \in \mathcal{L}_0$ のみならず一般の $H \in \mathcal{L}_2(\langle M \rangle)$ に対して確率積分 $H \bullet M_t = \int_0^t H_s dM_s$ を次の手順で定義する。

(Step 1) まず補題 29 (つまり \mathcal{L}_0 の $\mathcal{L}_2(\langle M \rangle)$ における稠密性) より $H^{(n)} \in \mathcal{L}_0$ で $\|H^{(n)} - H\|_T \rightarrow 0$ となる列 $\{H^{(n)}\}$ が取れる。すなわち、

$$\lim_{n \rightarrow \infty} E \int_0^T |H_s^{(n)} - H_s|^2 d\langle M \rangle_s = 0$$

となるものが取れる。そこで

$$X_t^{(n)} := \int_0^t H_s^{(n)} dM_s$$

とおく。ただし右辺の確率積分は 9.1 節で定義した通りである。

(Step 2) 補題 23 (つまり初等確率過程に対する確率積分の二次変分の公式) を使って、 $n, m \rightarrow \infty$ のとき

$$E|X_T^{(n)} - X_T^{(m)}|^2 = E \int_0^T |H_s^{(n)} - H_s^{(m)}|^2 d\langle M \rangle_s \rightarrow 0$$

である。すなわち $\|X^{(n)} - X^{(m)}\|_T \rightarrow 0$ であり、 $\{X^{(n)}\}$ は \mathcal{H}^2 のコーシー列である。

(Step 3) 補題 27 (つまり \mathcal{H}^2 の完備性) より、ある $X \in \mathcal{H}^2$ であって

$$\|X^{(n)} - X\|_T \rightarrow 0$$

となるものが存在する。この X は $H^{(n)}$ の取り方に依存しないことも証明できる。

(Step 4) このようにして決まる $X \in \mathcal{H}^2$ を H の M に関する確率積分といい、

$$H \bullet M_t = \int_0^t H_s dM_s := X_t$$

と書く。

10 半マルチンゲールの定義と例

定義 30 X が半マルチンゲールであるとは、ある有界変動適合過程 A と局所マルチンゲール M が存在して

$$X_t = X_0 + A_t + M_t$$

と書けるときにいう。この分解は一意ではない。

定義 31 X が特別半マルチンゲールであるとは、ある予測可能有界変動適合過程 A と局所マルチンゲール M が存在して

$$X_t = X_0 + A_t + M_t$$

と書けるときにいう。この分解は一意である。

例 32 (計数過程) 強度 $t \mapsto \lambda_t$ をもつ計数過程 $t \mapsto N_t$ は特別半マルチンゲールである。実際、

$$N_t = \int_0^t \lambda_s ds + (N_t - \int_0^t \lambda_s ds)$$

と書ける。

例 33 (拡散過程) β, σ を \mathbb{R} 上で定義された適当な条件を満たす関数であるとする。 $t \mapsto W_t$ を標準ウィナー過程であるとする。このとき

$$X_t = X_0 + \int_0^t \beta(X_s) ds + \int_0^t \sigma(X_s) dW_s$$

を満たす特別半マルチンゲール X が存在する。このような X は拡散過程の一例であり、 β をドリフト係数、 σ を拡散係数という。

半マルチンゲールの重要な点は、少なくとも次の2つある。まず、確率積分が自然に定義できる確率過程である点である：

$$\int_0^t H_s dX_s := \int_0^t H_s dA_s + \int_0^t H_s dM_s$$

$$H \bullet X_t := H \bullet A_t + H \bullet M_t$$

ただし右辺の第1項は実現値 ω ごとのルベーク・スティルチェス積分であり、第2項は前節で定義した確率積分である。

もうひとつは、次の節で紹介する「伊藤の公式」が使える点である。

11 連続半マルチンゲールに対する伊藤の公式

$X = (X^1, \dots, X^d)$ は連続な d -次元半マルチンゲールであるとする。つまり、各 $i = 1, \dots, d$ に対し

$$X_t^i = X_0^i + A_t^i + M_t^i$$

は連続な半マルチンゲールであるとする。 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ は2回連続的微分可能な関数であるとする。

定理 34 (伊藤の公式) 上述の設定のもとで、

$$f(X_t) - f(X_0) = \sum_{i=1}^d D_i f(X) \bullet X_t^i + \frac{1}{2} \sum_{i,j=1}^d D_{ij} f(X) \bullet \langle M^i, M^j \rangle_t$$

が成り立つ。ただし $D_i = \frac{\partial}{\partial x_i}$, $D_{ij} = \frac{\partial^2}{\partial x_i \partial x_j}$

証明の細部には触れないが、出発点はテイラー展開である。分点 $0 = t_0 < t_1 < \dots < t_n = t$ をとったとき

$$f(X_{t_k}) - f(X_{t_{k-1}}) = \sum_{i=1}^d D_i f(X_{t_{k-1}}) (X_{t_k}^i - X_{t_{k-1}}^i) + \frac{1}{2} \sum_{i,j=1}^d D_{ij} f(\xi_k) (X_{t_k}^i - X_{t_{k-1}}^i) (X_{t_k}^j - X_{t_{k-1}}^j)$$

が成り立つ。ただし ξ_k は $X_{t_{k-1}}$ と X_{t_k} を結ぶ線分上の点である。

実解析のテイラー展開はこの ξ という一種の「近似」が入ってしまうが、確率解析の伊藤の公式は近似が一切入らない恒等式であることが著しい点である。

伊藤の公式はジャンプがある半マルチンゲールにまで一般化されている。

12 連続マルチンゲールに対する中心極限定理

伊藤の公式の応用例として、連続マルチンゲールに対する中心極限定理を述べる。

定理 35 $M^n = (M^{n,(1)}, \dots, M^{n,(q)})$ は連続なマルチンゲールの q -次元ベクトルの列であるとし、 T_n を停止時刻の列であるとする。 $n \rightarrow \infty$ とするとき

$$(1) \quad \langle M^{n,(j)}, M^{n,(k)} \rangle_{T_n} \xrightarrow{p} C^{(j,k)} \quad (\text{定数})$$

であることを仮定する。このとき、

$$M_{T_n}^n \rightarrow^d N(0, C)$$

が成り立つ。ただし $C = (C^{(j,k)})_{j,k=1}^q$

証明の概略：一次元の場合を考える。多次元の場合はクラメル・ウォルドのデバイスによる。以下 $M^n = M^{n,(1)}$, $C = C^{(1,1)}$ と読み、積率母関数を計算する。 $z \in \mathbb{R}$ を固定する。

$$X_t = -\frac{z^2}{2} \langle M^n, M^n \rangle_t + z M_t^n$$

とおく。以下

$$E \exp(X_{T_n}) \approx 1$$

を示すことによって

$$E \exp(z M_{T_n}^n) \approx \exp\left(\frac{z^2}{2} C\right)$$

を示す。

さて、 $t \mapsto X_t$ は連続な半マルチンゲールであり

$$\langle z M^n, z M^n \rangle_t = z^2 \langle M^n, M^n \rangle_t$$

である。よって伊藤の公式を関数 $f(x) = \exp(x)$ に対して適用することにより

$$\begin{aligned} \exp(X_t) - 1 &= \exp(X) \bullet X_t + \frac{z^2}{2} \exp(X) \bullet \langle M^n, M^n \rangle_t \\ &= -\frac{z^2}{2} \exp(X) \bullet \langle M^n, M^n \rangle_t + z \exp(X) \bullet M_t^n \\ &\quad + \frac{z^2}{2} \exp(X) \bullet \langle M^n, M^n \rangle_t \\ &= z \exp(X) \bullet M_t^n \end{aligned}$$

である。右辺はマルチンゲールである。ここで停止時刻

$$S_n = \inf\{t \in [0, \infty) : \langle M^n, M^n \rangle_t \geq C\}$$

を導入する。このとき

$$\langle M^n, M^n \rangle_{S_n} = C$$

である。(ここは少し不正確である。 $\langle M^n, M^n \rangle_t$ の値が C に至らないこともあり得るからである。修正は演習問題とする。) よって

$$E \exp(X_{S_n}) - 1 = 0$$

より

$$E \exp(zM_{S_n}^n) = \exp\left(\frac{z^2}{2}C\right)$$

が得られた。従って

$$M_{S_n}^n \sim N(0, C)$$

である。後は

$$M_{T_n}^n - M_{S_n}^n \xrightarrow{p} 0$$

を示せばよいが、これは

$$Y_t^n = M_{t \wedge T_n}^n - M_{t \wedge S_n}^n$$

によって定義される二乗可積分マルチンゲール $t \mapsto Y_t^n$ に対して Lenglart の不等式 (付録 1) を適用し、仮定 (1) より

$$\begin{aligned} \langle Y^n, Y^n \rangle_\infty &= \langle M^n, M^n \rangle_{T_n} + \langle M^n, M^n \rangle_{S_n} - 2\langle M^n, M^n \rangle_{T_n \wedge S_n} \\ &\xrightarrow{p} 0 \end{aligned}$$

であることを用いればよい。 □

付録 1 : Lenglart の不等式

定理 36 X は局所劣マルチンゲールであるとし、その予測可能カンペンセイターを A とする。 T は停止時刻であるとする。このとき、任意の $\eta, \delta > 0$ に対し

$$P\left(\sup_{t \in [0, T]} X_t > \eta\right) \leq \frac{\delta}{\eta} + P(A_T > \delta)$$

が成り立つ。

特に M が局所二乗可積分マルチンゲールであるとき、 $t \mapsto M_t^2$ は局所劣マルチンゲールであり、その予測可能カンペンセイターは $\langle M, M \rangle$ であるから

$$\begin{aligned} P\left(\sup_{t \in [0, T]} |M_t| > \eta\right) &= P\left(\sup_{t \in [0, T]} |M_t|^2 > \eta^2\right) \\ &\leq \frac{\delta}{\eta^2} + P(\langle M, M \rangle_T > \delta) \end{aligned}$$

が成り立つ。

これらの不等式が著しいのは、左辺において t に関する supremum が確率の中に入っている点である。使い方としては「二乗可積分マルチンゲールが 0 に確率収束することを示したいときには、その二次変分が 0 に確率収束することを示せばよい」と覚えておけばよい。なお、劣マルチンゲール性の仮定は「Lenglart の支配性質」と呼ばれるものに置き換えることができる。

以下の証明では「予測可能時刻」の概念を用いる。それは停止時刻の特別な場合（アナウンス列の存在と呼ばれる良い性質をもつ停止時刻）であるが、詳しくは、例えば Jacod and Shiryaev (1987) を参照されたい。

証明：前半のみ示せば十分である。 $T_n = T \wedge n$ とおく。このとき

$$\lim_{n \rightarrow \infty} P \left(\sup_{t \in [0, T_n]} X_t > \eta \right) = P \left(\sup_{t \in [0, T]} X_t > \eta \right)$$

かつ

$$\lim_{n \rightarrow \infty} P \left(\sup_{t \in [0, T_n]} A_{T_n} > \delta \right) = P(A_T > \delta)$$

であるから、各 T_n に対して不等式を示せばよい。いいかえれば、停止時刻 T が有界であると仮定してよい。

$R = \inf(s : X_s > \eta)$ および $S = \inf(s : A_s > \delta)$ とおく。このとき、 R は停止時刻であり、 S は予測可能時刻である。 $\{\sup_{s \leq T} X_s > \eta\} \subset \{A_T > \delta\} \cup \{R \leq T < S\}$ であるから、

$$P \left(\sup_{t \in [0, T]} X_t > \eta \right) \leq P(R \leq T < S) + P(A_T > \delta)$$

である。ここで S は予測可能時刻であるから、ある停止時刻の列 (S_n) が存在して $S_n < S$ a.s. on $S > 0$ かつ $\lim_n S_n = S$ a.s. が成り立つ。(これをアナウンス列とよぶ。Jacod and Shiryaev (1987) の 1.2.16 参照。) よって

$$\begin{aligned} P(R \leq T < S) &\leq \lim_n P(R \leq T < S_n) \\ &\leq \lim_n P(X_{R \wedge T \wedge S_n} > \eta) \\ &\leq \frac{1}{\eta} \lim_n E(X_{R \wedge T \wedge S_n}) \\ &\leq \frac{1}{\eta} \lim_n E(A_{R \wedge T \wedge S_n}) \end{aligned}$$

である。(最後の不等式で X が劣マルチンゲールであることを使った。前述のように、この部分は「Lenglart の支配性質」に置き換えることができる。) さらに $A_{R \wedge T \wedge S_n} \leq A_{S_n} \leq \delta$ である (後の不等式でアナウンス列を導入したことを使った) から、不等式が証明された。
□

付録 2：拡散過程の母数推定（最も基本的なモデルとして）

確率微分方程式

$$X_t = X_0 + \int_0^t \beta(X_s, \theta) ds + \int_0^t \sigma(X_s) dW_s$$

を考える。ただし $\theta \in \Theta \subset \mathbb{R}^q$ は未知母数であるとし、拡散係数 σ は既知であるとする。

エルゴード性を仮定し、その不変分布を F とする。すなわち、任意の F -可積分関数 h に対し

$$\frac{1}{T} \int_0^T h(X_t) dt \longrightarrow \int_{\mathbb{R}} h(x) F(dx) \quad \text{a.s. as } T \rightarrow \infty$$

が成り立つとする。

この母数モデルにおいて、対数尤度比は

$$\log \frac{dP_\theta}{dP_{\theta_0}} \Big|_{\mathcal{F}_T} = l_T(\theta) - l_T(\theta_0)$$

となる。ただし

$$l_T(\theta) = \int_0^T \frac{\beta(X_t, \theta)}{\sigma(X_t)^2} dX_t - \frac{1}{2} \int_0^T \frac{\beta(X_t, \theta)^2}{\sigma(X_t)^2} dt$$

である。微分と積分の順序交換ができることを仮定すると、スコア統計量

$$U_T(\theta) = (U_T^1(\theta), \dots, U_T^q(\theta))$$

は

$$\begin{aligned} U_T^j(\theta) &= \frac{\partial}{\partial \theta_j} l_T(\theta) \\ &= \int_0^T \frac{\frac{\partial}{\partial \theta_j} \beta(X_t, \theta)}{\sigma(X_t)^2} dX_t - \int_0^T \frac{\beta(X_t, \theta) \frac{\partial}{\partial \theta_j} \beta(X_t, \theta)}{\sigma(X_t)^2} dt \end{aligned}$$

となる。最尤推定量 $\hat{\theta}_T$ は等式

$$U_T^j(\hat{\theta}_T) = 0, \quad j = 1, \dots, q$$

の解として定義される。観測フィッシャー情報量行列 $\mathcal{I}_T = (I_T^{jk})_{j,k=1}^q$ はスコア統計量をもう一回微分したもの（つまり対数尤度を二回微分したもの）にマイナスをつけたものである。

$$I_T^{jk}(\theta) = -\frac{\partial}{\partial \theta_k} U_T^j(\theta) = -\frac{\partial^2}{\partial \theta_j \partial \theta_k} l_T(\theta)$$

真値 θ_0 のもとで、これが

$$\frac{1}{T} I_T^{jk}(\theta_0) \rightarrow^p \sigma^{jk}(\theta_0) \quad \text{as } T \rightarrow \infty$$

となることを仮定し、

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow^d N(0, \Sigma^{-1}) \quad \text{as } T \rightarrow \infty$$

となることを証明する。ただし $\Sigma = (\sigma^{jk}(\theta_0))_{j,k=1}^q$ 。

この結果を基にして、

$$(\hat{\theta}_T - \theta_0)^\top \mathcal{I}_T(\hat{\theta}_T)(\hat{\theta}_T - \theta_0) \quad \text{Wald 検定統計量}$$

$$U_T(\theta_0)^\top \mathcal{I}_T(\theta_0)^{-1} U_T(\theta_0) \quad \text{スコア検定統計量}$$

$2(l_T(\hat{\theta}_T) - l_T(\theta_0))$ 尤度比検定統計量

が漸近的に自由度 q の χ^2 -分布に従うことが証明できる。
では主要部分に入ろう。

定理 37 (漸近正規性) $\hat{\theta}_T$ は $U_T(\theta) = 0$ の *consistent* な解であるとする。このとき、

$$\sqrt{n}(\hat{\theta}_T - \theta_0) \rightarrow^d N(0, \Sigma^{-1}) \quad \text{as } T \rightarrow \infty$$

が成り立つ。

証明の概略: $\theta \mapsto U_T^j(\theta)$ の θ_0 のまわりにおけるテイラー展開により、

$$\frac{1}{\sqrt{T}} U_T^j(\hat{\theta}_T) = \frac{1}{\sqrt{T}} U_T^j(\theta_0) - \sum_{k=1}^q \sqrt{T}(\hat{\theta}_T^k - \theta_{0,k}) \frac{1}{T} I_T^{jk}(\theta^*)$$

が得られる。ただし θ^* は $\hat{\theta}_T$ と θ_0 を結ぶ線分上にある。

左辺がゼロであることに注意すると、

$$\frac{1}{\sqrt{T}} U_T(\theta_0) \rightarrow^d N(0, \Sigma)$$

と

$$\frac{1}{T} I_T^{jk}(\theta^*) \rightarrow^p \sigma^{jk}(\theta_0)$$

を示せば十分である。後者は、もう一回テイラー展開した後、正則条件を用いて示す。

前者は、次のようにマルチンゲール中心極限定理を用いて示す。

$$\begin{aligned} \frac{1}{\sqrt{T}} U_T^j(\theta_0) &= \frac{1}{\sqrt{T}} \int_0^T \frac{\frac{\partial}{\partial \theta_j} \beta(X_t, \theta_0)}{\sigma(X_t)^2} dX_t \\ &\quad - \frac{1}{\sqrt{T}} \int_0^t \frac{\beta(X_t, \theta_0) \frac{\partial}{\partial \theta_j} \beta(X_t, \theta_0)}{\sigma(X_t)^2} dt \\ &= \frac{1}{\sqrt{T}} \int_0^T \frac{\frac{\partial}{\partial \theta_j} \beta(X_t, \theta_0)}{\sigma(X_t)^2} \sigma(X_t) dW_t \\ &= \frac{1}{\sqrt{T}} \int_0^T \frac{\frac{\partial}{\partial \theta_j} \beta(X_t, \theta_0)}{\sigma(X_t)} dW_t \end{aligned}$$

このマルチンゲールの二次変分は

$$\frac{1}{T} \int_0^T \frac{\frac{\partial}{\partial \theta_j} \beta(X_t, \theta_0) \frac{\partial}{\partial \theta_k} \beta(X_t, \theta_0)}{\sigma(X_t)^2} dt$$

となり、エルゴード性より、その極限はフィッシャー情報量

$$\int_{\mathbb{R}} \frac{\frac{\partial}{\partial \theta_j} \beta(x, \theta_0) \frac{\partial}{\partial \theta_k} \beta(x, \theta_0)}{\sigma(x)^2} F(dx)$$

に一致する。 □

付録 3 : 文献紹介

マルチンゲールの入門～中級といった部分を勉強したい方には Billingsley (1986) の 35, 37 節をお薦めする。

本稿は、連続時間確率過程論としてのマルチンゲール理論を解説することを意図したため、中級といった感じになった。執筆にあたって主として Jacod and Shiryaev (1987) の I 章と II 章を参考にした。同書は証明もほぼ全て書いてあり、自己完結的である。ただし記述は極めて抽象的である。半マルチンゲールに対する極限定理を本格的に勉強したければ同書が最もよい。極限がガウス過程である場合のみならず、一般の半マルチンゲールである場合まで、体系的に書いてある。

確率積分と伊藤の公式を深く勉強したければ Ikeda and Watanabe (1989) が最もよい。確率微分方程式の権威的文献である。限られたページ数でマルチンゲールの本質が理解できるように書いてある文献としては Durrett (1999) がある。

和書の中では長井 (1999) を一番にすすめる。後ろの方の章で確率制御や数理ファイナンスへの応用が述べられている。定評ある良書の和訳としては、カラザスとシュレーブ (2001) およびエクセンダー (1999) があり、いずれも記述が詳しく丁寧である。入門的にわかりやすく書いてある文献としては松原 (2003) がある。

なお、付録 2 において拡散過程の母数推定に触れたが、詳しくは Kutoyants (2004) を参照されたい。

参考文献

- [1] Billingsley, P. (1986): Probability and Measure. (2nd ed.) Wiley
- [2] Durrett, R. (1999): Essentials of Stochastic Processes. Springer
- [3] Ikeda, N. and Watanabe, S. (1989): Stochastic Differential Equations and Diffusion Processes. (2nd ed.) North-Holland/Kodansha
- [4] Jacod, J. and Shiryaev, A.N. (1987): Limit Theorems for Stochastic Processes. Springer
- [5] カラザス, I. と シュレーブ, S.E. [渡邊壽夫・訳] (2001): ブラウン運動と確率積分. シュプリンガー・フェアラーク東京
- [6] Kutoyants, Yu. (2004): Statistical Inference for Ergodic Diffusion Processes. Springer
- [7] 松原望 (2003): 入門確率過程. 東京図書
- [8] 長井英生 (1999): 確率微分方程式. 共立出版
- [9] エクセンダー, B. [谷口説男・訳] (1999): 確率微分方程式. シュプリンガー・フェアラーク東京