

# 自己励起POTモデルを用いた条件付きリ スク尺度について

天野裕介

東京大学大学院経済学研究科

## 概要

数百年に一度の大雨に備え、安全なダムや堤防の設計を行うという目的とともに極値統計学は発展してきた。通常の統計学の興味が降水量の平均や分散であるのに対し、ダムや堤防の設計のためには「最大の降水量」や「一定以上の降水の回数」に注目する必要がある。すなわち、「将来の降水量が平均的にどの程度なのか？」ではなく「将来見込まれる最大の降水量はどのくらいか？」、あるいは「将来一定水準以上の降水量は何回あるのか？」といった極端な事象（“極値事象”）に関する情報をデータから推測することが求められる。「降水量」を「クレーム額」や「金融資産の価格変動」などと読み替えることにより、将来起こりうる最大の（極端な）事象の振る舞いをとらえるために発展してきた極値統計学は保険や金融の分野でも研究が進められている。

本研究では日経平均、およびハンセン指数に関して、その値が指定された高い水準を超える下落をすることを極値事象であるととらえ、極値統計学の手法である POT (Peak Over Threshold, 閾値超過) モデル、特に自身の過去の閾値超過によってその後の閾値超過の起きやすさが増大するという性質、すなわち自己励起性をもつクラスを用いた分析を行う。この分析によって局所的なリスクの高まりを反映する柔軟なリスク尺度の構成が可能となる。

# 自己励起過程を用いた条件付きリスク尺度について

天野 裕介\*

## 1 導入

大雨に備え, 安全なダムや堤防の設計といった水文学に関連して極値統計学は発展してきた. 通常の統計学の興味が降水量の平均や分散であるのに対し, ダムや堤防の設計のためには「最大の降水量」や「一定以上の降水の回数」に注目する必要がある. すなわち, 「将来の降水量が平均的にどの程度なのか?」ではなく「将来見込まれる最大の降水量はどのくらいか?」, あるいは「将来一定水準以上の降水量は何回あるのか?」といった極端な事象(“極値事象”)に関する情報をデータから推測する必要がある.

「降水量」を「クレーム額」や「金融資産の価格変動」などと読み替えることにより, 将来起こりうる最大の(極端な)事象の振る舞いをとらえるための考え方(極値統計学)は保険や金融の分野でもリスク管理のための研究が進められている. 本研究では, 日経平均に関して, 価格の大幅な下落(-2%以上の変動)を極値事象であるにとらえ, 極値統計学の手法である POT (Peak Over Threshold, 閾値超過) モデル, 特に自身の過去の閾値超過によって超過の起きやすさが変化するクラスであるホークス過程を用いた分析を行う.

ホークス過程を用いて分析を行う対象となるデータについて簡単な解説を行う. Figure 1.1 は 1990 年 1 月から 2015 年 8 月までの各営業日の日経平均の収益率(終値-始値)/始値)を描いたものである. この図からは「平均的には」収益率はおおよそ 0 であることが読み取れる. 一方で本研究においては極値事象, すなわち Figure 1.2 に示されている「-2% (図の下側の水平線) より大きな下落」の大きさと発生タイミングに注目する.

本研究ではリスク管理の観点から, 以下の 2 点に注目して解説を行う.

1. 下落幅の程度  
⇒ “一般化パレート分布”
2. 閾値超過の起きるタイミング  
⇒ “点過程モデル”

---

\*東京大学大学院経済学研究科

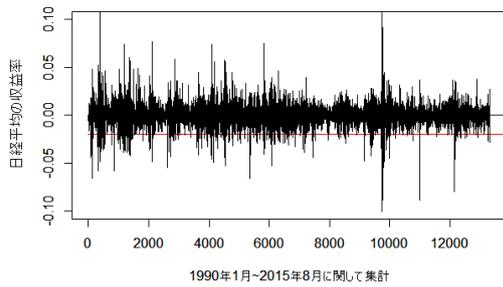


Figure 1.1: 日経平均の各日の収益率

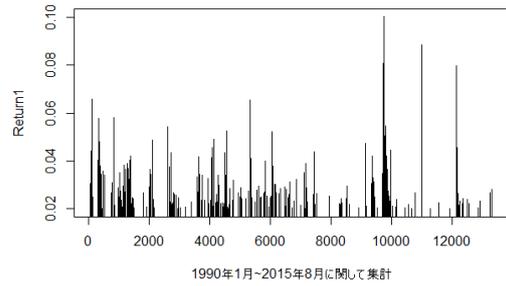


Figure 1.2: 極値事象のタイミングと絶対値

## 2 一般化パレート分布

保険、金融分野の統計的モデリングの際によく用いられる損失分布に従うデータのうち、閾値を超過したものの振る舞い方（分布）は、以下の”一般化パレート分布”を用いてモデル化される。

**定義 2.1.** (a) (一般化パレート分布)

閾値を  $\mu$  とする一般化パレート分布の確率分布関数は以下で与えられる。

$$F_{\mu, \sigma, \xi}(x) = \begin{cases} 1 - \left(1 + \xi \frac{(x-\mu)}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left(-\frac{(x-\mu)}{\sigma}\right) & \xi = 0 \end{cases}$$

各パラメータ  $\sigma, \xi$  はそれぞれ、尺度パラメータ、形状パラメータと呼ばれる。

(b) (超過分布)

データ  $X$  の従う損失分布を  $F$  とするとき、閾値  $\mu$  の超過分布  $F_\mu$  を

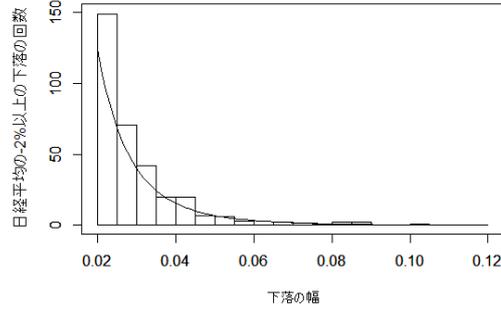
$$F_\mu(x) = P(X - \mu \leq x | X > \mu) = \frac{F(x + \mu) - F(\mu)}{1 - F(\mu)}$$

と定義する。

閾値を超過したデータの分布として一般化パレート分布を用いることの妥当性は Pickands-Balkema-de Haan の定理によって保証される。この定理はガンマ分布や対数正規分布といった保険、金融分野において多く用いられる分布に対して、その超過分布が閾値を上げるにしたがって一般化パレート分布へと収束することを意味する。詳細に関しては Embrechts, Klüppelberg and Mikosch (1997) を参照。

実際 1991 年 1 月から 2015 年 7 月までの日経平均の閾値超過のデータに対して一般化パレート分布をあてはめたときの分布形とそのヒストグラムを比べてみるとよく適合していることがわかる。

閾値超過データのヒストグラム



日経平均収益率の閾値超過データに対して一般化パレート分布を当てはめた場合の各パラメータは最尤法を用いて  $\sigma = 8.0 \times 10^{-3}$ ,  $\xi = 1.7 \times 10^{-2}$  と推定される. 直感的な解釈としては  $\sigma$  は実現値のスケール,  $\xi$  は特に大きな実現値の出やすさである.

**Remark 2.1.** 参考のために香港株式市場のハンセン指数の同様の期間のデータに対しても一般化パレート分布を当てはめてみると,  $\sigma = 8.1 \times 10^{-3}$ ,  $\xi = 1.5 \times 10^{-2}$  と推定される. 日本のものと比較してみると, スケールに関してはほぼ同様であるが,  $\xi$  が小さいことから日本よりは大きな変動は起こりにくい傾向にあることがわかる.

この特定化によって VaR (Value at Risk) などの種々のリスク指標の計算が行うことができる.

### 3 点過程モデル

次に閾値超過の時間軸に注目してみると, 閾値超過のタイミングは互いに無関係でないように考えられる. 特に閾値超過のタイミングは密集 (クラスタリング) する傾向がみとれる. このように閾値超過のようなイベントをその大きさにのみ注目するのではなく, 時間軸にも注目し, その発生の間隔をモデル化するための道具として POT (peaks over threshold) モデルが伝統的に用いられてきた. この章ではまず元データが i.i.d. である場合の最も単純な場合に関する POT モデルの解説を行う. 一方で多くの金融時系列データはこの仮定を満たさない, そのためより柔軟かつ発展的なモデルが必要となる. そのようなモデルの一つとしてホークス過程の解説を行う.

#### 3.1 i.i.d. データの閾値超過

閾値超過データの統計的なモデル化の際に用いられる点過程の理論について基本的な事柄を確認する. クレーム額などの損失を表す i.i.d 変数列  $(X_i)_{i \in \mathbb{N}}$  を考える. このとき共通の損失分布  $F$  はある極値分布の最大吸引域  $\text{MDA}(H_\xi)$ <sup>1</sup> に入っているとす. 閾値を

<sup>1</sup>したがってある非退化な極限分布  $H_\xi$  と基準化定数列  $a_n, b_n$  が存在して,

$$\lim_{n \rightarrow \infty} n \log F(a_n x + b_n) = \log H_\xi(x) \iff \lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = -\log H_{\xi}(x)$$

となる.

$u_n(x) = a_n x + b_n$  とおくと、 $X_1, \dots, X_n$  のうち  $u_n(x)$  を超えるものの個数は試行回数  $n$ , 確率パラメータ  $1 - F(u_n(x))$  の二項分布に従う。したがってポアソンの小数の法則から  $n \rightarrow \infty$  で、閾値を超えるものの個数の分布は平均  $-\log H_\varepsilon(x)$  のポアソン分布へと収束することがわかる。

次に点過程, 特にポアソン点過程についての数学的な定義を確認する。  $A$  を  $\mathbb{R}^n$  の任意の集合とする。このとき, 確率ベクトル  $X = (X_1, \dots, X_n)$  に対して,

$$N(A) := \sum_{i=1}^n I_A(X_i) \quad (3.1)$$

で定義される  $N(A)$  という確率変数を定義する。いくつかの技術的な条件のもとで,  $\mathbb{N}$  は点過程と呼ばれる。ただし  $I_A(\cdot)$  は集合  $A$  の指示関数であり,

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \in A^c \end{cases}$$

を意味する。直観的には  $N(A)$  は  $X_1, \dots, X_n$  のうちで領域  $A$  に入るものの個数を意味すると考えられる。

### 定義 3.1. (ポアソン点過程)

点過程  $N(\cdot)$  が以下の条件を満たすとき, 生起度測度  $\Lambda$  のポアソン点過程と呼ぶ。

1.  $A \subset \mathbb{R}$  と  $k \in \mathbb{N}$  に対して

$$P(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!} & \Lambda(A) < \infty \\ 0 & \Lambda(A) = \infty \end{cases}.$$

2. 任意の  $m \in \mathbb{N}$  に対して,  $A_1, \dots, A_m \subset \mathbb{R}$  が互いに排反であれば,  $N(A_1), \dots, N(A_m)$  は独立。

このとき,  $\Lambda(A) = \int_A \lambda(x) dx$  となるような,  $\lambda$  が存在すれば, それを点過程  $N(\cdot)$  の生起度関数という。

## 3.2 ホークス過程

与えられた時系列データが i.i.d. の場合に関して, 注目する期間における閾値超過の回数がポアソン点過程に従うことを確認した。しかし一方で現実の金融時系列データが i.i.d. であるとは考えにくい。そこで過去の閾値超過の履歴によって, 将来の閾値超過の起こりやすさが影響を受けるという性質<sup>2</sup>を持つ点過程の一つのクラスであるホークス過程に近年注目が集まっている。自己励起性をもつ過程へ最初に言及したのは Hawkes (1971) であり, 地震の発生モデリングに応用されてきたが, その柔軟性から近年金融分野での注目が高まっている。

<sup>2</sup>この性質を自己励起性という。

**定義 3.2.** (ホークス過程)

ホークス過程とは条件付き生起度が以下で与えられる点過程である.

$$\lambda(t) = \lambda_0 + \int_{-\infty}^t h(X_s, t, s) dN_s^*$$

すなわち, 時刻  $t$  までの時点  $t_1, \dots, t_n$  で大きさがそれぞれ  $X_1, \dots, X_n$  の閾値超過が起こったとすると, 条件付き生起度  $\lambda(t)$  は

$$\lambda(t) = \lambda_0 + \sum_{i=1}^n h(X_{t_i}, t, t_i)$$

で与えられる

**Remark 3.1.** 条件付き生起度はその時点における瞬間的は閾値超過の起こりやすさであると解釈できるので, 上記のホークス過程の定義は過去の履歴に従って将来の閾値超過の起こりやすさが影響を受けるという求めていた性質を反映できていることがわかる.

本論文では  $h(X_s, t, s) = \alpha X_s e^{-\gamma(t-s)}$  と特定し,

$$\lambda(t) = \lambda_0 + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^* \quad (3.2)$$

というモデルを用いる. このモデルを用いることで, 注目する時系列の閾値超過のタイミングが過去の履歴に影響から影響を受けるか否かが以下のようにして統計学的に検定できる.

1. 観測データとして閾値超過のタイミング  $t_1, \dots, t_n$  とその大きさ  $X_1, \dots, X_n$  が与えられたとき, 条件付き生起度が (3.2) で与えられるとするとその条件付き尤度関数は

$$L(\lambda_0, \alpha, \gamma \mid data) = \exp\left(-\int_0^{t_n} \lambda(s) ds\right) \prod_{i=1}^n \lambda(t_i)$$

で与えられる. 点過程の条件付き尤度関数の導出に関しては Daley and Vere-Jones (2007) に詳しい.

2. 数値計算プログラムによって上記の条件付き尤度関数を  $(\lambda_0, \alpha, \gamma)$  に関して最大化し, 最大値を達成する最尤推定量  $(\hat{\lambda}_0, \hat{\alpha}, \hat{\gamma})$  を求める.
- 3-(1). 最尤推定量における条件付き尤度関数のヘシアンから各最尤推定量の標準偏差をもとめ,  $t$ -値を計算する.
- 3-(2). あるいは各パラメータのとりうる値を制限した上で条件付き尤度関数の最大化を行い, 条件なしで最大化された尤度関数との比較による尤度比検定を行う.

**Remark 3.2.** 特に  $\alpha = 0$  であるとする, 条件付き生起度関数が  $\lambda(t) = \lambda_0$  となって過去の経歴にも時刻にも依存しない (閾値超過は過去の系列と無関係に起こる.) という場合に対応する. すなわち  $\alpha = 0$  という仮説の検定を行えば, 自己励起性の有無を確認できる.

実際のデータ（日経平均）に対して、上記のホークス過程のモデルをあてはめてみると以下ようになる。()内は  $t$ -値。

$$\lambda_0 = 9.0 \times 10^{-3} \text{ (6.5)}, \alpha = 5.7 \times 10^{-1} \text{ (5.9)}, \gamma = 2.6 \times 10^{-2} \text{ (5.8)}.$$

$\alpha$  の  $t$ -値に注目すれば、日経平均の閾値超過には自己励起性があること、言い換えれば過去の閾値超過によって瞬間的な閾値超過の可能性が高まることによるクラスタリングの傾向をもつことが統計的に確認できた。

**Remark 3.3.** 参考のためハンセン指数についても同じホークス過程のモデルをあてはめてみると以下ようになる。()内は  $t$ -値。

$$\lambda_0 = 4.8 \times 10^{-3} \text{ (5.0)}, \alpha = 4.3 \times 10^{-1} \text{ (5.2)}, \gamma = 1.6 \times 10^{-2} \text{ (5.1)}.$$

日経平均のものと比べると  $\lambda_0$ , すなわち条件付き生起度関数の定常水準が小さく、比較的閾値超過が起こりにくいということがわかる。また  $\gamma$  の値も日経平均のものより小さく、これは過去に起きた閾値超過の影響が日本に比べて減衰しにくいことを示している。

## 4 自己励起過程を用いた条件付きリスク尺度

前節において金融時系列の閾値超過にはクラスタリングの傾向を持ちうることを確認した。そこで McNeil et al. (2015) の第7章の議論を参考に、局所的なリスクの高まりを反映するような動学的リスク尺度の提案を行う。そのために過去の履歴に尺度パラメータが依存するような一般化パレート分布を持つことを許す、予測可能なマークを持つモデルを考える。

**定義 4.1.** (予測可能なマーク)

状態空間  $\mathcal{X} = (0, \infty] \times (\mu, \infty)$  上の点過程で、条件付き生起度

$$\lambda(t, x) = \frac{\lambda_0 + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^*}{\sigma + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^*} \left( 1 + \xi \frac{x - \mu}{\beta + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^*} \right)^{-1/\xi - 1}$$

を持つものを予測可能なマークを持つモデルという。

この予測可能なマークを持つモデルを用いて、動学的な条件付きリスク尺度の提案を行うことができる。たとえば代表的なリスク尺度である  $z\%$ -VaR は再帰水準問題として

$$\begin{aligned} \Lambda(x) &= \int_x^\infty \lambda(t, y) dy \\ &= \left( \lambda_0 + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^* \right) \left( 1 + \xi \frac{x - \mu}{\beta + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^*} \right)^{-1/\xi} \\ &= 1 - z \end{aligned}$$

を  $x \geq \mu$  に関して解くことによって得られる。

**Proposition 4.1.** (動学的条件付き VaR)

時点  $t$  までの観測値に基づく動学的条件付き VaR は

$$VaR_z^t = \mu + \frac{\left(\lambda_0 + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^*\right)}{\xi} \left\{ \left( \frac{\lambda_0 + \alpha \int_{-\infty}^t X_s e^{-\gamma(t-s)} dN_s^*}{1-z} \right)^{1/\xi} - 1 \right\}$$

で与えられる.

また同様に, 損失額が VaR を超えた場合における損失額の平均値である期待ショートフォールに関する式

$$\frac{1}{1-\alpha} \int_z^1 VaR_y^t dy$$

を計算することにより動学的なリスク尺度へと拡張できることがわかる.

## 5 まとめ

本研究ではまず金融時系列のホークス過程によるモデリングによって金融データの極値事象の発生におけるクラスタリング傾向の有無の確認方法について解説し, 日経平均における閾値超過のクラスタリング傾向の確認を行った. 加えてさらにクラスタリングによって局所的に高まりうるリスクをとらえるための指標として動学的 VaR の紹介を行った.

## References

- Daley, D.J. and Vere-Jones, D. (2007). "An Introduction to Point Processes: Volume I: Elementary theory and Methods, Second Edition." *Springer*.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). "Modeling Extremal Events for Insurance and Finance." *Springer, Berlin*.
- Grothe, O. Korniiichuk, V. Manner, H. (2014). "Modeling multivariate extreme events using self-exciting point process." *Jornal of Econometrics* 182 269-289.
- Hawkes, A.G. (1971). "Point spectra of some mutuality exciting point processes.." *J. R. Stat. Socm Ser. B (Stat. Methodol.)*, Vol. 26, No. 1.
- McNeil, A. J. et.al (2015) "Quantitative Risk Management: Concepts, Techniques and Tools." *Princeton Series in Finance*.
- Resnick, S.I. (2007). "Heavy-tail phenomena: probabilistic and statistical modeling." *Journal of the American Statistical Association*.