

# 非対称分布族と一般化双曲型分布族の有限混合による 株式時価総額の統計モデリング

地 道 正 行

## 要 旨

本稿では、地道、阪（2021-a, b）で扱われた対数株式時価総額のデータに対して、探索的データ解析の視点に立ち、データの可視化から得られた知見にもとづいて、非対称分布族と一般化双曲型分布族に属する分布を当てはめることによって統計モデリングを行う。とくに、国のタイプ（先進国、新興国）に属する企業に分類した場合も考慮し、有限混合分布の当てはめや、経年変化を考慮した考察も行う。なお、本研究は動的文書生成によって再現可能研究の立場から実施される。

キーワード：財務データ (Financial Data), 非対称分布族 (Family of Skew-Symmetric Distributions), 一般化双曲型分布族 (Family of Generalized Hyperbolic Distributions), 探索的データ解析 (Exploratory Data Analysis), 再現可能研究 (Reproducible Research)

## I はじめに

地道、阪（2021-a）では、Bureau van Dijk社から提供されるデータベース Osiris から抽出された世界154カ国の上場企業の企業財務データ<sup>1)</sup>と FTSE Russell社の ESG レーティングデータの結合について議論されており、さらに地道、阪（2021-b）では、このデータにおける株式時価総額の対数（対数

---

1) Osiris から抽出されたデータの事前処理や統計モデリングについては、地道（2018-a, b）を参照のこと。

株式時価総額)の分布と両対数モデルによる統計モデリングについて詳細に考察されている。この論文において、対数株式時価総額の可視化の結果から、データは非対称分布、特に、非対称ティータ分布の当てはまりが良いという結果を得ており、さらに先進国と新興国(国のタイプ)のそれぞれについても、非対称分布に従っていることが推察されている。このことから、国タイプ別に分類(分割)した企業の対数株式時価総額が、それぞれ非対称分布に従うことが示唆され、それぞれ非対称分布に従う独立な変量(今の場合は対数株式時価総額)を合併したものが、再び同族の非対称分布に従うかどうかという命題を提起していることになる。この命題は、確率分布に関する無限分解可能性<sup>2)</sup>(infinitely divisible)や再生性<sup>3)</sup>(reproductive property)と関連があるが、残念ながら、非対称正規分布は無限分解可能でなく、再生性ももたないことも地道、阪(2021-b)では指摘されている<sup>4)</sup>。

この問題について本稿では、一般化双曲型分布族に属する再生性を持つ分布が当てはめに利用できるかを検討するとともに、非対称分布族と一般化双曲型分布族に属する分布の有限混合を利用した統計モデリングを考察する。

本稿の構成は以下のようなものである。まず、2018会計年度の対数株式時価総額のデータ及び、国のタイプ別に分類したものを可視化する(II節)ことによって得られる分布構造についての知見を使って、分布の統計モデリングを行う(III節)。モデリングに際しては、非対称正規分布、非対称ティータ分布(非対称分布族)、正規逆ガウス分布(一般化双曲型分布族)を考え、情報量規準によってモデル選択を行う。次に、国のタイプ(先進国、新興国)別に分けられたデータに同様の分布を当てはめ、それぞれの場合に対して、

2) 確率分布に関する無限分解可能性の定義については、増田(2002)等を参照されたい。

3) 一般に、再生性とは、母数 $\theta$ をもつある分布 $P_\theta$ に対して、分布族 $\mathcal{P}_\theta := \{P_\theta | \theta \in \Theta\}$ を考え、確率変数 $X, Y$ が、それぞれ独立に分布 $P_{\theta_1}, P_{\theta_2}$ に従うとき、和 $X+Y$ の分布が再び同一の分布族に属する分布 $P_{\theta_1+\theta_2} \in \mathcal{P}_\theta$ に従うことをいう。ここで、 $\Theta$ は母数空間である。

4) 非対称正規分布が無限分解可能でないことは、Kozubowska and Nolan(2008)のProposition 2.1で示されており、再生性をもたないことはAzzalini and Capitanio(2014)のp. 27を参照されたい。

情報量規準によってモデル選択を行う (IV 節). ここでは, 国タイプ別に分類した企業の対数株式時価総額の再生性に関する命題の検証も行う. 以上の考察を踏まえて, 国タイプ別データに当てはめた分布を混合した有限混合分布を当てはめることを検討する (V 節). さらに, ここまで検討したモデルの当てはまりの経年変化を考察する (VI 節). 最後に本稿で得られた結果の総括と, 今後の展望について述べる (VII 節).

追加情報として, 本稿で利用しているデータについての要約や変数の説明, 先進国, 新興国の分類, 欠測情報を与える (付録 A). また, 非対称分布族 (付録 B), 一般化双曲型分布族 (付録 C), 混合分布 (付録 D) についての簡単な解説と一般化双曲型分布族と深く関わりをもつ変形ベッセル関数についての性質を与える (付録 E). さらに, 本稿の結果を再現するための R スクリプト (付録 F) と, 利用したコンピュータ環境について与えている (付録 G).

本研究は, データ解析環境 R を利用しており, Tukey (1977) による探索的データ解析 (Exploratory Data Analysis: EDA) の視点に立って行われている<sup>5)</sup>. また, 本稿の執筆には, Sweave<sup>6)</sup> による動的文書生成によって再現可能研究を実施することを試みている<sup>7)</sup>.

## II データ可視化

本稿で利用するデータは, 地道, 阪 (2021-a) で扱われた財務データと ESG レーティングデータを結合したものである. なお, 付録 A にもデータに関する情報を与えているので参照されたい.

本節では, データを幾つかの観点から可視化することによって, 問題点の

---

5) R については, 例えば, 地道 (2018-c) を参照されたい. また, 財務データに関する EDA の実行については, 地道 (2018-a, b) も参照されたい.

6) <https://stat.ethz.ch/R-manual/R-devel/library/utils/doc/Sweave.pdf>

7) Sweave を利用した動的文書生成と再現可能研究については, 例えば, 地道 (2018-b) を参照されたい.

指摘を行う。

### 1. 対散布図による可視化

2018会計年度の財務データの対数をとったものとESGレーティングデータの相関構造を対散布図によって可視化したものを図1に与える。この図では、国のタイプ別に色分け（新興国：E（暖色），先進国：G（寒色））しており，それによって，新興国と先進国の分布と相関構造の違いが見て取れる。

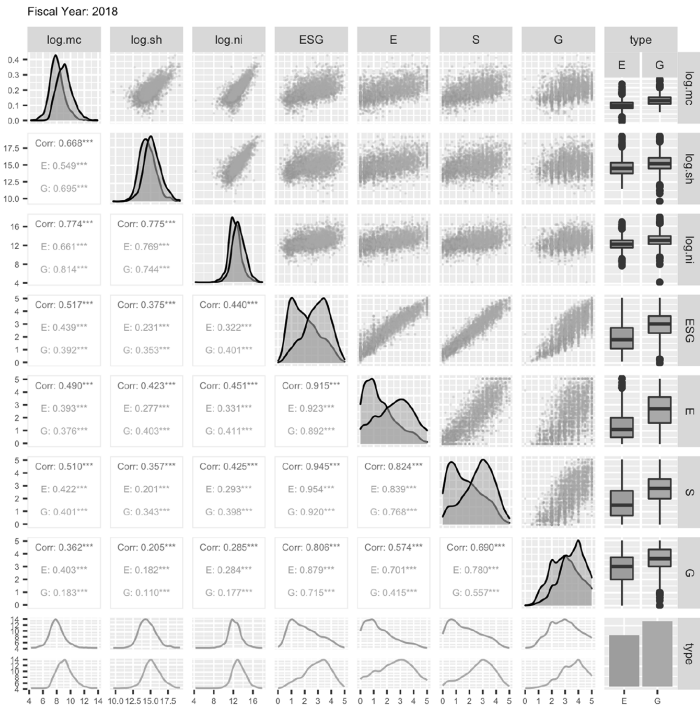


図1：対散布図：対数スケール

特に，図1の(1, 1)ブロックの対数株式時価総額 (log.mc) の推定された密度関数の相違を考慮した統計モデリングを行うことが，本稿のテーマである。

## 2. 対数株式時価総額の分布に関する可視化

2018会計年度の対数株式時価総額のヒストグラムと推定された密度関数、正規 Q-Q プロットを描くことによって、分布構造を考察する。まず、推定された密度関数をヒストグラムに重ねて描いたプロットを図 2 に与える。

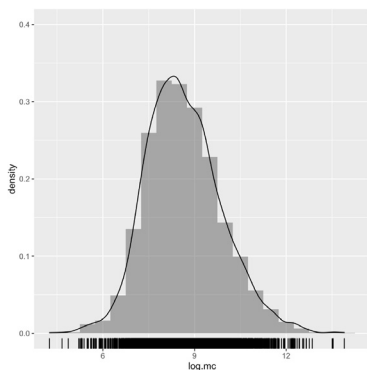


図 2：対数株式時価総額のヒストグラムと推定された密度関数：ラグ付き

図 2 より、対数株式時価総額は正規分布に近い構造を持つが、若干右に歪んでいることが観察される。

このことをさらに詳細に調べるために、対数株式時価総額の正規 Q-Q プロットを図 5 に与える。

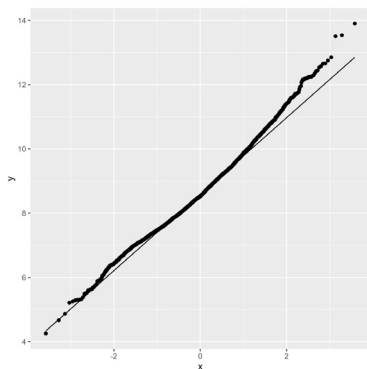


図 3：対数株式時価総額の正規 Q-Q プロット

正規 Q-Q プロットは、右裾の部分での当てはまりが顕著に悪いことがわかる。さらに、対数株価時価総額の歪度を計算すると0.39となり、右に歪んでいることがわかる。以上の結果から、対数株価時価総額は正規分布より若干右に歪んだ分布に従うことがわかる。なお、地道、阪（2021-b）では、この結果を利用して、非対称正規分布と非対称ティー分布を当てはめている。

一方、対数株式時価総額のヒストグラムと推定された密度関数を国のタイプ（新興国：E、先進国：G）別にプロットしたものを図4に与える（対散布図1の（1,1）ブロックも参照のこと）。

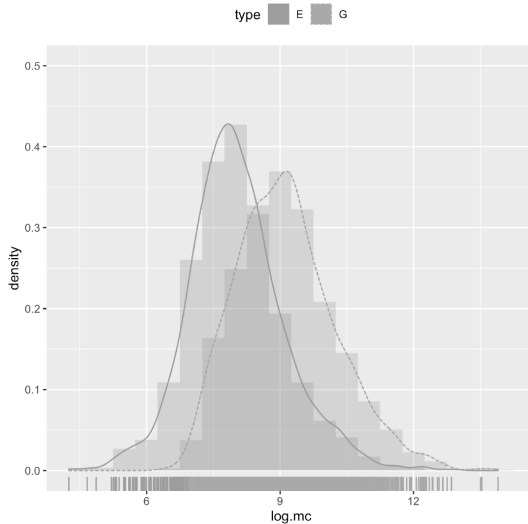


図4：対数株式時価総額のヒストグラムと推定された密度：国タイプ別（新興国：E（実線）、先進国：G（点線））、ラグ付き

この結果から、新興国（実線）と先進国（点線）に対する対数株式時価総額の分布は、共に正規分布ではないことが推察される。このことを対数株式時価総額の正規 Q-Q プロットを描くことによって精査する。

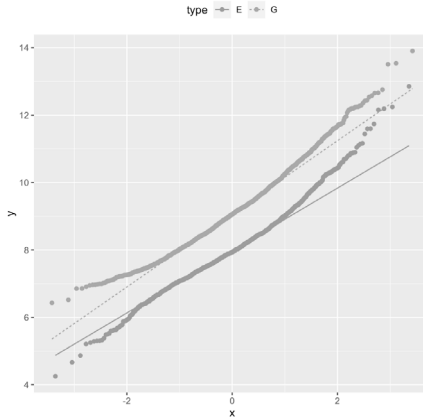


図5：対数株式時価総額の正規 Q-Q プロット：国タイプ別

この結果（図5）からも，新興国（実線）と先進国（点線）に対する対数株式時価総額の分布は，共に正規分布に従っていない，歪んだものであることがわかる。

次節では，これらの結果から得られた知見を利用して対数株式時価総額の分布の統計モデリングを行う。

### III 対数株式時価総額の分布に関する統計モデリング

ここで，前節で得られた対数株式時価総額の分布が歪んだ分布であることから，非対称分布族として，非対称正規分布，非対称ティー分布を考え（cf. 地道（2017-a, b），Jimichi *et al.*（2018），地道，阪（2021-b）），さらに，一般化双曲型分布族に属する正規逆ガウス分布も当てはめる。これは，国タイプ別に分けた対数株式時価総額のそれぞれの分布と，それらを合算した分布が同じ分布族に属する（再生性）という現象を，ある種の再生性を持つ非対称分布である正規逆ガウス分布を用いて説明できる可能性があるからである。なお，非対称正規分布と非対称ティー分布を含む非対称分布族については，付録Bを，正規逆ガウス分布を含む一般化双曲型分布族については，付録Cに簡単な説明を与えているので参照されたい。

## 1. 非対称正規分布の当てはめ

ここでは、2018会計年度末の対数株式時価総額  $\log(\text{market\_cap})$  に非対称正規分布  $\text{SN}(\xi, \omega^2, \alpha)$  を当てはめる。非対称正規分布の直接母数 (direct parameter)  $(\xi, \omega, \alpha)$  の最尤推定値は以下のように与えられる：

$$(\hat{\xi}, \hat{\omega}, \hat{\alpha}) = (7.48, 1.7, 1.73)$$

これらの推定値を確率密度関数の母数に代入した統計モデル (statistical model) は、

$$\begin{aligned} & f_{\text{SN}}(\log(\text{market\_cap}) | \hat{\xi}, \hat{\omega}, \hat{\alpha}) \\ &= \frac{2}{\hat{\omega}} \phi\left(\frac{\log(\text{market\_cap}) - \hat{\xi}}{\hat{\omega}}\right) \Phi\left(\hat{\alpha} \frac{\log(\text{market\_cap}) - \hat{\xi}}{\hat{\omega}}\right) \\ &= 1.18 \phi\left(\frac{\log(\text{market\_cap}) - 7.48}{1.7}\right) \Phi\left(1.73 \frac{\log(\text{market\_cap}) - 7.48}{1.7}\right) \end{aligned} \quad (1)$$

となる。ここで、 $\phi(\cdot)$  と  $\Phi(\cdot)$  は、それぞれ、標準正規分布の確率密度関数と累積分布関数である (定義については、付録 B.1 を参照のこと)。

ヒストグラムに統計モデル (1) を重ねて描いたものと、標準化直接母数残差の2乗の P-P プロットを、それぞれ、図 6, 7 に与える。

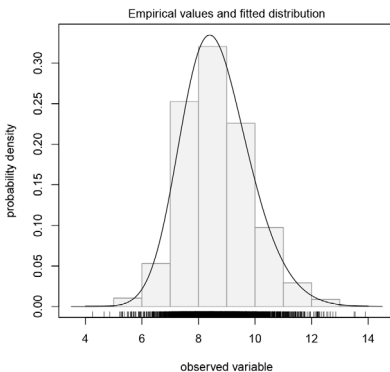


図 6：株式時価総額の対数のヒストグラムと非対称正規分布にもとづく統計モデル

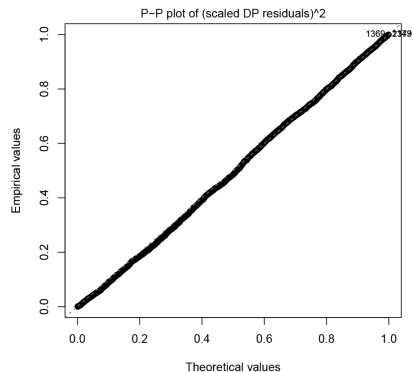


図 7：株式時価総額の対数に非対称正規分布を当てはめたときの標準化直接母数残差の2乗の P-P プロット



これらの可視化の結果から、非対称正規分布は対数株式時価総額によく当てはまっていることがわかる。

## 2. 非対称ティー分布の当てはめ

2018会計年度末の対数株式時価総額  $\log(\text{market\_cap})$  に非対称ティー分布  $ST(\xi, \omega^2, \alpha, \nu)$  を当てはめる。直接母数  $(\xi, \omega, \alpha, \nu)$  の最尤推定値は以下のように与えられる：

$$(\hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu}) = (7.57, 1.55, 1.53, 19.54)$$

これらの推定値より、統計モデルは、

$$\begin{aligned} & f_{ST}(\log(\text{market\_cap}) | \hat{\xi}, \hat{\omega}, \hat{\alpha}, \hat{\nu}) \\ &= \frac{2}{\hat{\omega}} f_t \left( \frac{\log(\text{market\_cap}) - \hat{\xi}}{\hat{\omega}} \middle| \hat{\nu} \right) \cdot \\ & F_t \left( \hat{\alpha} \frac{\log(\text{market\_cap}) - \hat{\xi}}{\hat{\omega}} \sqrt{\frac{\hat{\nu} + 1}{\left( \frac{\log(\text{market\_cap}) - \hat{\xi}}{\hat{\omega}} \right)^2 + \hat{\nu}}} \middle| \hat{\nu} + 1 \right) \\ &= 1.29 f_t \left( \frac{\log(\text{market\_cap}) - 7.57}{1.55} \middle| 19.54 \right) \cdot \\ & F_t \left( 1.53 \frac{\log(\text{market\_cap}) - 7.57}{1.55} \sqrt{\frac{20.54}{\left( \frac{\log(\text{market\_cap}) - 7.57}{\hat{\omega}} \right)^2 + 19.54}} \middle| 20.54 \right) \end{aligned} \tag{2}$$

となる。ここで、 $f_t(\cdot)$  と  $F_t(\cdot)$  は、それぞれ、ティー分布の確率密度関数と累積分布関数である（定義については、付録 B.2 を参照のこと）。

ヒストグラムに統計モデル (2) を重ねて描いたものと、標準化直接母数残差の 2 乗の P-P プロットを、それぞれ、図 8, 9 に与える。

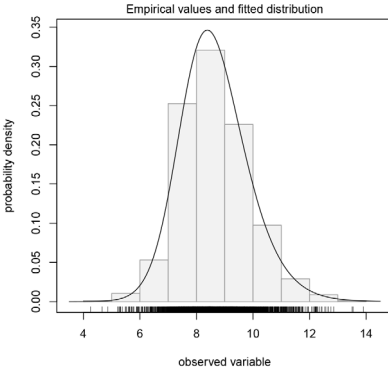


図8：株式時価総額の対数のヒストグラムと非対称ティー分布にもとづく統計モデル

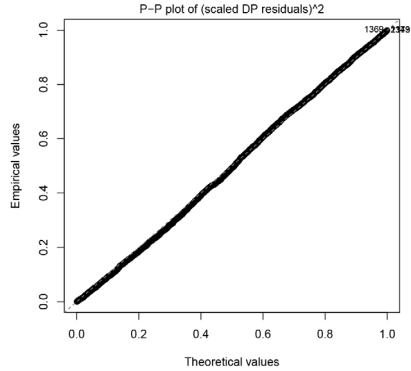


図9：株式時価総額の対数に非対称ティー分布を当てはめたときの標準化直接母数残差の2乗のP-Pプロット

これらの可視化の結果から、非対称ティー分布も非対称正規分布と同様に対数株式時価総額によく当てはまっていることがわかり、可視化の結果だけでは優劣は付けがたいことがわかる。

### 3. 正規逆ガウス分布の当てはめ

2018会計年度末の対数株式時価総額  $\log(\text{market\_cap})$  に正規逆ガウス分布  $\text{NIG}(\alpha, \beta, \delta, \mu)$  を当てはめる。母数  $(\alpha, \beta, \delta, \mu)$  の最尤推定値は以下のように与えられる<sup>8)</sup>：

$$(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu}) = (2.38, 0.92, 2.83, 7.45) \quad (3)$$

これらの推定値より、統計モデルは、

$$f_{\text{NIG}}(\log(\text{market\_cap}) | \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu})$$

8) ここで、正規逆ガウス分布  $\text{NIG}(\alpha, \beta, \delta, \mu)$  の母数の最尤推定は、Rパッケージ **GeneralizedHyperbolic** に付属する `nigFit` 関数を利用している。この関数が返す結果は、推定値の順序が

$$(\hat{\mu}, \hat{\delta}, \hat{\alpha}, \hat{\beta}) = (7.45, 2.83, 2.38, 0.92)$$

となっており、ヒストグラムに統計モデルを描くプロット(図10)に付与される母数の推定値 (`param = (...)`) の順序となっていることに注意する必要がある。

$$\begin{aligned}
 &= \frac{\hat{\alpha}\hat{\delta}}{\pi} \exp\{\hat{\delta}\sqrt{\hat{\alpha}^2 - \hat{\beta}^2} + \hat{\beta}(\log(\text{market\_cap}) - \hat{\mu})\} \cdot \\
 &\quad \frac{K_1(\hat{\alpha}\sqrt{(\log(\text{market\_cap}) - \hat{\mu})^2 + \hat{\delta}^2})}{\sqrt{(\log(\text{market\_cap}) - \hat{\mu})^2 + \hat{\delta}^2}} \\
 &= \frac{6.73}{\pi} \exp\{6.2 + 0.92(\log(\text{market\_cap}) - 7.45)\} \cdot \\
 &\quad \frac{K_1(2.38\sqrt{(\log(\text{market\_cap}) - 7.45)^2 + 8.02})}{\sqrt{(\log(\text{market\_cap}) - 7.45)^2 + 8.02}} \tag{4}
 \end{aligned}$$

となる。ここで、 $K_\nu(\cdot)$  は第2種変形ベッセル関数である（定義については、付録Eを参照のこと）。

ヒストグラムに統計モデル（4）を重ねて描いたものと、対数株式時価総額のP-Pプロットを、それぞれ、図10、11に与える。

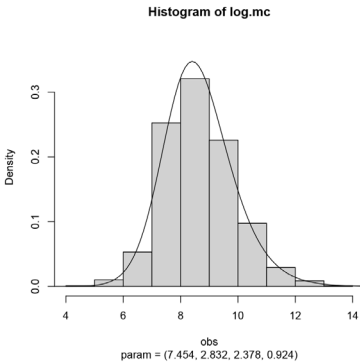


図10：対数株式時価総額のヒストグラムと正規逆ガウス分布にもとづく統計モデル

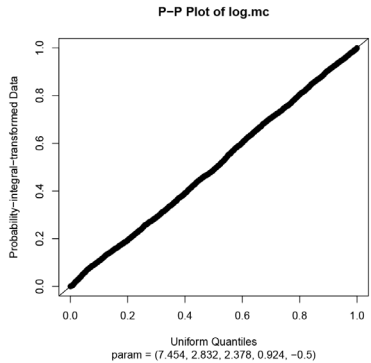


図11：対数株式時価総額に正規逆ガウス分布を当てはめたときのP-Pプロット

これらの可視化の結果からも、非対称正規分布、非対称ティー分布の場合と同様に対数株式時価総額によく当てはまっていることがわかり、可視化の結果だけでは優劣は付けがたいことがわかる。以上のことから、次に情報量規準によって比較する。

#### 4. 対数株式時価総額の分布に関するモデル選択

ここでは、対数株式時価総額に対して非対称正規 (Skew-normal: SN) 分布, 非対称ティー (Skew-t: ST) 分布, 正規逆ガウス (Normal Inverse Gaussian: NIG) 分布を当てはめの良さを赤池情報量規準 (Akaike's Information Criterion: AIC), ベイズ情報量規準 (Bayesian Information Criterion: BIC) を用いて比較・検討する<sup>9)</sup>.

表 1: AIC, BIC 表: 対数株式時価総額への分布  
当てはめに関する情報量規準の比較

	dim	AIC	BIC
SN	3	9306.21	9324.11
ST	4	9299.10	9322.97
NIG	4	9301.20	9325.06

表 1 から, 非対称ティー (ST) 分布を当てはめた場合が最も良いという結果 (AIC, BIC の値が小さい) が得られた. よって, (先進国と新興国を合わせた企業の) 対数株式時価総額の分布としては非対称ティー分布の当てはまりがよいということがわかった.

#### IV 対数株式時価総額の国タイプ別分布に関する統計モデリング

前節では, 先進国と新興国を合わせた企業の対数株式時価総額に対して非対称正規分布, 非対称ティー分布, 正規逆ガウス分布を当てはめた. 本節では, データを国のタイプ (先進国: G, 新興国: E) に分けたものに対して, これらの分布を当てはめることによって, より詳細に分布構造を探る.

##### 1. 非対称正規分布の国タイプ別当てはめ

###### 先進国

ここでは, 先進国 (G) の2018会計年度末の対数株式時価総額  $\log(\text{mar-}$

9) 情報量規準全般についての詳細は, Konishi and Kitagawa (2008) を参照されたい.

cket\_cap<sub>g</sub>) に非対称正規分布  $SN(\xi_g, \omega_g^2, \alpha_g)$  を当てはめる。非対称正規分布の直接母数  $(\xi_g, \omega_g, \alpha_g)$  の最尤推定値は以下のように与えられる：

$$(\hat{\xi}_g, \hat{\omega}_g, \hat{\alpha}_g) = (7.9, 1.67, 2.49)$$

これらの推定値を確率密度関数の母数に代入した統計モデルは、

$$\begin{aligned} & f_{SN}(\log(\text{market\_cap}_g) | \hat{\xi}_g, \hat{\omega}_g, \hat{\alpha}_g) \\ &= \frac{2}{\hat{\omega}_g} \phi\left(\frac{\log(\text{market\_cap}_g) - \hat{\xi}_g}{\hat{\omega}_g}\right) \Phi\left(\hat{\alpha}_g \frac{\log(\text{market\_cap}_g) - \hat{\xi}_g}{\hat{\omega}_g}\right) \\ &= 1.2 \phi\left(\frac{\log(\text{market\_cap}_g) - 7.9}{1.67}\right) \Phi\left(2.49 \frac{\log(\text{market\_cap}_g) - 7.9}{1.67}\right) \quad (5) \end{aligned}$$

となる。

ヒストグラムに統計モデル (5) を重ねて描いたものと、標準化直接母数残差の2乗の P-P プロットを、それぞれ、図12、13に与える。

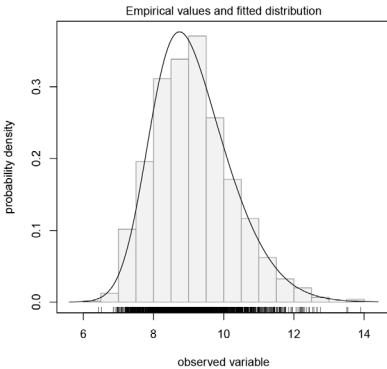


図12：株式時価総額対数のヒストグラムと非対称正規分布にもとづく統計モデル：先進国

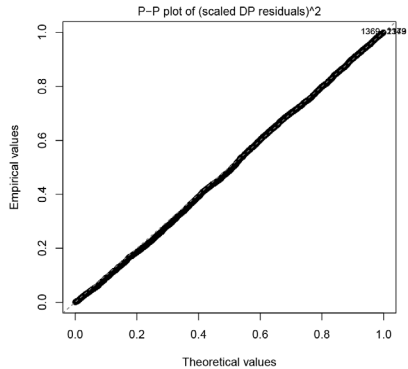


図13：株式時価総額対数に非対称正規分布を当てはめたときの標準化直接母数残差の2乗の P-P プロット：先進国

これらの可視化の結果から、総じて非対称正規分布は先進国の企業に対する対数株式時価総額に当てはまっていることがわかる。

### 新興国

ここでは、新興国 (E) の2018会計年度末の対数株式時価総額  $\log(\text{mar-}$

cket\_cap<sub>E</sub>) に非対称正規分布  $SN(\xi_E, \omega_E^2, \alpha_E)$  を当てはめる。非対称正規分布の直接母数  $(\xi_E, \omega_E, \alpha_E)$  の最尤推定値は以下のように与えられる：

$$(\hat{\xi}_E, \hat{\omega}_E, \hat{\alpha}_E) = (7.05, 1.45, 1.59)$$

これらの推定値を確率密度関数の母数に代入した統計モデルは、

$$\begin{aligned} & f_{SN}(\log(\text{market\_cap}_E) | \hat{\xi}_E, \hat{\omega}_E, \hat{\alpha}_E) \\ &= \frac{2}{\hat{\omega}_E} \phi\left(\frac{\log(\text{market\_cap}_E) - \hat{\xi}_E}{\hat{\omega}_E}\right) \Phi\left(\hat{\alpha}_E \frac{\log(\text{market\_cap}_E) - \hat{\xi}_E}{\hat{\omega}_E}\right) \\ &= 1.38 \phi\left(\frac{\log(\text{market\_cap}_E) - 7.05}{1.45}\right) \Phi\left(1.59 \frac{\log(\text{market\_cap}_E) - 7.05}{1.45}\right) \end{aligned} \quad (6)$$

となる。

ヒストグラムに統計モデル (6) を重ねて描いたものと、標準化直接母数残差の 2 乗の P-P プロットを、それぞれ、図14, 15に与える。

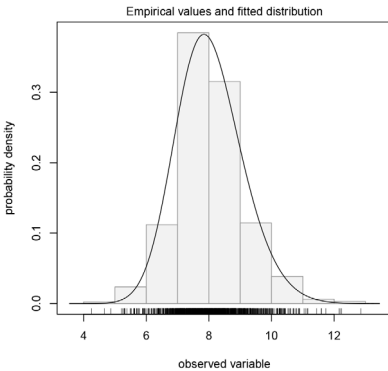


図14：株式時価総額の対数のヒストグラムと非対称正規分布にもとづく統計モデル：新興国

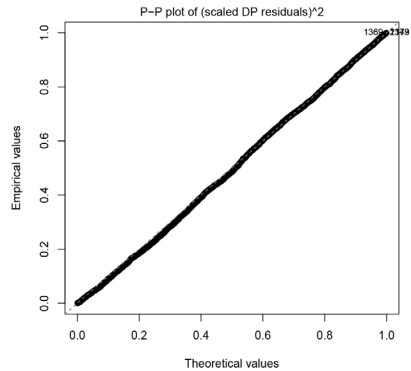


図15：株式時価総額の対数に非対称正規分布を当てはめたときの標準化直接母数残差の 2 乗の P-P プロット：新興国

これらの可視化の結果から、総じて非対称正規分布は新興国の企業に対する対数株式時価総額に当てはまっていることがわかる。

## 2. 非対称ティー分布の国タイプ別当てはめ

## 先進国

先進国 (G) の2018会計年度末の対数株式時価総額  $\log(\text{market\_cap}_G)$  に対数非対称ティー分布  $LST(\xi_G, \omega_G^2, \alpha_G, \nu_G)$  を当てはめる。

対数非対称ティー分布の直接母数  $(\xi_G, \omega_G, \alpha_G, \nu_G)$  の最尤推定は以下のよう  
に与えられる<sup>10)</sup>：

$$(\hat{\xi}_G, \hat{\omega}_G, \hat{\alpha}_G, \hat{\nu}_G) = (7.9, 1.67, 2.49, 650564.79)$$

これらの推定値より、統計モデルは、

$$\begin{aligned} & f_{ST}(\log(\text{market\_cap}_G) | \hat{\xi}_G, \hat{\omega}_G, \hat{\alpha}_G, \hat{\nu}_G) \\ &= \frac{2}{\hat{\omega}_G} f_t \left( \frac{\log(\text{market\_cap}_G) - \hat{\xi}_G}{\hat{\omega}_G} \middle| \hat{\nu}_G \right) \cdot \\ & F_t \left( \hat{\alpha}_G \frac{\log(\text{market\_cap}_G) - \hat{\xi}_G}{\hat{\omega}_G} \sqrt{\frac{\hat{\nu}_G + 1}{\left( \frac{\log(\text{market\_cap}_G) - \hat{\xi}_G}{\hat{\omega}_G} \right)^2 + \hat{\nu}_G}} \middle| \hat{\nu}_G + 1 \right) \\ &= 1.2 f_t \left( \frac{\log(\text{market\_cap}_G) - 7.9}{1.67} \middle| 650564.79 \right) \cdot \\ & F_t \left( 2.49 \frac{\log(\text{market\_cap}_G) - 7.9}{1.67} \sqrt{\frac{650565.79}{\left( \frac{\log(\text{market\_cap}_G) - 7.9}{\hat{\omega}} \right)^2 + 650564.79}} \middle| 650565.79 \right) \end{aligned} \quad (7)$$

となる。

ヒストグラムに統計モデル (7) を重ねて描いたものと、標準化直接母数  
残差の2乗のP-Pプロットを、それぞれ、図16, 17に与える。

10) 先進国の対数時価総額に非対称ティー分布を当てはめた場合の自由度の推定値  $\hat{\nu}_G = 650564.79$  は非常に大きなものとなっており、推定値の標準誤差が推定できない等の問題もあるため、参考程度と考えることが妥当であろう。

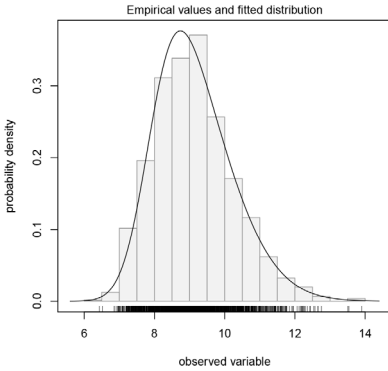


図16：株式時価総額の対数のヒストグラムと非対称ティー分布にもとづく統計モデル：先進国

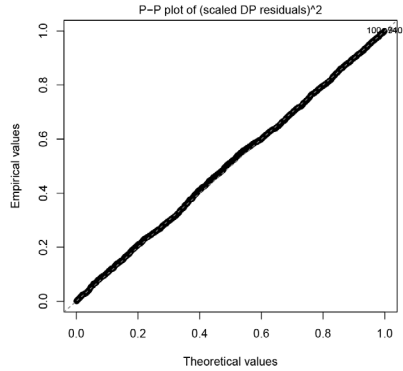


図17：株式時価総額の対数に非対称ティー分布を当てはめたときの標準化直接母数残差の2乗のP-Pプロット：先進国

これらの可視化の結果から、非対称ティー分布も非対称正規分布と同様に対数株式時価総額によく当てはまっていることがわかるが、自由度の推定値に関する問題 ( $\hat{\nu}_E$  の値) があることから、非対称正規分布の方が適切と考えられる。

新興国

新興国 (E) の2018会計年度末の対数株式時価総額  $\log(\text{market\_cap}_E)$  に対数非対称ティー分布  $LST(\xi_E, \omega_E^2, \alpha_E, \nu_E)$  を当てはめる。

対数非対称ティー分布の直接母数  $(\xi_E, \omega_E, \alpha_E, \nu_E)$  の最尤推定は以下のよう  
に与えられる：

$$(\hat{\xi}_E, \hat{\omega}_E, \hat{\alpha}_E, \hat{\nu}_E) = (7.34, 1.09, 1.01, 7.04)$$

これらの推定値より、統計モデルは、

$$f_{ST}(\log(\text{market\_cap}_E) | \hat{\xi}_E, \hat{\omega}_E, \hat{\alpha}_E, \hat{\nu}_E) = \frac{2}{\hat{\omega}_E} f_t\left(\frac{\log(\text{market\_cap}_E) - \hat{\xi}_E}{\hat{\omega}_E} \middle| \hat{\nu}_E\right)$$



$$\begin{aligned}
 & F_t \left( \hat{\alpha}_E \frac{\log(\text{market\_cap}_E) - \hat{\xi}_E}{\hat{\omega}_E} \sqrt{\frac{\hat{\nu}_E + 1}{\left( \frac{\log(\text{market\_cap}_E) - \hat{\xi}_E}{\hat{\omega}_E} \right)^2 + \hat{\nu}_E}} \right) \\
 &= 1.83 f_t \left( \frac{\log(\text{market\_cap}_E) - 7.34}{1.09} \middle| 7.04 \right) \cdot \\
 & F_t \left( 1.01 \frac{\log(\text{market\_cap}_E) - 7.34}{1.09} \cdot \sqrt{\frac{8.04}{\left( \frac{\log(\text{market\_cap}_E) - 7.34}{\hat{\omega}} \right)^2 + 7.04}} \middle| 8.04 \right)
 \end{aligned} \tag{8}$$

となる。

ヒストグラムに統計モデル (8) を重ねて描いたものと、標準化直接母数残差の2乗の P-P プロットを、それぞれ、図18, 19に与える。

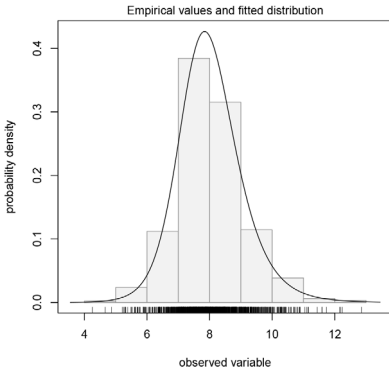


図18：株式時価総額の対数のヒストグラムと非対称ティー分布にもとづく統計モデル：新興国

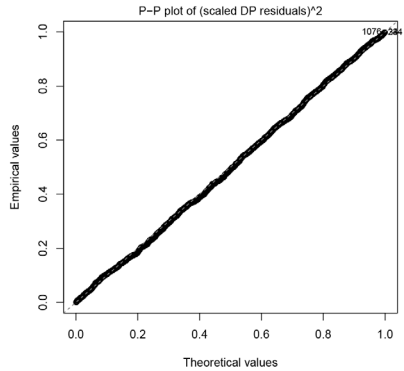


図19：株式時価総額の対数に非対称ティー分布を当てはめたときの標準化直接母数残差の2乗の P-P プロット：新興国

これらの可視化の結果から、非対称ティー分布も非対称正規分布と同様に対数株式時価総額によく当てはまっていることがわかる。

3. 正規逆ガウス分布の国タイプ別当てはめ  
先進国

先進国 (G) の2018会計年度末の対数株式時価総額  $\log(\text{market\_cap}_c)$  に正規逆ガウス分布  $\text{NIG}(\alpha_c, \beta_c, \delta_c, \mu_c)$  を当てはめる。母数  $(\alpha_c, \beta_c, \delta_c, \mu_c)$  の最尤推定値は以下のように与えられる：

$$(\hat{\alpha}_c, \hat{\beta}_c, \hat{\delta}_c, \hat{\mu}_c) = (64.06, 61.98, 1.3, 4.16) \tag{9}$$

これらの推定値より、統計モデルは、

$$\begin{aligned} & f_{\text{NIG}}(\log(\text{market\_cap}_c) | \hat{\alpha}_c, \hat{\beta}_c, \hat{\delta}_c, \hat{\mu}_c) \\ &= \frac{\hat{\alpha}_c \hat{\delta}_c}{\pi} \exp\{\hat{\delta}_c \sqrt{\hat{\alpha}_c^2 - \hat{\beta}_c^2} + \hat{\beta}_c (\log(\text{market\_cap}_c) - \hat{\mu}_c)\} \cdot \\ & \quad \frac{K_1(\hat{\alpha}_c \sqrt{(\log(\text{market\_cap}_c) - \hat{\mu}_c)^2 + \hat{\delta}_c^2})}{\sqrt{(\log(\text{market\_cap}_c) - \hat{\mu}_c)^2 + \hat{\delta}_c^2}} \\ &= \frac{83.2}{\pi} \exp\{21 + 61.98(\log(\text{market\_cap}_c) - 4.16)\} \\ & \quad \frac{K_1(64.06 \sqrt{(\log(\text{market\_cap}_c) - 4.16)^2 + 1.69})}{\sqrt{(\log(\text{market\_cap}_c) - 4.16)^2 + 1.69}} \end{aligned} \tag{10}$$

ヒストグラムに統計モデル (10) を重ねて描いたものと、対数株式時価総額の P-P プロットを、それぞれ、図20, 21に与える。

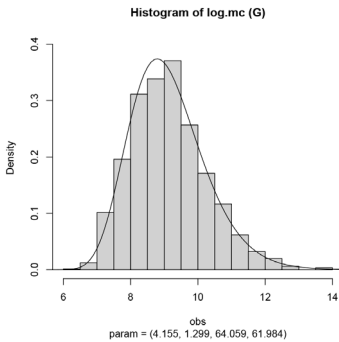


図20：対数株式時価総額のヒストグラムと正規逆ガウス分布にもとづく統計モデル：先進国

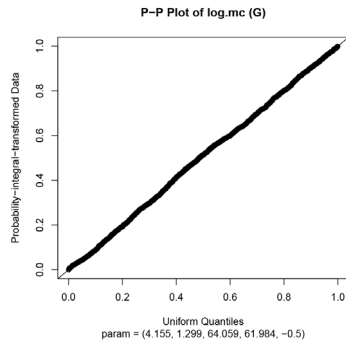


図21：対数株式時価総額に正規逆ガウス分布を当てはめたときの P-P プロット：先進国

これらの可視化の結果からも、非対称正規分布、非対称ティー分布の場合と同様に対数株式時価総額によく当てはまっていることがわかり、可視化の結果だけでは優劣は付けがたいことがわかる。

## 新興国

新興国 (E) の2018会計年度末の対数株式時価総額  $\log(\text{market\_cap}_E)$  に正規逆ガウス分布  $\text{NIG}(\alpha_E, \beta_E, \delta_E, \mu_E)$  を当てはめる。母数  $(\alpha_E, \beta_E, \delta_E, \mu_E)$  の最尤推定値は以下のように与えられる：

$$(\hat{\alpha}_E, \hat{\beta}_E, \hat{\delta}_E, \hat{\mu}_E) = (1.45, 0.4, 1.51, 7.6) \quad (11)$$

これらの推定値より、統計モデルは、

$$\begin{aligned} & f_{\text{NIG}}(\log(\text{market\_cap}_E) | \hat{\alpha}_E, \hat{\beta}_E, \hat{\delta}_E, \hat{\mu}_E) \\ &= \frac{\hat{\alpha}_E \hat{\delta}_E}{\pi} \exp \left\{ \hat{\delta}_E \sqrt{\hat{\alpha}_E^2 - \hat{\beta}_E^2} + \hat{\beta}_E (\log(\text{market\_cap}_E) - \hat{\mu}_E) \right\} \cdot \\ & \quad \frac{K_1 \left( \hat{\alpha}_E \sqrt{(\log(\text{market\_cap}_E) - \hat{\mu}_E)^2 + \hat{\delta}_E^2} \right)}{\sqrt{(\log(\text{market\_cap}_E) - \hat{\mu}_E)^2 + \hat{\delta}_E^2}} \\ &= \frac{2.18}{\pi} \exp \{ 2.09 + 0.4 (\log(\text{market\_cap}_E) - 7.6) \} \cdot \\ & \quad \frac{K_1 \left( 1.45 \sqrt{(\log(\text{market\_cap}_E) - 7.6)^2 + 2.27} \right)}{\sqrt{(\log(\text{market\_cap}_E) - 7.6)^2 + 2.27}} \quad (12) \end{aligned}$$

ヒストグラムに統計モデル (12) を重ねて描いたものと、対数株式時価総額の P-P プロットを、それぞれ、図22, 23に与える。

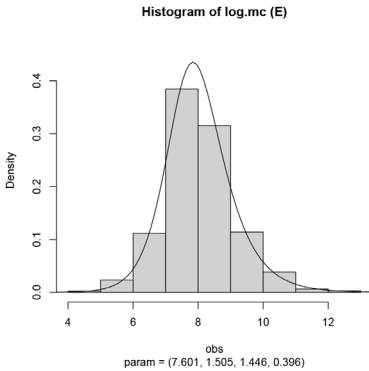


図22：対数株式時価総額のヒストグラムと正規逆ガウス分布にもとづく統計モデル：新興国

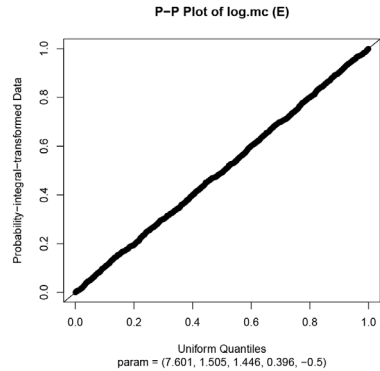


図23：対数株式時価総額に正規逆ガウス分布を当てはめたときのP-Pプロット：新興国

これらの可視化の結果から、非対称正規分布、非対称ティー分布の場合よりも、対数株式時価総額によく当てはまっているようであるが、可視化の結果だけでは判断が難しいため、次節で情報量規準を用いて比較・検討する。

#### 4. 対数株式時価総額の国タイプ別分布に関するモデル選択

ここでは、情報量規準であるAICとBICを利用して、対数株式時価総額の分布に対して統計モデルの当てはまりの比較を行う。

まず、先進国の企業の対数株式時価総額については、表2から、AICでは正規逆ガウス分布が最も良く、BICでは非対称正規分布が最も良いことがわかった<sup>11)</sup>。

11) 先進国の企業の対数株式時価総額については非対称ティー分布は最も当てはまりは悪いが、この結果は母数の最尤推定値の問題とも整合する。

表 2 : AIC, BIC 表 : 先進国の企業の対数株式時価総額への分布当てはめに関する情報量規準の比較

	dim	AIC	BIC
SN	3	4864.62	4880.78
ST	4	4866.62	4888.16
NIG	4	4861.89	4883.43

次に、新興国の企業の対数株式時価総額については、表 3 から、AIC, BIC とも正規逆ガウス (NIG) 分布が最も良いことがわかった。

表 3 : AIC, BIC 表 : 新興国の企業の対数株式時価総額への分布当てはめに関する情報量規準の比較

	dim	AIC	BIC
SN	3	3756.16	3771.59
ST	4	3729.94	3750.52
NIG	4	3727.19	3747.77

以上の考察から、企業を国タイプ別に分けた場合の対数株式時価総額の分布としては、(特に AIC を利用した場合は) 正規逆ガウス (NIG) 分布の当てはまりが良いという結果を得た。

## 5. 再生性のチェック

正規逆ガウス分布は再生性を持つ (付録 C における命題 6 を参照) ことから、この分布に対数株式時価総額が従っていることが、本稿の冒頭で指摘したように、国タイプ別に分類した企業の対数株式時価総額を合併したものが再生性を持つということの理論的根拠となることが期待される。ここでは、このことを以下の命題のように定式化する：

**命題 1 (対数株式時価総額に関する再生性)**  $\log(\text{market\_cap}_e)$  と  $\log(\text{market\_cap}_e)$  をそれぞれ、先進国と新興国の対数株式時価総額とすると、条件：

$$(C1) \log(\text{market\_cap}_G) \sim \text{NIG}(\alpha_G, \beta_G, \delta_G, \mu_G)$$

$$(C2) \log(\text{market\_cap}_E) \sim \text{NIG}(\alpha_E, \beta_E, \delta_E, \mu_E)$$

$$(C3) \log(\text{market\_cap}_G) \perp \log(\text{market\_cap}_E) \text{ (独立性)}$$

$$(C4) \alpha = \alpha_G = \alpha_E$$

$$(C5) \beta = \beta_G = \beta_E$$

のもとで、

$$\log(\text{market\_cap}_G) + \log(\text{market\_cap}_E) \sim \text{NIG}(\alpha, \beta, \delta_G + \delta_E, \mu_G + \mu_E) \quad (13)$$

が成り立つ。

命題 (1) における条件のうち、これまでの考察から、条件 (C1), (C2), (C3) はある程度妥当といえるかもしれないが、最尤推定値 (3), (9), (11) より、

$$\hat{\alpha}_G = 64.06, \hat{\alpha}_E = 1.45, \hat{\alpha} = 2.38,$$

$$\hat{\beta}_G = 61.98, \hat{\beta}_E = 0.4, \hat{\beta} = 0.92$$

となり、この結果から、条件 (C4), (C5) がみたされているとは考えにくい<sup>12)</sup>。よって、残念ながら、再生性が成り立つとはいえない。

この問題に対して、次節では有限混合分布を利用した当てはめを行う。

## V 有限混合分布の当てはめ

ここでは、対数株式時価総額  $\log(\text{market\_cap})$  へ有限混合分布  $\text{Mix}(\theta_G, \theta_E, \varpi_G, \varpi_E)$  を当てはめることを考える。なお、付録 D に有限混合分布の導出と、最尤推定、情報量規準に関する事項をまとめているので参照されたい。

有限混合分布の確率密度関数を以下で定義する：

$$\begin{aligned} f_{\text{Mix}}(\log(\text{market\_cap}) | \theta_G, \theta_E, \varpi_G, \varpi_E) : \\ = \varpi_G f_G(\log(\text{market\_cap}) | \theta_G) + \varpi_E f_E(\log(\text{market\_cap}) | \theta_E) \end{aligned}$$

ここで、 $f_G(\log(\text{market\_cap}) | \theta_G)$ ,  $f_E(\log(\text{market\_cap}) | \theta_E)$  は、それぞれ、先進国と新興国の対数株式時価総額  $\log(\text{market\_cap})$  の分布に対する確

12) より精確には、母数の同等性の検定を考える必要があるかもしれないが、数値的に母数は明らかに異なっていると思われる。

率密度関数であり、

$$(\varpi_G, \varpi_E) = (\varpi_G, 1 - \varpi_G), \quad \varpi_G + \varpi_E = 1, \quad \varpi_G, \varpi_E \in [0, 1]$$

は混合比率である。

混合比率  $(\varpi_G, \varpi_E)$  の最尤推定値は、付録 D の定義式 (51) と、表 11 の 2018 会計年度の企業数より<sup>13)</sup>、

$$\begin{aligned} & (\text{総企業数, 先進国の企業数, 新興国の企業数}) \\ & = (n, n_G, n_E) = (2882, 1613, 1269) \end{aligned}$$

となるので、

$$(\hat{\varpi}_G, \hat{\varpi}_E) := \left( \frac{n_G}{n}, \frac{n_E}{n} \right) = \left( \frac{1613}{2882}, \frac{1269}{2882} \right) = (0.56, 0.44)$$

で与えられる。よって、母数ベクトル  $\theta_G, \theta_E$  の最尤推定値ベクトルを、それぞれ、 $\hat{\theta}_G, \hat{\theta}_E$  とすると、統計モデル

は以下で与えられる：

$$\begin{aligned} & f_{\text{mix}}(\log(\text{market\_cap}) | \hat{\theta}_G, \hat{\theta}_E, \hat{\varpi}_G, \hat{\varpi}_E): \\ & = \hat{\varpi}_G f_G(\log(\text{market\_cap}) | \hat{\theta}_G) + \hat{\varpi}_E f_E(\log(\text{market\_cap}) | \hat{\theta}_E) \\ & = 0.56 f_G(\log(\text{market\_cap}) | \hat{\theta}_G) + 0.44 f_E(\log(\text{market\_cap}) | \hat{\theta}_E) \quad (14) \end{aligned}$$

先進国 (G) と新興国 (E) の統計モデル  $f_G(\log(\text{market\_cap}) | \hat{\theta}_G), f_E(\log(\text{market\_cap}) | \hat{\theta}_E)$  の選択としては、情報量規準による分布の選択結果 (表 2, 表 3) より、先進国 (G) に対しては、正規逆ガウス (NIG) 分布 (統計モデル (10)), 非対称正規 (SN) 分布 (統計モデル (5)) が候補となろう。一方、新興国 (E) に対しては、正規逆ガウス (NIG) 分布 (統計モデル (12)), 非対称ティール (ST) 分布 (統計モデル (8)) が候補となろう。

これらの候補の分布を対数株式時価総額に当てはめたときの AIC, BIC の値を表 4 に与える。

13) 今回利用しているデータは、財務データのみであることから、企業数は、ESG 指標に存在する欠測値の影響は受けないことに留意する必要がある。

表4：AIC, BIC表：対数株式時価総額への有限混合分布の当てはめに関する情報量規準の比較

	G	E	df	AIC	BIC
1	NIG	NIG	9	12545.22	12598.91
2	NIG	ST	9	12547.97	12601.66
3	SN	NIG	8	12547.95	12595.68
4	SN	ST	8	12550.70	12598.43

表4における情報量規準の値の比較から、AICでは、先進国と新興国の企業の株式時価総額ともに、正規逆ガウス(NIG)分布(統計モデル(10)、(12))を混合した分布が適していることがわかる。一方、BICでは、先進国に対しては、非対称正規(SN)(統計モデル(5))と新興国に対しては正規逆ガウス(NIG)分布(統計モデル(12))を混合した分布が適しているという結果を得た。

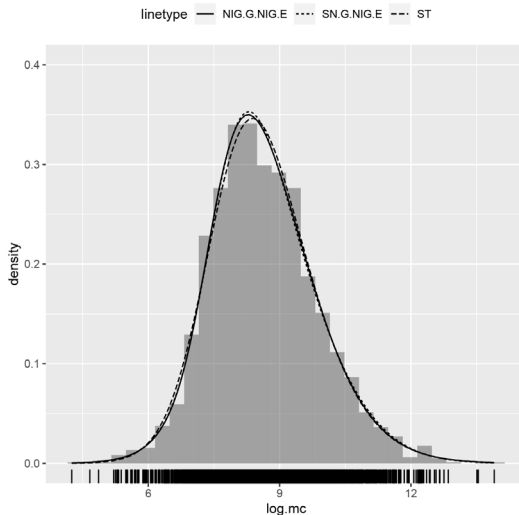


図24：対数株式時価総額のヒストグラムと統計モデル：実線が先進国と新興国の場合ともに正規逆ガウス(NIG)分布を当てはめそれらを混合した場合であり、点線が先進国に非対称正規(SN)分布、新興国に正規逆ガウス(NIG)分布を当てはめそれらを混合した場合である。また、破線はデータをタイプに分けず非対称ティータ(ST)分布を全体に当てはめた場合である。



以上の結果から、先進国については、正規逆ガウス (NIG) 分布または非対称正規 (SN) 分布のどちらを選択するかは判断が分かれるが、新興国については正規逆ガウス (NIG) 分布がよいと判断ができればよい。なお、データをタイプに分けず全企業の株式時価総額へ単一の分布を当てはめる場合については、表1から、非対称ティー (ST) 分布の当てはまりが良かったことを踏まえて、これらの分布の候補の統計モデルを可視化した結果を図24に与える。

## VI 経年変化を考慮した考察

これまでの議論では、2018会計年度に限定して、対数株式時価総額への各種の分布の当てはめについて考察してきた。ここでは経年変化を考慮し、考察を行う。

### 1. 経年変化を考慮した対数株式時価総額への分布当てはめに関するモデル選択

単年度毎に非対称正規分布、非対称ティー分布、正規逆ガウス分布をそれぞれ当てはめたときの情報量規準 AIC, BIC の経年変化をみる。

図25, 26からは、AIC 値, BIC 値について3種類の分布間に顕著な差異は見られないが、企業数で除した値 (1社あたりの AIC 値, BIC 値) では、非対称正規分布に比べて、非対称ティー分布と正規逆ガウス分布の当てはまりが良いことがわかる。さらに、表5から、図25, 26からは判断が難しかったが、全ての年度において非対称ティー分布の AIC 値 (AIC\_ST) が、その他の場合に比べて (若干であるが) 小さいことがわかる。また、BIC 値 (BIC\_ST) も幾つかの例外 (会計年度: 2015, 2017) があるが、非対称ティー分布の当てはまりが良いことがわかる。この結果は、2018会計年度における結果 (表1) がその他の年度においても成り立つことを表している。

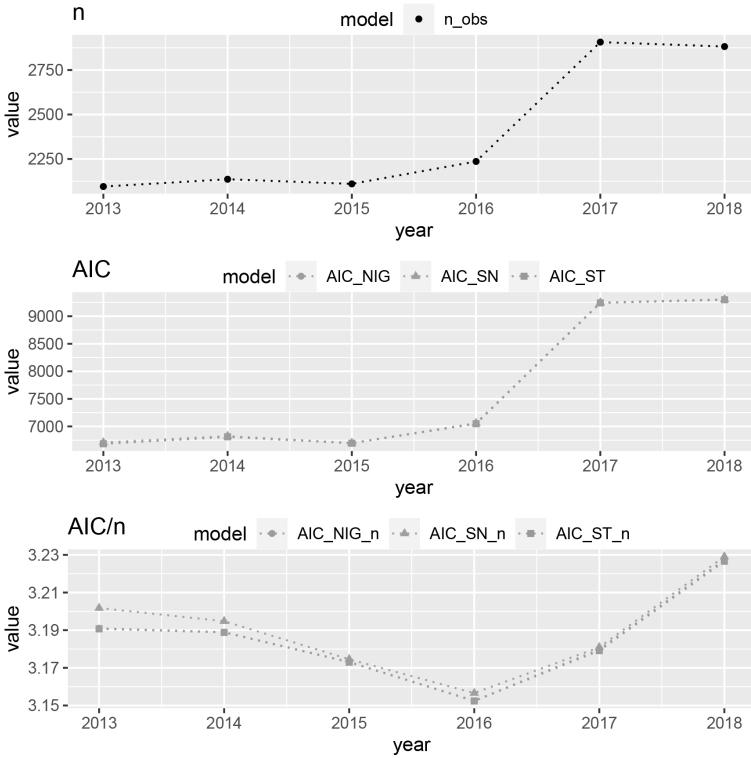


図25：2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と非対称正規分布，非対称テーパー分布，正規逆ガウスの AIC 値  $AIC$ （中段），1社あたりの AIC 値  $AIC/n$ （下段）の経年変化のプロット：SN（非対称正規分布），ST（非対称テーパー分布），NIG（正規逆ガウス分布）

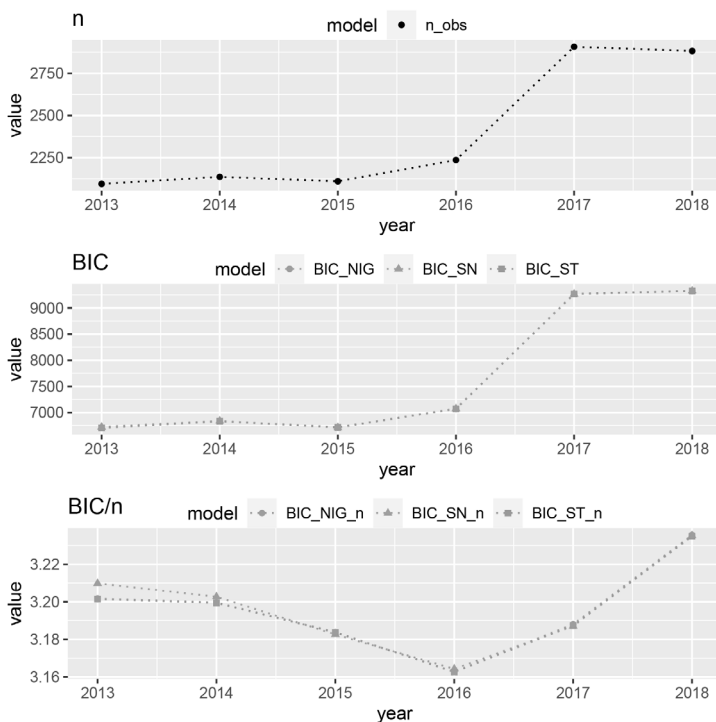


図26：2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と非対称正規分布，非対称ティータ分布，正規逆ガウスの  $BIC$  値  $BIC$ （中段），1社あたりの  $BIC$  値  $BIC/n$ （下段）の経年変化のプロット：SN（非対称正規分布），ST（非対称ティータ分布），NIG（正規逆ガウス分布）

表 5：AIC, BIC 表：対数株式時価総額への分布当てはめに関する情報量規準の経年変化

year	n_obs	AIC_SN	AIC_ST	AIC_NIG	BIC_SN	BIC_ST	BIC_NIG
2013	2095	6707.66	6684.49	6685.17	6724.61	6707.08	6707.76
2014	2136	6824.03	6811.33	6811.65	6841.03	6834.00	6834.32
2015	2110	6698.35	6694.98	6695.36	6715.31	6717.60	6717.97
2016	2236	7058.14	7048.69	7049.03	7075.27	7071.54	7071.88
2017	2907	9246.97	9241.81	9243.64	9264.90	9265.70	9267.54
2018	2882	9306.21	9299.10	9301.20	9324.11	9322.97	9325.06

## 2. 国タイプ別の経年変化を考慮した対数株式時価総額への分布当てはめに関するモデル選択

### 先進国

先進国の企業に関して単年度毎に非対称正規分布、非対称ティー分布、正規逆ガウス分布をそれぞれ当てはめたときの情報量規準 AIC, BIC の経年変化をみる。図27, 28からは、AIC 値, BIC 値について3種類の分布間に顕著な差異は見られないが、企業数で除した値（1社あたりの AIC 値, BIC 値）からは、AIC に関しては、総じて、非対称正規分布と非対称ティー分布に比べて正規逆ガウス分布の当てはまりが良いことがわかる。一方、BIC に関しては、非対称ティー分布と正規逆ガウス分布に比べて非対称正規分布の当てはまりが良いことがわかる。表6からも、これらの結果が数値的に確認できる。以上の結果は、2018会計年度における結果（表2）がその他の年度においても成り立つことを表している。

表6：AIC, BIC 表：先進国における企業の対数株式時価総額への分布当てはめに関する情報量規準の経年変化

year	n_obs	AIC_SN	AIC_ST	AIC_NIG	BIC_SN	BIC_ST	BIC_NIG
2013	1466	4453.80	4455.80	4452.32	4469.67	4476.97	4473.48
2014	1489	4583.39	4584.30	4585.75	4599.31	4605.53	4606.97
2015	1463	4436.22	4438.22	4435.18	4452.08	4459.37	4456.34
2016	1538	4610.26	4612.26	4607.98	4626.28	4633.62	4629.34
2017	1625	4859.26	4861.26	4858.26	4875.44	4882.83	4879.83
2018	1613	4864.62	4866.62	4861.89	4880.78	4888.16	4883.43

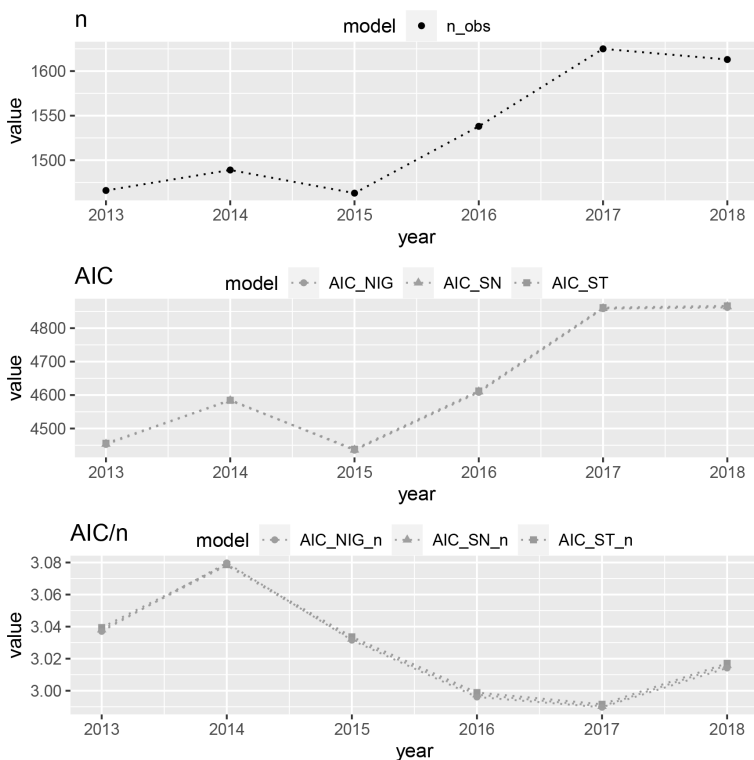


図27：先進国における2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と非対称正規分布，非対称ティータ分布，正規逆ガウスのAIC値  $AIC$ （中段），1社あたりのAIC値  $AIC/n$ （下段）の経年変化のプロット：SN（非対称正規分布），ST（非対称ティータ分布），NIG（正規逆ガウス分布）

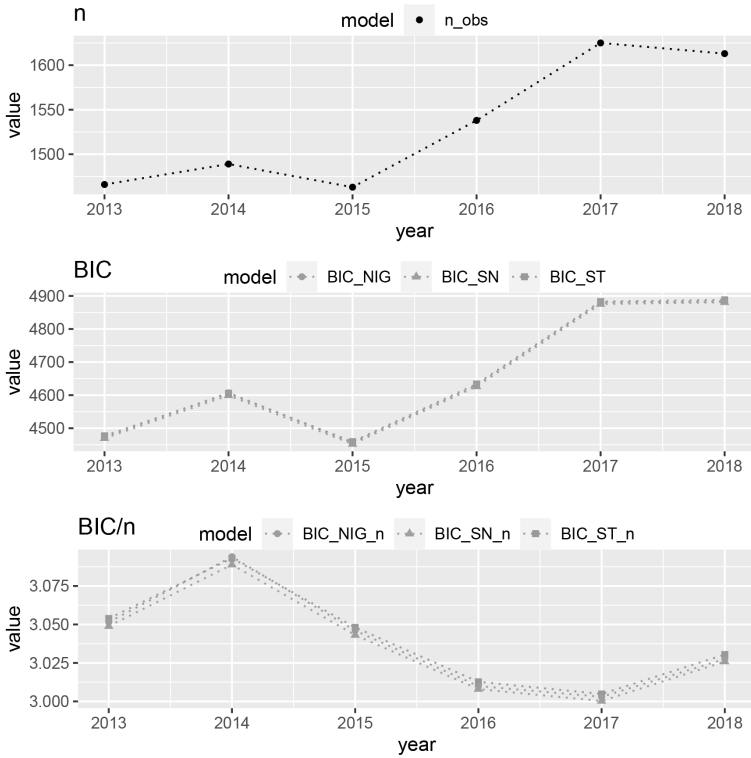


図28：先進国における2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と非対称正規分布，非対称ティータ分布，正規逆ガウスの  $BIC$  値  $BIC$ （中段），1社あたりの  $BIC$  値  $BIC/n$ （下段）の経年変化のプロット：SN（非対称正規分布），ST（非対称ティータ分布），NIG（正規逆ガウス分布）

## 新興国

新興国の企業に関して単年度毎に非対称正規分布，非対称ティー分布，正規逆ガウス分布をそれぞれ当てはめたときの情報量規準 AIC, BIC の経年変化をみる．図29, 30からは，AIC 値，BIC 値について3種類の分布間に顕著な際は見られないが，企業数で除した値（1社あたりの AIC 値，BIC 値）からは，AIC, BIC とともに，非対称正規分布と非対称ティー分布に比べて正規逆ガウス分布の当てはまりが良いことがわかる．表7からも，これらの結果が数値的に確認できる．以上の結果は，2018会計年度における結果（表3）がその他の年度においても成り立つことを表している．

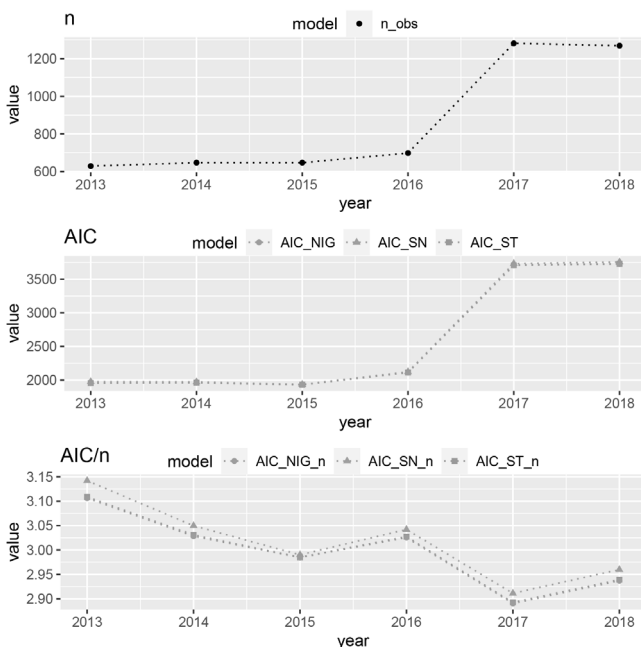


図29：新興国における2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と非対称正規分布，非対称ティー分布，正規逆ガウスの AIC 値  $AIC$ （中段），1社あたりの AIC 値  $AIC/n$ （下段）の経年変化のプロット：SN（非対称正規分布），ST（非対称ティー分布），NIG（正規逆ガウス分布）

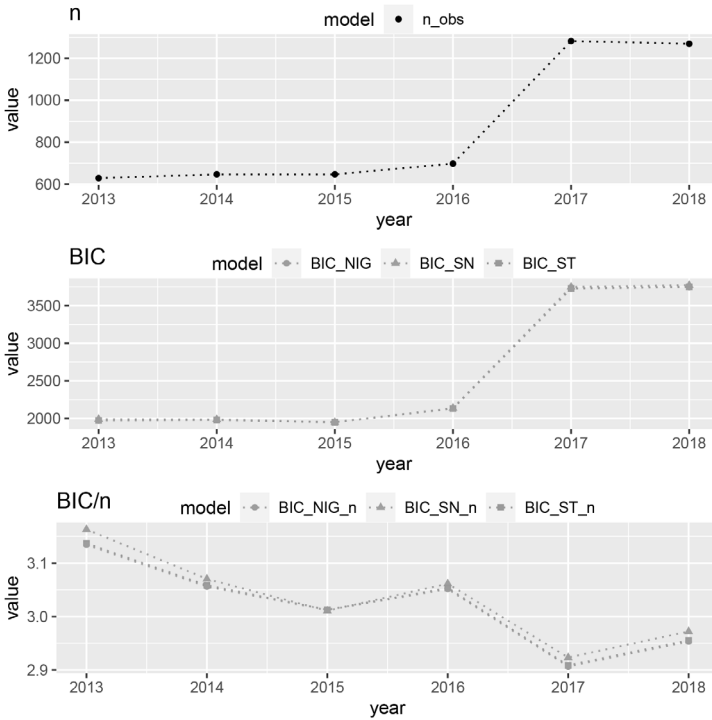


図30：新興国における2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と非対称正規分布，非対称ティータ分布，正規逆ガウスの  $BIC$  値  $BIC$ （中段），1社あたりの  $BIC$  値  $BIC/n$ （下段）の経年変化のプロット：SN（非対称正規分布），ST（非対称ティータ分布），NIG（正規逆ガウス分布）

表7：AIC, BIC表：新興国における企業の対数株式時価総額への分布当てはめに関する情報量規準の経年変化

year	n_obs	AIC_SN	AIC_ST	AIC_NIG	BIC_SN	BIC_ST	BIC_NIG
2013	629	1976.35	1955.44	1953.81	1989.68	1973.21	1971.59
2014	647	1973.18	1961.00	1959.34	1986.60	1978.89	1977.23
2015	647	1934.60	1931.37	1931.12	1948.02	1949.26	1949.01
2016	698	2123.30	2113.13	2111.84	2136.94	2131.32	2130.04
2017	1282	3732.31	3708.33	3705.72	3747.78	3728.95	3726.34
2018	1269	3756.16	3729.94	3727.19	3771.59	3750.52	3747.77



### 3. 経年変化を考慮した対数株式時価総額への有限混合分布の当てはめに関するモデル選択

単年度毎に有限混合分布を当てはめたときの情報量規準 AIC, BIC の経年変化をみる。ここでは、前小節の結果から、先進国の企業に対する対数株式時価総額には正規逆ガウス分布と非対称正規分布を、新興国には正規逆ガウス分布を選択した場合の有限混合分布を考える。

図31における上段のプロットから、企業数が近年（2017年、2018年）増加していることがわかるが、これは、新興国の企業数（ $n_E$ ）が増加していることによることがわかる。このことは、表8から具体的に数値で確認することができ、2013年から2016年までは先進国の企業数が1,500社程度に対して、新興国は600社程度であり、（混合）比率の推定値では、 $(\hat{\omega}_G, \hat{\omega}_E) \simeq (0.7, 0.3)$  だったものが、2017年、2018年では  $(\hat{\omega}_G, \hat{\omega}_E) \simeq (0.56, 0.44)$  とほぼ対等まで企業数が変化していることが確認できる。

また、図31の中段と下段の AIC 値、1社あたりの AIC 値について2種類の有限混合分布間に顕著な差異は見られないが、表9から、先進国に正規逆ガウス分布、新興国に正規逆ガウス分布を当てはめた場合の有限混合分布 (NIG.G\_NIG.E) の AIC 値が若干小さいことがわかり、この基準のものでは、先進国・新興国とも正規逆ガウス分布を有限混合した分布が良いことがわかる。

一方、図32からは、中段の BIC 値について2種類の有限混合分布間に顕著な差異は見られないが、1社あたりの BIC 値から、先進国に非対称正規分布、新興国に正規逆ガウス分布) を当てはめた場合の有限混合分布 (SN.G\_NIG.E) の BIG 値が若干小さいことがわかり、このことは表9からも確認できる。よって、BIC 値からは先進国に非対称正規分布、新興国に正規逆ガウス分布を当てはめた場合の有限混合分布が良いことがわかる。

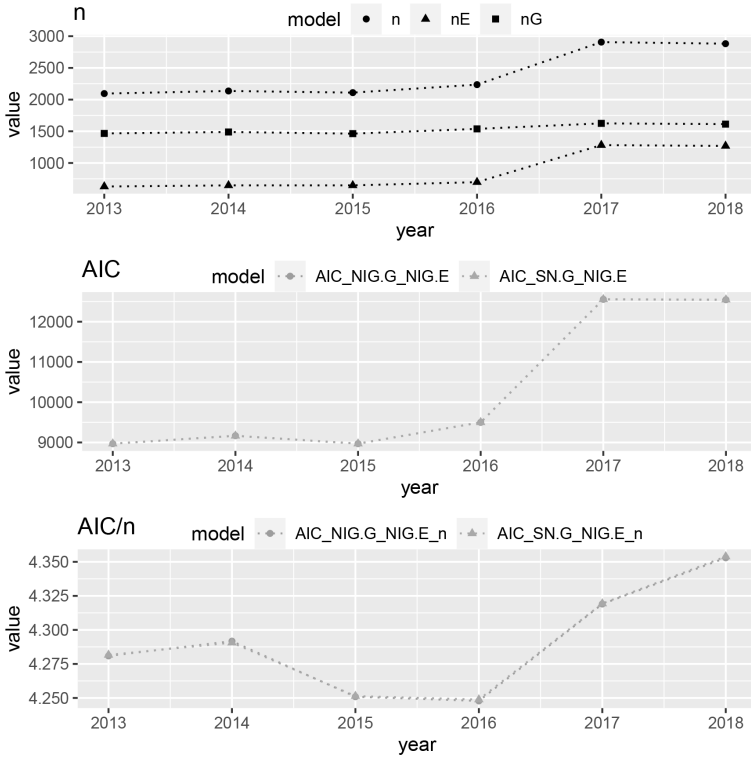


図31：2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と有限混合分布の AIC 値  $AIC$ （中段），1社あたりの AIC 値  $AIC/n$ （下段）の経年変化のプロット：NIG.G\_NIG.E（有限混合分布（先進国：正規逆ガウス分布，新興国：正規逆ガウス分布）を当てはめた場合），SN.G\_NIG.E（有限混合分布（先進国：非対称正規分布，新興国：正規逆ガウス分布）を当てはめた場合）

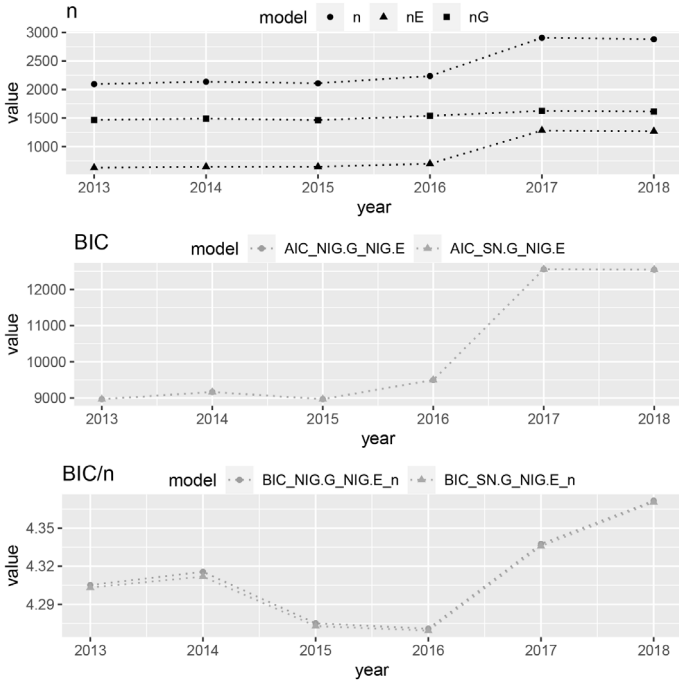


図32：2013～2018年（会計年度）のそれぞれの時点における企業数  $n$ （上段）と有限混合分布の BIC 値  $BIC$ （中段），1社あたりの BIC 値  $BIC/n$ （下段）の経年変化のプロット：NIG.G\_NIG.E（有限混合分布（先進国：正規逆ガウス分布，新興国：正規逆ガウス分布）を当てはめた場合），SN.G\_NIG.E（有限混合分布（先進国：非対称正規分布，新興国：正規逆ガウス分布）を当てはめた場合）

表 8：企業数と混合比率の推定値の経年変化：企業数 ( $n$ )，先進国企業数 ( $nG$ )，新興国企業数 ( $nE$ )，先進国の混合比率推定値 ( $varpiG$ )，新興国の混合比率推定値 ( $varpiE$ )

year	n	nG	nE	varpiG	varpiE
2013	2095	1466	629	0.70	0.30
2014	2136	1489	647	0.70	0.30
2015	2110	1463	647	0.69	0.31
2016	2236	1538	698	0.69	0.31
2017	2907	1625	1282	0.56	0.44
2018	2882	1613	1269	0.56	0.44

表9：AIC, BIC表：対数株式時価総額への有限混合分布の当てはめに関する情報量規準の経年変化

year	AIC_NIG.G_NIG.E	AIC_SN.G_NIG.E	BIC_NIG.G_NIG.E	BIC_SN.G_NIG.E
2013	8968.50	8969.98	9019.32	9015.16
2014	9167.12	9164.77	9218.12	9210.10
2015	8969.43	8970.46	9020.32	9015.70
2016	9498.14	9500.42	9549.55	9546.12
2017	12555.37	12556.37	12609.14	12604.17
2018	12545.22	12547.95	12598.91	12595.68

## VII おわりに

本稿では、世界の上場企業に対する対数株式時価総額に対して非対称正規分布族と一般化双曲型分布族を考慮した分布のモデリングについて考察した。また、国のタイプ（先進国，新興国）別に分類された企業の対数株式時価総額にも同様の分布の当てはめ，さらに，有限混合分布も当てはめた。これらの考察から以下のような結果がえられた：

- (R1) 単一分布を当てはめる場合は，経年変化を考慮しても，非対称ティータ分布の当てはまりが良い
- (R2) 企業を先進国と新興国に分類した場合：
- 先進国については，AIC では正規逆ガウス分布の当てはまりが良く，BIC では非対称正規分布の当てはまりが良い
  - 新興国については，AIC, BIC とも正規逆ガウス分布の当てはまりが良い
  - 正規逆ガウス分布を仮定した場合でも，国タイプ別に分類した企業の対数株式時価総額を合併したものが再生性を持つことを肯定することはできない
- (R3) 有限混合分布を当てはめる場合：
- 企業を先進国と新興国に分類した場合の結果が受け継がれる

- AIC, BIC の値は単一の分布を当てはめる場合よりも大きくなる

以上の結果をどのように判断するかは議論が分かれるところであろうが、いわゆる、「ケチの原理」からは、モデルがシンプルな非対称ティー分布が適切であり、本稿で考察したより複雑なモデルを当てはめるよりも、AIC, BIC の値は小さい。一方、データを同質のグループに「分割」し、適切な分布を当てはめ（「適用」）、結果を「統合」という観点、いわゆる、分割・適用・統合戦略（split-apply-combine strategy）（cf. Wickham (2011)）からは、企業を先進国と新興国に分類したものに適切な分布を当てはめ、結果を有限混合したものが指示される。このときは、AIC を利用する場合は、先進国、新興国ともに正規逆ガウス分布を混合したものが良く、また、BIC を利用する場合は、非対称正規分布（先進国）と、正規逆ガウス分布（新興国）を混合したものが良いが、いずれの場合も AIC, BIC の値は非対称ティー分布をデータ全体に当てはめた場合よりも大きくなることに留意する必要がある。なお、AIC と BIC の性質の違いについては、赤池（1996）を参照されたい。

なお、本稿では、地道、阪（2021-b）において指摘された問題を検討するという立場から、財務データと ESG レーティングデータを結合したデータを利用した。今後は、財務データのみを利用した考察や、有限混合分布を利用した回帰モデリングを考察する予定である。

（筆者は関西学院大学商学部教授）

#### 参考文献

- [1] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, *Proceedings of the 2nd International Symposium on Information Theory*, Petrov, B. N., and Csaki, F. (eds.), Akademiai Kiado, Budapest: pp. 267-281.
- [2] 赤池弘次 (1996) 「AIC と MDL と BIC」, オペレーションズリサーチ, 第41巻, 第7号, pp. 375-378.
- [3] Azzalini, A. (1985) A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, Vol. 12, No. 2, pp. 171-178.

- [4] Azzalini, A. with the collaboration of A. Capitanio (2014) *The Skew-Normal and Related Families*, Cambridge University Press, Institute of Mathematical Statistics Monographs.
- [5] Barndorff-Nielsen, O. E. (1977) Exponential decreasing distributions for the logarithm of particle size, *Proceedings of The Royal Society A: Mathematics, Physical and Engineering Sciences*, Vol 353, pp. 401-419.
- [6] Barndorff-Nielsen, O. E. (1988) *Parametric Statistical Models and Likelihood*, Lecture Notes in Statistics, 50, Springer-Verlag.
- [7] Barndorff-Nielsen, O. E., J. Kent, and M. Sørensen (1982) Normal variance-mean mixtures and  $z$  distributions, *International Statistical Review*, Vol. 50, pp. 145-159.
- [8] Bishop, C. M. (2006) *Pattern Recognition and Machine Learning*, Springer-Verlag.  
(邦訳) C. M. ビショップ (2006, 2007) 元田 浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田 昇 (共訳) パターン認識と機械学習 (上, 下) - ベイズ理論による統計の予測 -, シュプリンガー・ジャパン株式会社.
- [9] Cohen, A. C. (1967) Estimation in mixtures of two normal distributions, *Technometrics*, Vol. 9, No. 1, pp. 15-28.
- [10] Day, N. E. (1969) Estimating the components of a mixture of normal distributions, *Biometrika*, Vol. 56, pp. 463-474.
- [11] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39 No. 1 pp. 1-38.
- [12] FTSE Russell (2019) *ESG Data Model 6th Research Cycle, Methodology (2019/20)*, <https://qsd.ftserussell.com/Docs/ESG/FTSE%20Russell%20ESG%20Data%20Model%20Methodology%20-%20April-2019-March-2020-Rev2.pdf>
- [13] Hastie, T., R. Tibshirani, J. Friedman (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*, Springer.  
(邦訳) 杉山 将・井手 剛・神嶋敏弘・栗田多喜夫・前田英作 (監訳) (2014) 『統計的学習の基礎: データマイニング・推論・予測』, 共立出版.
- [14] 稲垣宣生 (2003) 『数理統計学-改訂版-』, 裳華房.
- [15] James, G., D. Witten, T. Hastie, and R. Tibshirani (2021) *An Introduction to Statistical Learning with Applications in R, Second Edition*, Springer.  
(初版邦訳) 落海浩, 首藤信通 (共訳) (2018) 『Rによる統計的学習入門』, 朝倉書店.
- [16] 地道正行 (2017-a) 『Rによる対数非対称正規線形モデルによる財務データの統計モデリング』, 商学論究, 第64巻, 第5号, pp. 159-185, 関西学院大学商学研究会.
- [17] 地道正行 (2017-b) 『Rを利用した非対称分布族にもとづく財務データの統計モデリング』, 経済学論究, 第71巻, 第2号, pp. 141-174, 関西学院大学経済学部研究会.
- [18] 地道正行 (2018-a) 『探索的財務ビッグデータ解析-前処理, データラングリング, 再現可能性-』, 商学論究, 第66巻, 第1号, pp. 1-32, 関西学院大学商学研究会.
- [19] 地道正行 (2018-b) 『探索的財務ビッグデータ解析-データ可視化, 統計モデリン

- グ, モデル選択, モデル評価, 動的文書生成, 再現可能研究-』, 商学論究, 第66巻, 第2号, pp. 1-41, 関西学院大学商学研究会.
- [20] 地道正行 (2018-c) 『データサイエンスの基礎: Rによる統計学独習』, 裳華房.
- [21] Jimichi, M., D. Miyamoto, C. Saka, and S. Nagata (2018) Visualization and statistical modeling of financial big data: Double-log modeling with skew-symmetric error distributions, *Japanese Journal of Statistics and Data Science*, Vol. 1, No. 2, pp. 347-371, <https://doi.org/10.1007/s42081-018-0019-1>
- [22] 地道正行, 阪 智香 (2021-a) 『財務データと ESG レーティングデータの前処理と結合』, 商学論究, 第68巻, 第3号, pp. 79-116, 関西学院大学商学研究会.
- [23] 地道正行, 阪 智香 (2021-b) 『財務データと ESG レーティングデータによる株式時価総額の統計モデリング』, 商学論究, 第69巻, 第2号, pp. 1-64, 関西学院大学商学研究会.
- [24] 金森敬文, 竹之内高志, 村田 昇 (2009) 『Rで学ぶデータサイエンス5: パターン認識』, 共立出版.
- [25] Konishi, S. and G. Kitagawa (2008) *Information Criteria and Statistical Modeling*, Springer.
- [26] 小西貞則, 越智義道, 大森裕浩 (2008) 『計算機統計学の方法-ブートストラップ・EM アルゴリズム・MCMC-』, 朝倉書店.
- [27] Kozubowska, T. J. and J. P. Nolanb (2008) Infinite divisibility of skew Gaussian and Laplace laws, *Statistics & Probability Letters*, Vol. 78, pp. 654-660.
- [28] 増田弘毅 (2002) 『GIG 分布と GH 分布に関する解析』, 統計数理, 第50巻, 第2号, pp. 165-199, 統計数理研究所.
- [29] McNeil, A. J., R. Frey, and P. Embrechts (2005) *Quantitative risk management: concepts, techniques and tools*, Princeton University Press.
- (翻訳) 塚原英敦, 小林 俊, 三浦良造, 川崎能典, 内山浩嗣, 中山秀敏 (共訳) (2008) 『定量的リスク管理-基礎概念と数理技法-』, 共立出版.
- [30] McLachlan, G. and D. Peel (2000) *Finite Mixture Models*, Wiley-Interscience.
- [31] Olver, F. W. J., D. W. Lozier, R. F. Boisvert, and C. W. Clark (ed.) (2010) *NIST Handbook of Mathematical Functions*, Cambridge University Press.
- [32] Pfaff, B. (2016) *Financial Risk Modeling and Portfolio Optimization with R, Second Edition*, Wiley.
- [33] Prause, K. (1999) *The generalized hyperbolic models: Estimation, financial derivative and risk measurement*, Ph.D Thesis, Mathematical Faculty, University of Freiburg.
- [34] Pearson, K. (1894) Contributions to the mathematical theory of evolution, *Philosophical Transactions of the Royal Society of London. A*, Vol. 185., pp. 71-110.
- [35] Redner, R. A. and H. F. Walker (1984) Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, Vol. 26, No. 2, pp. 195-239.
- [36] Scott, D. J., D. Würtz, C. Dong, and T. T. Tran (2011) Moments of the generalized


- hyperbolic distribution, *Computational Statistics*, Vol. 26, pp. 459–476.
- [37] 杉山 将 (2009) 『統計的機械学習—生成モデルに基づくパターン認識—』, オーム社.
- [38] Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley Publishing Co.
- [39] 梅津雄一, 中野貴広 (2018) 『RとShinyで作るWebアプリケーション』, C & R 研究所.
- [40] van der Loo, M. and E. de Jonge (2018) *Statistical Data Cleaning with Applications in R*, Wiley.
- (邦訳) 地道正行, 高橋雅夫, 藤野友和, 安川武彦, 和田かず美 (共訳) (2021) 『統計的データクリーニングの理論と実践』, 共立出版.
- [41] Vapnik, V. N. (1998) *Statistical Learning Theory*, Wiley-Blackwell.
- [42] Vapnik, V. N. (2000) *The Nature of Statistical Learning 2nd Edition*, Springer.
- [43] Wickham, H. (2011) The split-apply-combine strategy for data analysis, *Journal of Statistical Software*, Vol. 40, Issue 1.
- [44] Wickham, H. and G. Grolemund (2016) *R for Data Science*, O'Reilly.
- [45] Wu, C. F. J. (1983) On the convergence properties of the EM algorithm, *The Annals of Statistics*, Vol. 11, pp. 95–103.
- [46] 山田作太郎, 稲垣宣生 (1973) 『漁網の選択性について—選択を受けた資料からの推定法—』, 統計数理研究所シンポジウム記事, Vol. 5, pp. 23–29.


## 謝辞


本稿は、2022年6月10日(金)にオンライン開催された九州大学統計科学セミナーにおける講演内容をもとに加筆・修正したものである。このセミナーに招待いただいた九州大学の増田弘毅教授に感謝申し上げるとともに、セミナーに先だってコメントを頂いた関西学院大学商学部の永田修一准教授と、セミナー当日、様々なコメントをいただいた参加者の先生方に感謝の意を表する。

本研究の一部は以下の助成を得ている。

 **科研費** 科学研究費基盤研究 C: 「共有価値創造 (CSV) のための社会環境会計の構築」 (2019年~2022年), 課題番号: 19K02006

 **JHPCN** 2017年度~2022年度学際大規模情報基盤共同利用・共同研究点 (JHPCN) 課題: 「財務ビッグデータの可視化と統計モデリング」, 課題番号: jh171002-NWJ, jh181001-NWJ, jh191002-NWJ, jh201003-NWJ, jh211001-NWJ, jh221001

 **統計数理研究所** 2022年度統計数理研究所計思考院公募型人材育成事業ワークショップ 課題: 「探索的ビッグデータ解析と再現可能研究」, 課題番号: 2022-思考院-7005

 **統計数理研究所** 2022年度統計数理研究所公募型共同利用一般研究 2 課題: 「財務ビッグデータの時空間分析と可視化に関する研究」, 課題番号: 2022-ISMCRP-2030



関西学院大学図書館図書費 B, 研究設備費 (III), 個人研究費



何時もながら、Bureau van Dijk の増田歩氏にはデータ抽出に関してご足労を賜った。ここに感謝の意を表する。

## 付録 A データとその要約

本稿で利用するデータは、Bureau van Dijk 社のデータベース Osiris の 2020年 3 月版から上場企業（上場廃止企業含む）96,377社の主要財務情報（91項目）に関する連結（Consolidated）データを抽出したものを前処理したデータセット OsirisC2020 と FTSE Russell 社から提供される 2015～2020 年（6 年間）の各年の先進国 25カ国 2,167社・新興国 24カ国 1,778社の合計 49カ国 3,945社を対象とした ESG レーティングデータセット FTSE2020 を ISIN コード<sup>14)</sup> で結合したものである（詳細は、地道、阪（2021-a, b）を参照のこと）。

### A.1 データにおける変数仕様

データは表形式で与えられており、列（変数）の仕様は以下の通りである：

firmID: 企業名+BvD ID

year: 財務データセット OsirisC2020 の会計年度

year\_SPT: ESG レーティングデータセット FTSE2020 におけるデータ公表年

market\_cap: 株式時価総額（単位：100万 US ドル）

shareholders: 純資産（単位：1,000 US ドル）

14) 国際証券コード仕様 ISO 6166 で定められている全世界共通の証券系コードであり、各国一機関のみが自国の ISIN コードの付番権限を付与されている。基本構成は、国名コード（2桁）+基本コード（9桁）+チェックディジット（1桁）の計12桁からなる。（日本取引所グループ用語集 <https://www.jpx.co.jp/glossary/h-n/489.html> 参照）。

net\_income: 当期純利益 (単位: 1,000 US ドル)  
log\_mc: 株式時価総額の対数 (単位: 100万 US ドルの自然対数)  
log\_sh: 純資産の対数 (単位: 1,000 US ドルの自然対数値)  
log\_ni: 当期純利益の対数 (単位: 1,000 US ドルの自然対数)  
ESG: ESG レーティングのサマリー (summary) 指標  
E: ESG レーティングの環境 (Environment) 指標  
S: ESG レーティングの社会 (Social) 指標  
G: ESG レーティングのガバナンス (Governance) 指標  
type: 国のタイプ E: 新興国 (Emerging country), G: 先進国 (Global (developed) country)

ここで、株式時価総額 market\_cap (とその対数 log\_mc), 純資産 shareholders (とその対数 log\_sh), 当期純利益 net\_income (とその対数 log\_ni) がデータセット OsirisC2020 からのものであり、サマリー指標 ESG, 環境指標 E, 社会指標 S, ガバナンス指標 G, 国のタイプ type がデータセット FTSE2020 からのものである。

財務データセット OsirisC2020 の年情報 year については、会計年度が当てられているのに対して、ESG レーティングデータセット FTSE2020 の年情報 year\_SPT は、公表された年が当てられている。これらの年情報は表記上 2 年のラグがあり、この理由としては、ESG レーティングデータが公表されるまでに約 2 年の時間を要するため ESG レーティングの評価のようになった企業の ESG 情報の対象年度は、2 年前のものであると想定し、財務データの会計年度と ESG 情報の対象年度 (公表年の 2 年前の年度) を対応させたからである (地道, 阪 (2021-a) 参照)。

## A.2 データの要約

データの要約を以下に与える：

## データの要約

firmID	year	year_SPT	market_cap	shareholders
Length:14366	Min. :2013	Min. :2015	Min. : 32	Min. : 7874
Class :character	1st Qu.:2014	1st Qu.:2016	1st Qu.: 2749	1st Qu.: 1443453
Mode :character	Median :2016	Median :2018	Median : 5890	Median : 2997843
	Mean :2016	Mean :2018	Mean : 14724	Mean : 7285962
	3rd Qu.:2017	3rd Qu.:2019	3rd Qu.: 13060	3rd Qu.: 6746400
	Max. :2018	Max. :2020	Max. :1090308	Max. :294365704

net_income	log.mc	log.sh	log.ni	ESG
Min. : 74	Min. : 3.465	Min. : 8.971	Min. : 4.304	Min. :0.000
1st Qu.: 147567	1st Qu.: 7.919	1st Qu.:14.183	1st Qu.:11.902	1st Qu.:1.700
Median : 326486	Median : 8.681	Median :14.913	Median :12.696	Median :2.500
Mean : 921304	Mean : 8.758	Mean :14.975	Mean :12.760	Mean :2.473
3rd Qu.: 767808	3rd Qu.: 9.477	3rd Qu.:15.725	3rd Qu.:13.551	3rd Qu.:3.300
Max. :98616437	Max. :13.902	Max. :19.500	Max. :18.407	Max. :5.000
				NA's :885

E	S	G	type
Min. :0.000	Min. :0.000	Min. :0.000	E:5172
1st Qu.:1.000	1st Qu.:1.200	1st Qu.:2.500	G:9194
Median :2.000	Median :2.300	Median :3.100	
Mean :2.126	Mean :2.209	Mean :3.168	
3rd Qu.:3.100	3rd Qu.:3.200	3rd Qu.:4.000	
Max. :5.000	Max. :5.000	Max. :5.000	
NA's :885	NA's :885	NA's :885	

この結果において、欠測値が存在するが、全て ESG レーティングデータに由来するものであり、本稿の研究対象は株式時価総額（財務データ）であるので、欠測値に関して影響は受けない。

### A.3 国タイプ情報

本稿では、各国を先進国（G）と新興国（E）のタイプ別に分類したデータを利用しているが、以下に各国のタイプを与える。なお、この分類は、FTSE Russell 社から提供されている情報を利用している。

表10：国タイプ情報

	国名	ISO2c	国タイプ
1	United Arab Emirates	AE	E
2	Brazil	BR	E
3	Chile	CL	E
4	China	CN	E
5	Colombia	CO	E
6	Czechia	CZ	E
7	Egypt	EG	E
8	Greece	GR	E
9	Hungary	HU	E
10	Indonesia	ID	E
11	India	IN	E
12	Kuwait	KW	E
13	Mexico	MX	E
14	Malaysia	MY	E
15	Peru	PE	E
16	Philippines	PH	E
17	Pakistan	PK	E
18	Poland	PL	E
19	Qatar	QA	E
20	Russia	RU	E
21	Saudi Arabia	SA	E
22	Thailand	TH	E
23	Turkey	TR	E
24	Taiwan	TW	E
25	South Africa	ZA	E
26	Austria	AT	G
27	Australia	AU	G
28	Belgium	BE	G
29	Canada	CA	G
30	Switzerland	CH	G
31	Germany	DE	G
32	Denmark	DK	G
33	Spain	ES	G
34	Finland	FI	G
35	France	FR	G
36	United Kingdom	GB	G
37	Greece	GR	G
38	Hong Kong SAR China	HK	G
39	Ireland	IE	G
40	Israel	IL	G
41	Italy	IT	G
42	Japan	JP	G
43	South Korea	KR	G
44	Netherlands	NL	G
45	Norway	NO	G
46	New Zealand	NZ	G
47	Poland	PL	G
48	Portugal	PT	G
49	Sweden	SE	G
50	Singapore	SG	G
51	United States	US	G

### A.4 欠測情報

データが持つ欠測値の情報を可視化したものを図33に与える。プロットは、VIMパッケージ付属の `aggr` 関数を利用して描いたものであり、左図に変数の欠測値の割合をグレーの棒グラフで表しており、このデータの場合では、変数 `ESG`, `E`, `S`, `G` に、それぞれ、約6%の欠測値が存在することがわかる。また、右図からは、欠測が起こるパターン（グレー）として、変数 `ESG`, `E`, `S`, `G` の全てが欠測しているものだけであることを示している。この図からも、欠測は、ESGレーティングデータに由来するものであり、特に全てのESG関連の変数が全て欠測しているパターンのみであることもわかった<sup>15)</sup>。

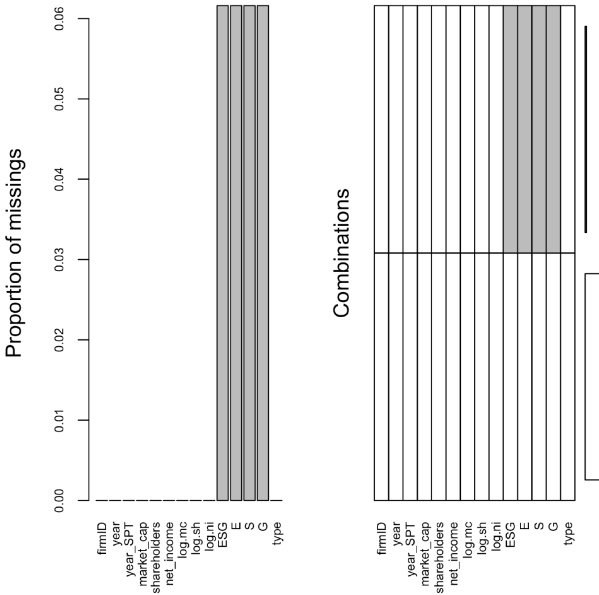


図33：欠測情報の可視化：VIMパッケージ付属の `aggr` 関数を利用

また、表11には、年度毎にデータに含まれる企業数（先進国企業数，新興国企業数，総企業数）と、表12には、欠測値を除去した年度毎の企業数（先

15) 欠測値の扱いを含むデータクリーニングについては、van der Loo and E. de Jonge (2018) が包括的で参考になる。

進国企業数，新興国企業数，総企業数）を与えている。

表11：年度別企業数

	会計年度	ESG 指標年度	先進国企業数	新興国企業数	総企業数
1	2013	2015	1466	629	2095
2	2014	2016	1489	647	2136
3	2015	2017	1463	647	2110
4	2016	2018	1538	698	2236
5	2017	2019	1625	1282	2907
6	2018	2020	1613	1269	2882

表12：欠測値除去後の年度別企業数

	会計年度	ESG 指標年度	先進国企業数	新興国企業数	総企業数
1	2013	2015	1422	594	2016
2	2014	2016	1465	633	2098
3	2015	2017	1457	641	2098
4	2016	2018	1527	679	2206
5	2017	2019	1607	724	2331
6	2018	2020	1605	1127	2732

## 付録 B 非対称分布族

Azzalini (1985) は、正規分布に関して、ある種の母数をもつ累積分布関数 (cumulative distribution function) を確率密度関数 (probability density function) に乗ずることによって、非対称性をもつ分布族を提案した。この分布族には典型的なものとして、非対称正規分布<sup>16)</sup> (skew-normal distribution) があり、その後、様々な拡張がなされている。本稿では、以下に非対称正規分布と非対称ティータ分布の定義を以下に与えるが、詳細は、Azzalini and Capitanio (2014) を参照されたい。

16) 非対称正規分布は歪正規分布と呼ばれることもある。

### B.1 非対称正規分布

**定義 1** (非対称正規分布) 確率変数  $X$  が確率密度関数

$$f_{\text{SN}}(x|\xi, \omega, \alpha) = \frac{2}{\omega} \phi\left(\frac{x-\xi}{\omega}\right) \Phi\left(\alpha \frac{x-\xi}{\omega}\right), x \in \mathbb{R} := (-\infty, \infty) \quad (15)$$

をもつとき、確率変数  $X$  は非対称正規分布  $\text{SN}(\xi, \omega^2, \alpha)$  に従うと呼ばれ、

$$X \sim \text{SN}(\xi, \omega^2, \alpha)$$

と書かれる。ここで、

$$\xi \in \mathbb{R}, \omega \in \mathbb{R}^+ := (0, \infty), \alpha \in \mathbb{R}$$

は未知母数であり、 $(\xi, \omega^2, \alpha)$  は直接母数 (direct parameters) と呼ばれる。また、

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \Phi(z) = \int_{-\infty}^z \phi(x) dx \quad (z \in \mathbb{R})$$

は、それぞれ、標準正規分布  $N(0, 1)$  の確率密度関数と累積分布関数である。

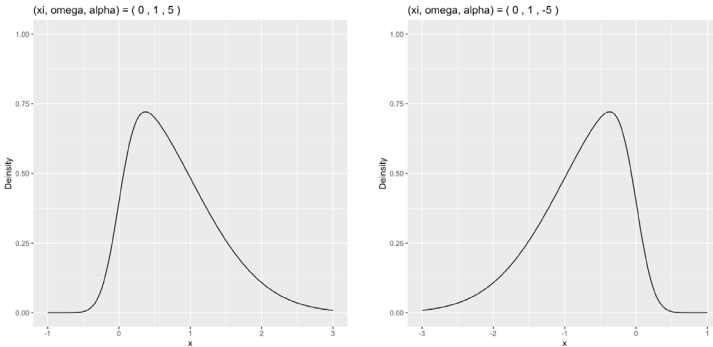


図34：非対称正規分布の pdf：左図： $(\xi, \omega, \alpha) = (0, 1, 5)$  の場合、右図： $(\xi, \omega, \alpha) = (0, 1, -5)$  の場合

### B.2 非対称テイー分布

**定義 2** (非対称テイー分布) 確率変数  $X$  が pdf:

$$f_{\text{ST}}(x|\xi, \omega, \alpha, \nu) = \frac{2}{\omega} f_t\left(\frac{x-\xi}{\omega} \middle| \nu\right) F_t\left(\alpha \frac{x-\xi}{\omega} \sqrt{\frac{\nu+1}{\left(\frac{x-\xi}{\omega}\right)^2 + \nu}} \middle| \nu+1\right), x \in \mathbb{R}$$

(16)

をもつとき、確率変数  $X$  は非対称ティー分布  $\text{ST}(\xi, \omega^2, \alpha, \nu)$  に従うと呼ばれ、

$$X \sim \text{ST}(\xi, \omega^2, \alpha, \nu)$$

と書かれる。ここで、

$$\xi \in \mathbb{R}, \quad \omega \in \mathbb{R}^+, \quad \alpha \in \mathbb{R}, \quad \nu \in \mathbb{R}^+$$

は未知母数であり、 $(\xi, \omega^2, \alpha, \nu)$  は直接母数と呼ばれる。また、

$$f_i(z|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad F_i(z|\nu) = \int_{-\infty}^z f_i(x|\nu) dx$$

は、それぞれ、自由度  $\nu$  のティー分布の確率密度関数と累積分布関数であり、

$$\Gamma(x) := \int_0^{\infty} t^{x-1} e^{-t} dt$$

はガンマ関数である。

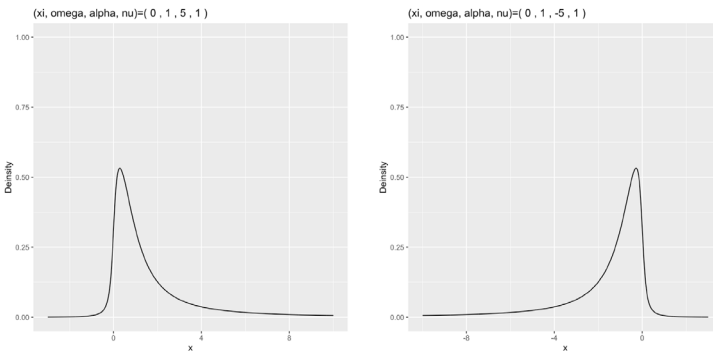


図35：非対称ティー分布の pdf：左図： $(\xi, \omega, \alpha, \nu) = (0, 1, 5, 1)$  の場合，右図： $(\xi, \omega, \alpha, \nu) = (0, 1, -5, 1)$  の場合

## 付録 C 一般化双曲型分布族

ここでは、正規尺度平均混合 (cf. Bandolff-Nielsen *et al.* (1982)) の定義からはじめ、一般化逆ガウス分布の正規尺度平均混合として、一般化双曲型



分布 (cf. Bandolff-Nielsen (1977)) が得られることを外観する<sup>17)</sup>。

さらに、一般化双曲型分布は、ある種の母数 ( $\lambda$ ) を変化させることによって分布族を構成するため、一般化双曲型分布族という名称も利用している。例えば、特別な場合 ( $\lambda = -1/2$ ) として、正規逆ガウス分布が導出されることを以下で述べる。なお、一般化双曲型分布 (族) についての詳細な調査が増田 (2002) でなされており非常に有益である。また、一般化双曲型分布の積率については、Scott *et al.* (2011) がコンパクトにまとまっている。さらに、一般化双曲型分布族の金融分野への応用としては、Prause (1999), McNeil *et al.* (2005) 等を参照されたい。

**定義 3 (正規尺度平均混合)**  $X$  を  $\mathbb{R}_+$  上で定義された確率変数とし、 $\varepsilon$  を  $X$  と独立に標準正規分布  $N(0, 1)$  に従う確率変数とする。このとき、 $\mu \in \mathbb{R}, \beta \in \mathbb{R}$  に対して以下のように確率変数を定義する：

$$Y := \mu + \beta X + \sqrt{X} \varepsilon \quad (17)$$

このとき、 $X$  を混合要素 (mixing distribution) といい、 $Y$  を  $X$  による正規尺度平均混合 (normal variancemean mixture) という。

**命題 2** 確率変数  $X$  が確率密度関数  $p_X(x)$  を持つと仮定する。定義 3 より、条件付き分布

$$Y|X=x \sim N(\mu + \beta x, x)$$

---

17) 本稿執筆中 (2022年7月15日) に、Ole Eiler Barndorff-Nielsen 先生 (以下、バンドルフ・ニールセン先生) が2022年6月26日にご逝去されたとの事実をメーリングリストで知った。筆者が、大阪大学基礎工学研究科在学中に、バンドルフ・ニールセン先生が数理教室を訪れる機会があり、微分幾何学を統計学へ応用する内容の講演を拝聴した。翌日助手の先生と一緒にバンドルフ・ニールセン先生が滞在するホテルにお迎えに上がり、広島行きの新幹線にお見送りをさせていただいたことが昨日のように思い出させる。そのとき、バンドルフ・ニールセン先生から、いろいろとお伺いしたこと (数学者の定義「数学者はコーヒーを定理に変換するマシンである。」や、伊藤清先生から囲碁を教わったという逸話) は今でも鮮明に記憶に残っている。後日、お送り頂いた御高書 (Barndorff-Nielsen (1988)) は生涯の宝物である。全て、感謝の念に堪えない。先生のご冥福をお祈りする。

が成り立つ。Y の確率密度関数は、

$$\begin{aligned} p_Y(y) &= \int_0^\infty \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{(y-\mu-\beta x)^2}{2x}\right\} p_X(x) dx \\ &= \frac{1}{\sqrt{2\pi}} e^{\beta(y-\mu)} \int_0^\infty x^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{(y-\mu)^2}{x} + \beta^2 x\right)\right\} p_X(x) dx \end{aligned}$$

で与えられる。

**定義 4 (一般化逆ガウス分布)**  $\mathbb{R}_+$  上で定義された確率変数  $X$  が確率密度関数

$$f_{\text{GIG}}(x|\lambda, \delta, \gamma) := \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\gamma\delta)} x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \gamma^2 x\right)\right\}, \quad x > 0 \quad (18)$$

を持つとき、 $X$  は一般化逆ガウス (Generalized Inverse Gaussian: GIG) 分布に従うといい、 $X \sim \text{GIG}(\lambda, \delta, \gamma)$  と書く。

**注意 1 (一般化逆ガウス分布の他の母数付け)** 一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  の他の母数付け (parameterization) として、以下のようなものがある (Bandolff-Nielsen (1977), (7.1) 式参照) :

$$f_{\text{gig}}(x|\lambda, \chi, \phi) := \frac{(\phi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\phi})} x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{x} + \phi x\right)\right\}, \quad x > 0 \quad (19)$$

この分布は記号として、 $\text{gig}(\lambda, \chi, \phi)$  と書かれる。

この表現は、一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  において、

$$\chi := \delta^2, \quad \phi := \gamma^2$$

とおいたものであり、この意味で、

$$\text{GIG}(\lambda, \delta, \gamma) = \text{GIG}(\lambda, \sqrt{\chi}, \sqrt{\phi}) = \text{gig}(\lambda, \chi, \phi)$$

が成り立つ。

**定義 5 (逆ガウス分布)** 一般化逆ガウス分布において、 $\lambda = -1/2$  の場合を、逆ガウス (Inverse Gaussian: IG) 分布と呼ぶ :

$$\text{IG}(\delta, \gamma) := \text{GIG}(-1/2, \delta, \gamma) \quad (20)$$

**命題 3** 確率変数  $X$  が  $\text{IG}(\delta, \gamma)$  に従うとき、確率密度関数は、

$$f_{\text{IG}}(x|\delta, \gamma) := \frac{\delta}{\sqrt{2\pi}} e^{\gamma x} x^{-3/2} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \gamma^2 x\right)\right\}, \quad x > 0 \quad (21)$$

で与えられる。

**定義 6** (一般化双曲型分布) 正規尺度平均混合の定義 3 において、 $X$  が一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  に従う場合、 $Y$  の分布を一般化双曲型 (Generalized Hyperbolic: GH) 分布といい、 $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  と記号的に表される<sup>18)</sup>。

**命題 4** 確率変数  $Y$  が一般化双曲型分布  $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  に従うとき、その確率密度関数は以下で与えられる：

$$f_{\text{GH}}(y|\lambda, \alpha, \beta, \delta, \mu) := \frac{(\alpha^2 - \beta^2)^{\lambda/2} \{h(y; \delta, \mu)\}^{\lambda-1/2}}{\sqrt{2\pi} \alpha^{\lambda-1/2} \delta^2 K_\lambda(\delta \sqrt{\alpha^2 - \beta^2})} K_{\lambda-1/2}(\alpha h(y; \delta, \mu)) \exp\{\beta(y - \mu)\} \quad (22)$$

ここで、 $K_\lambda(\cdot)$  は第 2 種変形ベッセル関数 (付録 E 参照) であり、

$$h(y; \delta, \mu) := \sqrt{(y - \mu)^2 + \delta^2} \quad (23)$$

である。また、一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  の正規尺度平均混合  $Y = \mu + \beta X + \sqrt{X} \varepsilon$  における母数  $\beta, \gamma$  を以下のような変換を行っている：

$$\alpha := \sqrt{\beta^2 + \gamma^2} \quad (24)$$

**注意 2** (一般化双曲型分布の母数空間と母数の役割) 一般化双曲型分布  $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  の母数空間  $\Theta$  は以下のように与えられる：

$$\Theta := \left\{ (\lambda, \alpha, \beta, \delta, \mu) \left| \begin{array}{l} \lambda > 0 \text{ のとき, } \delta \geq 0, \alpha > |\beta|, \\ \mu \in \mathbb{R}, \lambda = 0 \text{ のとき, } \delta > 0, \alpha > |\beta|, \\ \lambda < 0 \text{ のとき, } \delta > 0, \alpha \geq |\beta| \end{array} \right. \right\} \quad (25)$$

18) 一般化双曲型分布  $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  は、一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  による正規尺度平均混合から導かれることから、正規一般化逆ガウス (Normal Generalized Inverse Gaussian: NGIG) 分布と呼ばれることもある。

また、各母数は以下のような役割がある：

$\lambda$ : GH 分布族の決定

$\alpha$ : 各固定された  $\beta$  に対する峰点付近での集中度 (steepness: 尖り方の度合い)

$\beta$ : 非対称性

$\delta$ : 尺度

$\mu$ : 位置

ここで、 $\alpha$  は裾の厚さを決めることにも注意する必要がある。

**定義 7 (正規逆ガウス分布)**  $\lambda = -1/2$  のときの一般化双曲型分布  $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  を正規逆ガウス (Normal Inverse Gaussian: NIG) 分布と呼ぶ：

$$\text{NIG}(\alpha, \beta, \delta, \mu) := \text{GH}(-1/2, \alpha, \beta, \delta, \mu) \quad (26)$$

**命題 5** 確率変数  $Y$  が正規逆ガウス分布  $\text{NIG}(\alpha, \beta, \delta, \mu)$  に従うとき、その確率密度関数は以下で与えられる：

$$f_{\text{NIG}}(y|\alpha, \beta, \delta, \mu) := \frac{\alpha\delta}{\pi} \exp\{\delta\sqrt{\alpha^2 - \beta^2} + \beta(y - \mu)\} \frac{K_1(\alpha\sqrt{(y - \mu)^2 + \delta^2})}{\sqrt{(y - \mu)^2 + \delta^2}} \quad (27)$$

**命題 6 (正規逆ガウス分布の再生性)** 正規逆ガウス分布は再生性をもつ。つまり、確率変数  $X_1$  と  $X_2$  が、それぞれ、独立に正規逆ガウス分布  $\text{NIG}(\alpha, \beta, \delta_1, \mu_1)$ ,  $\text{NIG}(\alpha, \beta, \delta_2, \mu_2)$  に従うとき、

$$X_1 + X_2 \sim \text{NIG}(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2) \quad (28)$$

が成り立つ。

**注意 3 (正規逆ガウス分布の正規尺度平均混合からの導出)** 正規逆ガウス分布という名称は、正規尺度平均混合の定義 3 において、 $X$  が逆ガウス分布  $\text{IG}(\delta, \mu)$  に従うときの  $Y$  の分布であることから、この名称が付いたといわれる。

**注意 4** (正規尺度平均混合と一般化双曲型分布族の関係) これまでの議論から, 正規尺度平均混合 (17) において, 確率変数  $X$  が一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  に従うとき, 確率変数  $Y$  は一般化双曲型分布  $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  に従う。

また, 確率変数  $X$  が逆ガウス分布  $\text{IG}(\delta, \gamma)$  に従うとき, 確率変数  $Y$  は正規逆ガウス分布  $\text{NIG}(\alpha, \beta, \delta, \mu)$  に従う。

一般化逆ガウス分布  $\text{GIG}(\lambda, \delta, \gamma)$  において,  $\lambda = -1/2$  のときが, 逆ガウス分布  $\text{IG}(\delta, \gamma)$  であり, 一般化双曲型分布  $\text{GH}(\lambda, \alpha, \beta, \delta, \mu)$  において,  $\lambda = -1/2$  のときが, 正規逆ガウス分布  $\text{NIG}(\alpha, \beta, \delta, \mu)$  である。

以上の関係を図式化したものを以下に与える:

$$\begin{array}{ccc}
 X & \xrightarrow{\mu + \beta X + \sqrt{X} \varepsilon} & Y \\
 \text{GIG}(\lambda, \delta, \gamma) & \longrightarrow & \text{GH}(\lambda, \alpha, \beta, \delta, \mu) \\
 \lambda = -1/2 \downarrow & & \lambda = -1/2 \downarrow \\
 \text{IG}(\delta, \gamma) & \longrightarrow & \text{NIG}(\alpha, \beta, \delta, \mu)
 \end{array}$$

### C.1 一般化双曲型分布の確率密度関数可視化のための Web アプリケーション

一般化双曲型分布の母数を変化に伴って確率密度関数の形状がどのように変化するかは, その関数をみているだけでは難しい. そこで, 母数をインタラクティブに変化させるに伴って確率密度関数をダイナミックに可視化することができれば, 確率密度関数の形状に対するその母数の役割を視覚的に捉えることができる. ここでは, RStudio 社<sup>19)</sup> が開発している Shiny<sup>20)</sup> を利用することによって, インタラクティブ性をもつ Web アプリケーションを作成する<sup>21)</sup>. 以下にその手順を与える:

19) <https://www.rstudio.com/>

20) Shiny は R を用いて簡単に Web アプリケーションをつくるための R パッケージ (Shiny) である. 詳しくは, <https://www.rstudio.com/products/shiny/> を参照されたい. なお, 梅津, 中野 (2018) も丁寧な説明があるので参照されたい.

21) ここでは, 一般化双曲型分布の確率密度関数可視化のための Web アプリケーション

(S1) リスト1で与えられるスクリプトをRmdファイル `interactive-pdf-GH.Rmd`<sup>22)</sup> として保存

(S2) **RStudio**<sup>23)</sup> で `interactive-pdf-GH.Rmd` ファイルを開く (図36)

(S3)  ボタンをクリック

以上の手順によって、レンダリングが行われ、インタラクティブ機能をもつ専用のウィンドウが開く (図38).

コード1：一般化双曲型分布の確率密度関数をインタラクティブに描画するためのRmdファイル：`interactive-pdf-GH.Rmd`

```

1 ---
2 title: "Interactive_Visualization_of_GH_Distribution"
3 output: html_document
4 runtime: shiny
5 ---
6
7 ```{r setup, include=FALSE}
8 knitr::opts_chunk$set(echo = TRUE)
9 ```
10
11 ```{r echo=FALSE}
12 require(dplyr)
13 require(ggplot2)
14 require(GeneralizedHyperbolic)
15 ```
16
17 ### Plot of Density Function of Generalized Hyperbolic Distribution:
18 ### GH( $\lambda$ ,  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\mu$ )
19 ```{r histogram, echo=FALSE}
20 inputPanel(
21   sliderInput("lambda", label = "lambda:", min = -10, max = 10, value = 0, step = 0.01),
22   sliderInput("alpha", label = "alpha:", min = 0, max = 10, value = 1, step = 0.01),
23   sliderInput("beta", label = "beta:", min = -10, max = 10, value = 0, step = 0.01),
24   sliderInput("delta", label = "delta:", min = 0, max = 10, value = 1, step = 0.01),
25   sliderInput("mu", label = "mu:", min = -10, max = 10, value = 0, step = 0.01),
26   sliderInput("xmax", label = "xmax:", min = 10, max = 20, value = 10, step = 1),
27   sliderInput("ymax", label = "ymax:", min = 0.5, max = 2, value = 1, step = 0.1)
28 )
29 sliderValues <- reactive({
30   data.frame(
31     Name = c("lambda", "alpha", "beta", "delta", "mu"),

```

の作成について述べたが、正規逆ガウス分布についても同様に作成可能である。

- 22) Rmd (RMarkdown) ファイルは Markdown 言語に R のコードを埋め込むことを可能にしたものである。
- 23) RStudio は、RStudio 社で開発されている R の統合開発環境 (Integrated Development Environment: IDE) である。以下の URL を参照のこと：<https://www.rstudio.com/products/rstudio/>

```

32 Value = as.character(c(as.numeric(input$lambda),
33                       as.numeric(input$alpha),
34                       as.numeric(input$beta),
35                       as.numeric(input$delta),
36                       as.numeric(input$mu)),
37                       stringsAsFactors = FALSE))
38 })
39 # Show the values in an HTML table ----
40 renderTable({
41   sliderValues()
42 })
43 renderPlot({
44   x <- seq(-as.numeric(input$xmax), as.numeric(input$xmax), 0.01)
45   df <- data.frame(x = x,
46                   y = dghyp(x,
47                             param = c(as.numeric(input$mu),
48                                       as.numeric(input$delta),
49                                       as.numeric(input$alpha),
50                                       as.numeric(input$beta),
51                                       as.numeric(input$lambda))))
52   p <- df %>% ggplot(aes(x, y)) + geom_line()
53   p + coord_cartesian(xlim = c(-as.numeric(input$xmax), as.numeric(input$xmax)), ylim = c(0, as.numeric(input$ymax)))
54 })
55

```

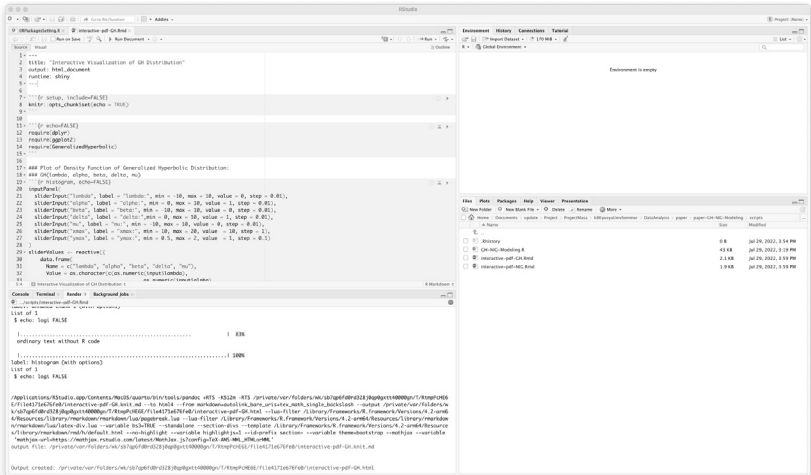


図36 : RStudio: ソースコードペイン上で Rmd ファイル (interactive-pdf-GH.Rmd) を開いたところ

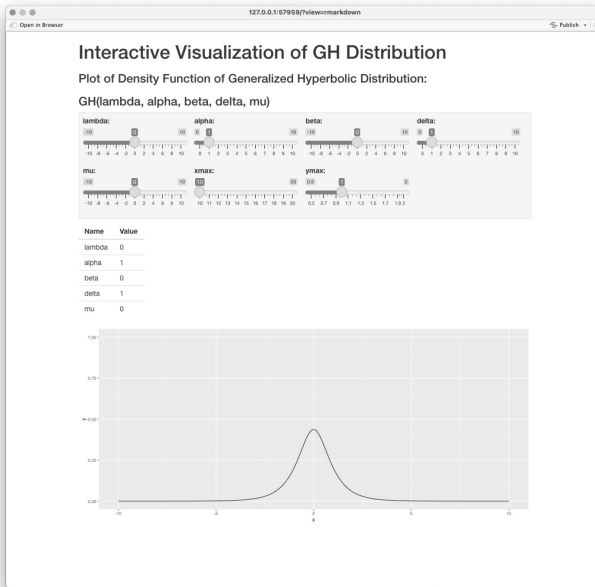


図37：専用ウィンドウ：インタラクティブ付き

さらに、専用ウィンドウ（図38）の **Open in Browser** ボタンをクリックすることによって、Web ブラウザが開き、インタラクティブ性をもつグラフィックスを表示することができる（図37）。このことは、一般化双曲型分布の母数をリアルタイムに変更し、そのときの確率密度関数を描画することによって、その結果を可視化する Web アプリケーションが作成されたことを表していることに注意しよう。なお、このアプリケーションを利用することによって、各母数の役割をダイナミックに把握することが可能となる。



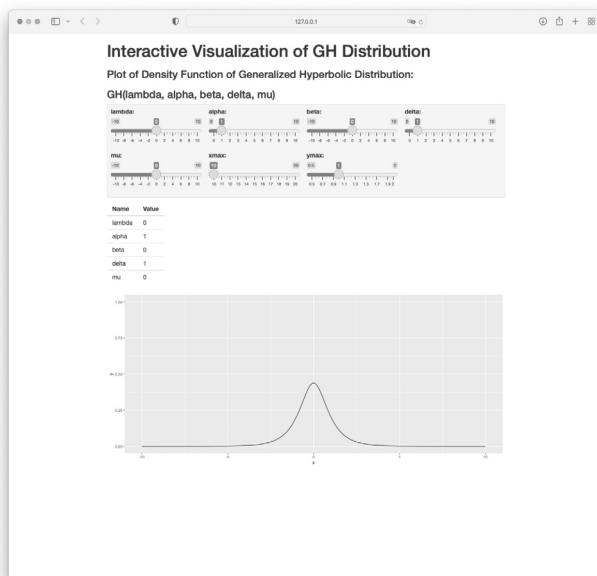


図38：一般化双曲型分布の確率密度関数の描画のための Web アプリケーション：各母数  $\lambda$  (**lambda**),  $\alpha$  (**alpha**),  $\beta$  (**beta**),  $\delta$  (**delta**),  $\mu$  (**mu**) の値をスライダーを調整することによって選択でき、その値に対する確率密度関数がダイナミックに描画される。なお、**xmax** と **ymax** のスライダーを調整することによって、横軸と縦軸の範囲（最小値，最大値）も変更できる。

## 付録 D 混合分布

ここでは、混合分布として、有限混合分布 (finite mixture distribution) の導出と、その母数に関する最尤推定を与え、さらに情報量規準を定義する。

### D.1 有限混合分布

確率空間  $(\Omega, \mathcal{A}, P)$  を考える。母集団  $\Omega$  に属する個体を  $\omega (\in \Omega)$  で表す。また、母集団  $\Omega$  が以下のようなカテゴリに分けられている (互いに素な有限部分集合  $C_1, \dots, C_k \in \mathcal{A}$  の層に直和分割されている) ものとする：

$$\Omega = C_1 + \cdots + C_k \iff \Omega = \bigcup_{j=1}^k C_j, C_j \neq \emptyset, C_j \cap C_{j'} = \emptyset (j \neq j') \quad (29)$$

以下が成り立つことに注意しよう：

$$P(C_1) + \cdots + P(C_k) = P(\Omega) = 1 \quad (30)$$

確率空間  $(\Omega, \mathcal{A}, P)$  上で定義された、2変量確率変数  $(X, Y) = (X(\omega), Y(\omega))$  を考え、ここで、 $L_c := \{1, \dots, k\}$  とし、確率変数  $Y$  は、

$$\begin{array}{ccc} Y : \Omega & \rightarrow & L_c \\ \cup & & \cup \\ \omega & \mapsto & y \end{array} \quad (31)$$

なる関数であり、第  $y$  カテゴリ  $C_y$  に属する個体  $\omega$  に対して  $y \in L_c$  を与えるものとする。よって、

$$C_y = \{\omega \in \Omega; Y(\omega) = y\} \quad (32)$$

が成り立つ。以上のことから  $Y$  は以下のような確率分布に従う：

$$g(y) := P(Y=y) := P(\{\omega \in \Omega; Y(\omega) = y\}) = P(C_y) =: \varpi_y \quad (33)$$

ここで、(30) より以下が成り立つ：

$$\varpi_1 + \cdots + \varpi_k = 1, \quad \varpi_y \in (0, 1) \quad (34)$$

次に、 $Y=y$  が与えられたもとの、 $X$  の条件付き確率密度関数を以下で定義する：

$$f(x|y) := f_y(x|\theta_y) \quad (35)$$

このとき、2変量確率変数  $(X, Y)$  の同時確率密度関数は以下のように与えられる：

$$f(x, y|\theta_y) := g(y)f(x|y) = \varpi_y f_y(x|\theta_y) \quad (36)$$

このとき、確率変数  $X$  の周辺確率密度関数は以下で与えられる：

$$f(x|\theta) := \sum_{y=1}^k f(x, y|\theta_y) = \sum_{y=1}^k \varpi_y f_y(x|\theta_y) \quad (37)$$

ここで、

$$\boldsymbol{\theta} := \begin{bmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_k \\ \varpi_1 \\ \vdots \\ \varpi_k \end{bmatrix} \quad (38)$$

は母数ベクトルである。

一般に、確率変数  $X$  が確率密度関数 (37) を持つとき、有限混合分布に従うと呼ばれ、

$$X \sim \text{Mix}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k, \varpi_1, \dots, \varpi_k) (= \text{Mix}(\boldsymbol{\theta}))$$

と表される。なお、 $\varpi_y$  は混合比率 (mixing proportion) と呼ばれる。

**注意 5** 制約式 (34) より、混合比率には以下の関係がある：

$$\varpi_k = 1 - \varpi_1 - \dots - \varpi_{k-1} \quad (39)$$

この関係から、 $\varpi_1, \dots, \varpi_k$  の中で、どれか一つは他のものから決定されることがわかる。このことは、いわゆる、「自由度」が  $k-1$  であるを表している<sup>24)</sup>。

**注意 6** (混合正規分布) 有限混合分布の最も基本的なものとして、 $Y=y$  が与えられたもとでの  $X$  の条件付き確率密度関数を正規分布

$$f(x|y) = f_y(x|\boldsymbol{\theta}_y) = f(x|\boldsymbol{\theta}_y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{(x-\mu_y)^2}{2\sigma_y^2}\right\} \quad (40)$$

とした場合に  $X$  の周辺分布として得られる

$$f(x|\boldsymbol{\theta}) = \sum_{y=1}^k \varpi_y f_y(x|\boldsymbol{\theta}_y) = \sum_{y=1}^k \varpi_y \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{(x-\mu_y)^2}{2\sigma_y^2}\right\} \quad (41)$$

24) ここで利用した自由度という用語は、カイ自乗分布に直接的に関係するものではなく、母数に関する制約に対するものである。R の AIC 等の出力結果として、母数の次元 (個数) に対するラベルとして df (degree of freedom) と表記される場合があるのは、この理由によるものと思われる。

があり、混合正規分布 (mixture of normal distribution) または正規混合分布 (normal mixture distribution, Gaussian mixture distribution) と呼ばれる。混合正規分布については、Peason (1894) までさかのぼることができ、理論・応用の両面で数多くの研究がある。たとえば、混合正規分布の母数に関するモーメント法による推定については、Peason (1894), Choen (1967) 等を、最尤法によるものについては Day (1969) を参照されたい。混合正規分布の興味深い応用としては、山田、稲垣 (1973) がある。なお、正規分布を含む一般的な分布の混合に関しては McLachlan and Peel (2000) が詳しい。

## D.2 最尤推定

2変量確率変数  $(X, Y)$  に対する観測値のセット

$$\{(x_i, y_i); i=1, \dots, n\} = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

が与えられたとき、第  $y$  カテゴリ  $C_y$  から抽出された個体番号  $i$  に関する添字集合を以下で定義する：

$$\mathcal{L}_y := \{i; y_i=y\} \quad (42)$$

このとき、観測された個体のうち第  $y$  カテゴリ  $C_y$  に属する個体数  $n_y$  は以下のように表される：

$$n_y := \#\mathcal{L}_y = \#\{i; y_i=y\} \quad (43)$$

ここで、 $\#A$  は集合  $A$  の濃度を表す。

以下の関係が成り立つ：

$$\begin{aligned} y \neq y' &\implies \mathcal{L}_y \cap \mathcal{L}_{y'} = \emptyset, \quad \bigcup_{y \in C} \mathcal{L}_y = \{1, \dots, n\}, \\ \#\mathcal{L}_1 + \dots + \#\mathcal{L}_k &= n_1 + \dots + n_k = n \end{aligned}$$

同時確率密度関数 (36) より、母数ベクトル  $\theta$  に対する対数尤度は、

$$\begin{aligned} \ell(\theta) &:= \log \prod_{i=1}^n f(x_i, y_i | \theta_{y_i}) = \log \prod_{i=1}^n \varpi_{y_i} f_{y_i}(x_i | \theta_{y_i}) \\ &= \sum_{i=1}^n \log(\varpi_{y_i} f_{y_i}(x_i | \theta_{y_i})) \\ &= \sum_{i=1}^n \log(\varpi_{y_i}) + \sum_{i=1}^n \log(f_{y_i}(x_i | \theta_{y_i})) \\ &= \sum_{y=1}^k \sum_{i \in \mathcal{L}_y} \log(\varpi_y) + \sum_{i=1}^k \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \theta_y)) \end{aligned}$$

$$= \sum_{y=1}^k n_y \log(\varpi_y) + \sum_{y=1}^k \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \theta_y)) \quad (44)$$

となる。ここで、混合比率に関する制約式 (34) を考慮すると、以下のような制約付き対数尤度を考える必要がある：

$$\ell(\boldsymbol{\theta}, \lambda) := \ell(\boldsymbol{\theta}) + \lambda \left( 1 - \sum_{y=1}^k \varpi_y \right) \quad (45)$$

よって、尤度方程式は、 $y \in \mathcal{L}_c = \{1, \dots, k\}$  に対して、

$$\frac{\partial \ell(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta}_y} = \frac{\partial}{\partial \boldsymbol{\theta}_y} \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \boldsymbol{\theta}_y)) = 0 \quad (46)$$

$$\frac{\partial \ell(\boldsymbol{\theta}, \lambda)}{\partial \varpi_y} = \frac{\partial}{\partial \varpi_y} n_y \log(\varpi_y) - \lambda = \frac{n_y}{\varpi_y} - \lambda = 0 \quad (47)$$

$$\frac{\partial \ell(\boldsymbol{\theta}, \lambda)}{\partial \lambda} = 1 - \sum_{y=1}^k \varpi_y = 0 \quad (48)$$

となる。尤度方程式 (46) は、第  $y$  カテゴリ  $C_y$  に属するデータ  $\{x_i, i \in \mathcal{L}_y\}$  を利用して、母数ベクトル  $\boldsymbol{\theta}_y$  について対数尤度を最大化するための方程式である。この方程式の解として、母数ベクトル  $\boldsymbol{\theta}_y$  に対する最尤推定値ベクトル  $\hat{\boldsymbol{\theta}}_y$  が与えられる。また、尤度方程式 (47) より、

$$\varpi_y = \frac{n_y}{\lambda} \quad (49)$$

が得られ、これを尤度方程式 (48) に代入することによって、

$$\sum_{y=1}^k \varpi_y = \sum_{y=1}^k \frac{n_y}{\lambda} = \frac{n}{\lambda} = 1 \iff \lambda = n \quad (50)$$

となり、この結果を、(49) に代入することによって、混合比率に対する最尤推定値

$$\hat{\varpi}_y := \frac{n_y}{n} \quad (51)$$

を得る。以上の結果より、母数ベクトル  $\boldsymbol{\theta}$  に対する最尤推定値ベクトルは、

$$\hat{\theta} := \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \\ \hat{\omega}_1 \\ \vdots \\ \hat{\omega}_k \end{bmatrix} = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \\ n_1/n \\ \vdots \\ n_k/n \end{bmatrix} \quad (52)$$

で与えられる.

**注意 7** ここでは, 2変量確率変数  $(X, Y)$  に対する観測値として完全セット

$$\{(x_i, y_i); i=1, \dots, n\} = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

が与えられた場合を考察した. これは, 統計的学習理論<sup>25)</sup> (statistical learning theory) の枠組みでは, 「教師あり学習」(supervised learning) と呼ばれる.

一方, 2変量確率変数  $(X, Y)$  に対する観測値として  $X$  に関するものしか観測されない不完全セット

$$\{x_i; i=1, \dots, n\} = \{x_1, \dots, x_n\}$$

から推測 (学習) することは「教師なし学習」(unsupervised learning) と呼ばれる. なお, 教師なし学習の枠組みからは, EM アルゴリズム (EM algorithm) を利用した推測が可能である. EM アルゴリズムは, 欠測値をもつ場合の推定理論・推定法として Dempster *et al.* (1977) によって提案されたものである. このアルゴリズムの収束性については, Wu (1983) で理論的に考察されている. また, 混合正規分布の母数の推定へ EM アルゴリズムの応用に関しては, Redner and Walker (1984) で詳細に議論されている. なお, Bishop (2006) の第 9 章, 金森ら (2009) の第 4 章はより洗練されていて包括的である. また, 小西ら (2008) には EM アルゴリズムを使った混合正規分布の母数推定をはじめ, ブートストラップ法等の様々なモンテカ

25) (統計的) 学習理論については, Vapnik (1998, 2000), Bishop (2006), Hastie *et al.* (2009), 杉山 (2009), James *et al.* (2021) 等を参照されたい.

ルロ法を使った手法について詳しく議論されている。

### D.3 最大対数尤度と情報量規準

有限混合分布に対する最大対数尤度は，(52) を (44) へ代入して，

$$\begin{aligned} \ell(\hat{\boldsymbol{\theta}}) &= \sum_{y=1}^k n_y \log(\hat{\varpi}_y) + \sum_{y=1}^k \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \hat{\boldsymbol{\theta}}_y)) \\ &= \sum_{y=1}^k n_y \log\left(\frac{n_y}{n}\right) + \sum_{y=1}^k \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \hat{\boldsymbol{\theta}}_y)) \end{aligned} \quad (53)$$

となる。また，情報量規準  $\text{AIC}_{\text{Mix}}$ ,  $\text{BIC}_{\text{Mix}}$  は，それぞれ，

$$\text{AIC}_{\text{Mix}} := -2\ell(\hat{\boldsymbol{\theta}}) + 2\left(\sum_{y=1}^k \dim(\boldsymbol{\theta}_y) + k - 1\right) \quad (54)$$

$$\text{BIC}_{\text{Mix}} := -2\ell(\hat{\boldsymbol{\theta}}) + \left(\sum_{y=1}^k \dim(\boldsymbol{\theta}_y) + k - 1\right) \log n \quad (55)$$

で与えられる。ここで， $\dim(\boldsymbol{\theta}_y)$  は母数ベクトル  $\boldsymbol{\theta}_y$  の次元を表し， $k-1$  は (自由に設定できる)  $\varpi_y$  の個数を表している (注意5 参照)。

情報量規準  $\text{AIC}_{\text{Mix}}$ ,  $\text{BIC}_{\text{Mix}}$  について以下の命題が成り立つ：

**命題 7** 有限混合分布に対する情報量規準  $\text{AIC}_{\text{Mix}}$ ,  $\text{BIC}_{\text{Mix}}$  に関して以下が成り立つ：

$$\text{AIC}_{\text{Mix}} = \sum_{y=1}^k \text{AIC}_y(\hat{\boldsymbol{\theta}}_y) - 2 \sum_{y=1}^k n_y \log(\hat{\varpi}_y) + 2(k-1), \quad (56)$$

$$\begin{aligned} \text{BIC}_{\text{Mix}} &= \sum_{y=1}^k \text{BIC}_y(\hat{\boldsymbol{\theta}}_y) - 2 \sum_{y=1}^k n_y \log(\hat{\varpi}_y) + (k-1) \log n \\ &\quad + \sum_{y=1}^k \dim(\boldsymbol{\theta}_y) (\log n - \log n_y) \end{aligned} \quad (57)$$

ここで，

$$\text{AIC}_y(\hat{\boldsymbol{\theta}}_y) := -2\ell_y(\hat{\boldsymbol{\theta}}_y) + 2\dim(\boldsymbol{\theta}_y) = -2 \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \hat{\boldsymbol{\theta}}_y)) + 2\dim(\boldsymbol{\theta}_y), \quad (58)$$

$$\begin{aligned} \text{BIC}_y(\hat{\boldsymbol{\theta}}_y) &:= -2\ell_y(\hat{\boldsymbol{\theta}}_y) + \dim(\boldsymbol{\theta}_y) \log n_y \\ &= -2 \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \hat{\boldsymbol{\theta}}_y)) + \dim(\boldsymbol{\theta}_y) \log n_y \end{aligned} \quad (59)$$

は，第  $y$  カテゴリ  $C_y$  からのデータ  $\{x_i, i \in \mathcal{L}_y\}$  にもとづく対数尤度  $\ell_y(\hat{\boldsymbol{\theta}}_y)$  を利用した AIC と BIC である。

**証明：**定義 (54), (55) から明らかであるが，簡単に示すと，

$$\text{AIC}_{\text{Mix}} = -2 \left\{ \sum_{y=1}^k n_y \log(\hat{\varpi}_y) + \sum_{y=1}^k \sum_{i \in \mathcal{L}_y} \log(f_y(x_i | \hat{\boldsymbol{\theta}}_y)) \right\} + 2 \left( \sum_{y=1}^k \dim(\boldsymbol{\theta}_y) + k - 1 \right)$$

$$\begin{aligned}
&= \sum_{y=1}^k \text{AIC}_y(\hat{\theta}_y) - 2 \sum_{y=1}^k n_y \log(\hat{\omega}_y) + 2(k-1) \\
\text{BIC}_{\text{Mix}} &= -2 \left\{ \sum_{y=1}^k n_y \log(\hat{\omega}_y) + \sum_{y=1}^k \sum_{i \in L_y} \log(f_y(x_i | \hat{\theta}_y)) \right\} + \left( \sum_{y=1}^k \dim(\theta_y) + k - 1 \right) \log n \\
&= \sum_{y=1}^k \text{BIC}_y(\hat{\theta}_y) - 2 \sum_{y=1}^k n_y \log(\hat{\omega}_y) + (k-1) \log n \\
&\quad + \sum_{y=1}^k \dim(\theta_y) (\log n - \log n_y)
\end{aligned}$$

となる. □

**注意 8** 命題 7 において, (56) と (57) を比較することによって, データ  $\{x_i, i \in L_y\}$  にもとづく対数尤度  $\ell_y(\hat{\theta}_y)$  を利用した情報量規準を利用して有限混合分布に対する情報量規準を求める際に, BIC の場合は, (57) の最終項  $\sum_{y=1}^k \dim(\theta_y) (\log n - \log n_y)$  を考慮する必要があることがわかる.

## 付録 E 変形ベッセル関数

ここでは, 増田 (2002), Olver *et al.* (2010) にしたがって, 変形ベッセル関数に関する定義や性質を与える. なお, ここで与える命題に関する証明については上記の文献を参照されたい.

変形ベッセル微分方程式:

$$z^2 \frac{d^2 w}{dz^2} + z \frac{dw}{dz} - (z^2 + \nu^2) w = 0, z \in \mathbb{C} \quad (60)$$

の解を変形ベッセル関数 (modified Bessel function) と呼ぶ. ここで, この解の一つが,

$$I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k! \Gamma(\nu + k + 1)} \quad (61)$$

で与えられ, 第 1 種変形ベッセル関数 (modified Bessel function of the first kind) という. また, もう一つの解は, 第 2 種変形ベッセル関数 (modified Bessel function of the second kind) と呼ばれ,  $K_\nu(z)$  という記号で表される. 第 2 種変形ベッセル関数  $K_\nu(z)$  は,



$$K_\nu(z) = \frac{1}{2} \pi \frac{I_{-\nu}(z) - I_\nu(z)}{\sin(\nu\pi)} \quad (62)$$

等の幾つかの表現を持っている。様々な文献の中では、第2種変形ベッセル関数は「第3種変形ベッセル関数」(modified Bessel function of the third kind) と呼ばれることがあるので十分注意する必要がある。

**命題 8**  $\nu \in \mathbb{R}$  かつ偏角  $\arg z = 0$  のとき、 $I_\nu(z)$ ,  $K_\nu(z)$  は実数であり、とくに、 $z (= x + i0 = x) \in \mathbb{R}^+ = (0, \infty)$ ,  $\nu \in \mathbb{R}$  のとき、以下の表現を持つ：

$$K_\nu(x) = \frac{1}{2} \int_0^\infty y^{\nu-1} \exp\left\{-\frac{x}{2}\left(y + \frac{1}{y}\right)\right\} dy \quad (63)$$

$$K_\nu(z) = K_{-\nu}(z) \quad (64)$$

**命題 9** (漸化式表現)  $K_\nu(z)$  について以下の漸化式が成り立つ：

$$K_{\nu+1}(z) = \frac{2\nu}{z} K_\nu(z) + K_{\nu-1}(z) \quad (65)$$

$$K_{\nu-1}(z) + K_{\nu+1}(z) = -2K'_\nu(z) \quad (66)$$

$$K'_{\nu-1}(z) = -K_{\nu-1}(z) - \frac{\nu}{z} K_\nu(z) \quad (67)$$

**命題 10**  $n \in \mathbb{N}$  に対して、

$$K_{n+\frac{1}{2}} = \sqrt{\frac{\pi}{2z}} e^{-z} \sum_{j=0}^n \frac{(n+j)!}{(n-j)!j!} (2z)^{-j} \quad (68)$$

が成り立ち、特に、

$$K_{\frac{1}{2}}(z) = K_{-\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}} e^{-z} \quad (69)$$

## 付録 F R スクリプト

本稿で利用した R スクリプトを以下に与える。

## コード2：本稿の全結果を再現するためのRスクリプト：GH-NIG-Modeling.R

```

1 #####
2 ## R パッケージの読み込み
3 #####
4 library(readr)
5 library(dplyr)
6 library(ggplot2)
7 library(xtable)
8 #####
9 ## データ読み込み, ラングリング
10 #####
11 finSPT <- read_csv("../data/finSPT2020.csv")
12 finSPT.sub <- finSPT %>%
13   filter(market_cap > 0, shareholders > 0, net_income > 0, month == 12) %>%
14   mutate(log.mc = log(market_cap), log.sh = log(shareholders), log.ni = log(net_income)) %>%
15   select(firmID, year, year_SPT,
16          market_cap, shareholders, net_income, log.mc, log.sh, log.ni, ESG, E, S, G, type)
17 finSPT.sub$type <- as.factor(finSPT.sub$type)
18 x <- finSPT.sub %>% filter(year == 2018)
19 finSPT.sub.G <- finSPT.sub %>% filter(type == "G")
20 finSPT.sub.E <- finSPT.sub %>% filter(type == "E")
21 #####
22 ## 企業数をカウントする関数
23 #####
24 count.obs <- function(){
25   require(dplyr)
26   a.T <- finSPT.sub %>% group_by(year) %>% summarise(n = n())
27   a.G <- finSPT.sub %>% filter(type == "G") %>% group_by(year) %>% summarise(G = n())
28   a.E <- finSPT.sub %>% filter(type == "E") %>% group_by(year) %>% summarise(E = n())
29   a <- a.G %>% left_join(a.E)
30   a <- a %>% mutate(year_SPT = seq(2015, 2020)) %>% select(year, year_SPT, G, E, n)
31   print(a)
32 }
33 }
34 #####
35 ## 欠測値を除去した企業数をカウントする関数
36 #####
37 count.obs.omit.NA <- function(){
38   require(dplyr)
39   a.T <- finSPT.sub %>% group_by(year) %>% na.omit() %>% summarise(n = n())
40   a.G <- finSPT.sub %>% filter(type == "G") %>% na.omit() %>% group_by(year) %>% summarise(G = n())
41   a.E <- finSPT.sub %>% filter(type == "E") %>% na.omit() %>% group_by(year) %>% summarise(E = n())
42   a <- a.G %>% left_join(a.E)
43   a <- a %>% left_join(a.T)
44   a <- a %>% mutate(year_SPT = seq(2015, 2020)) %>% select(year, year_SPT, G, E, n)
45   print(a)
46 }
47 }
48 #####
49 ## 対散布図を描くための関数
50 #####
51 finSPT.ggpairs <- function(df = finSPT.sub, yr = 2018, type = "raw")
52 {
53   require(GGally)
54   require(dplyr)
55   require(ggplot2)
56   if(type == "raw")
57   {
58     p <- df %>% filter(year == yr) %>%
59       select(market_cap, shareholders, net_income, ESG, E, S, G, type) %>%

```

```

59   ggpairs(upper = list(continuous = wrap("points", size = 0.1, alpha = 0.2), combo = "box"),
60           lower = list(continuous = wrap("cor", size=2), combo = wrap("facetdensity")),
61           mapping = aes(color = type),
62           diag = list(continuous = wrap("densityDiag", alpha = 0.5))) +
63   ggtitle(paste("Fiscal_Year:", as.character(yr))) +
64   theme(axis.text= element_text(size = 5),
65         legend.title = element_text(size = 7),
66         legend.text = element_text(size = 7),
67         axis.title = element_text(size = 7),
68         plot.title = element_text(size = 7),
69         strip.text = element_text(size = 7))
70   }
71   if(type == "log")
72   {
73     p <- df %>% filter(year == yr) %>%
74     select(log.mc, log.sh, log.ni, ESG, E, S, G, type) %>%
75     ggpairs(upper = list(continuous = wrap("points", size = 0.1, alpha = 0.2), combo = "box"),
76           lower = list(continuous = wrap("cor", size=2), combo = wrap("facetdensity")),
77           mapping = aes(color = type),
78           diag = list(continuous = wrap("densityDiag", alpha = 0.5))) +
79     ggtitle(paste("Fiscal_Year:", as.character(yr))) +
80     theme(axis.text = element_text(size = 5),
81           legend.title = element_text(size = 7),
82           legend.text = element_text(size = 7),
83           axis.title = element_text(size = 7),
84           plot.title = element_text(size = 7),
85           strip.text = element_text(size = 7))
86   }
87   if(type == "both")
88   {
89     p <- df %>% filter(year == yr) %>%
90     select(log.mc, log.sh, log.ni, market_cap, shareholders, net_income, ESG, E, S, G, type)
91     %>%
92     ggpairs(upper = list(continuous = wrap("points", size = 0.1, alpha = 0.2), combo = "box"),
93           lower = list(continuous = wrap("cor", size=2), combo = wrap("facetdensity")),
94           mapping = aes(color = type),
95           diag = list(continuous = wrap("densityDiag", alpha = 0.5))) +
96     ggtitle(paste("Fiscal_Year:", as.character(yr))) +
97     theme(axis.text= element_text(size = 5),
98           legend.title = element_text(size = 7),
99           legend.text = element_text(size = 7),
100          axis.title = element_text(size = 7),
101          plot.title = element_text(size = 7),
102          strip.text = element_text(size = 7))
103   }
104   print(p)
105   }
106   #####
107   ### 2018年度の株式時価総額に分布の当てはめ
108   #####
109   # 非対称正規分布
110   library(sn)
111   fitted.log.mc.SN <- sn::selm(log.mc ~ 1, data = x)
112   coef.fitted.log.mc.SN <- coef(fitted.log.mc.SN, param.type = "DP")
113   # 非対称テイー分布
114   fitted.log.mc.ST <- sn::selm(log.mc ~ 1, family = "ST", data = x)
115   coef.fitted.log.mc.ST <- coef(fitted.log.mc.ST, param.type = "DP")
116   #
117   BIC.SN <- function(obj) unname(-2*obj@logL + obj@size[3]*log(obj@size[4]))
118   BIC.ST <- function(obj) unname(-2*obj@logL + obj@size[3]*log(obj@size[4]))
119   #

```

```

119 library("GeneralizedHyperbolic")
120 fitted.log.mc.NIG <- nigFit(x$log.mc)
121 AIC.NIG <- function(obj) -2*obj$maxLik + 2*length(obj$params)
122 BIC.NIG <- function(obj) -2*obj$maxLik + length(obj$params)*log(length(obj$obs))
123 #
124 fitted.log.mc.SN.G <- x %>% filter(type == "G") %>% sn::selm(log.mc ~ 1, family = "SN", data = .)
125 fitted.log.mc.SN.E <- x %>% filter(type == "E") %>% sn::selm(log.mc ~ 1, family = "SN", data = .)
126 #
127 fitted.log.mc.ST.G <- x %>% filter(type == "G") %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
128 fitted.log.mc.ST.E <- x %>% filter(type == "E") %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
129 #
130 fitted.log.mc.NIG.G <- x %>% filter(type == "G") %>% pull(log.mc) %>% nigFit()
131 fitted.log.mc.NIG.E <- x %>% filter(type == "E") %>% pull(log.mc) %>% nigFit()
132 #####
133 ### 対数分布のプロット (2018会計年度)
134 #####
135 finSPT.ggpairs(type = "log")
136 #####
137 ### 対数株式時価総額ヒストグラムプロット (2018会計年度)
138 #####
139 p.log.mc <- x %>% ggplot(aes(log.mc))
140 p.log.mc +
141   geom_histogram(aes(y = ..density..), binwidth = 0.5, alpha = 0.5) +
142   geom_density() +
143   geom_rug() + coord_cartesian(ylim=c(0, 0.4))
144 #####
145 ### 対数株式時価総額正規Q-Qプロット (2018会計年度)
146 #####
147 x %>% select(market_cap) %>%
148   ggplot(aes(sample = log(market_cap))) + geom_qq() + geom_qq_line()
149 #####
150 ### 対数株式時価総額歪度 (2018会計年度)
151 #####
152 library(e1071)
153 skewness.log.mc2018 <- x %>% pull(log.mc) %>% skewness()
154 #####
155 ### 対数株式時価総額(国タイプ別)ヒストグラムプロット (2018会計年度)
156 #####
157 p.log.mc + geom_histogram(aes(y = ..density.., fill = type), position = "identity", binwidth =
158   0.5, alpha = 0.3) +
159   geom_density(aes(color = type, linetype = type)) +
160   geom_rug(aes(color = type)) +
161   # scale_colour_grey() +
162   coord_cartesian(ylim=c(0, 0.5)) +
163   theme(legend.position="top")
164 #####
165 ### 対数株式時価総額(国タイプ別)正規Q-Qプロット (2018会計年度)
166 #####
167 x %>% ggplot(aes(sample = log.mc, color = type, linetype = type)) +
168   geom_qq() + geom_qq_line() + theme(legend.position="top")
169 #####
170 ### 対数株式時価総額ヒストグラムと非対称正規分布の当てはめ (2018会計年度)
171 #####
172 plot(fitted.log.mc.SN, which = 2, ylim = c(0, 0.4))
173 #####
174 ### 非対称正規分布を当てはめたとときのP-Pプロット (2018会計年度)
175 #####
176 plot(fitted.log.mc.SN, which = 4)
177 #####
178 ### 対数株式時価総額ヒストグラムと非対称テーパー分布の当てはめ (2018会計年度)
179 #####

```

```

179 plot(fitted.log.mc.ST, which = 2, ylim = c(0, 0.4))
180 #####
181 ### 非対称ティ-分布を当てはめたときの P-P プロット (2018 会計度)
182 #####
183 plot(fitted.log.mc.ST, which=4)
184 #####
185 ### 対数株式時価総額のヒストグラムと正規逆ガウス分布の当てはめ (2018 会計度)
186 #####
187 fitted.log.mc.NIG %>%
188 plot(which = 1, plotTitles = c("Histogram_of_log.mc", "Log-Histogram_of_log.mc", "Q-Q_Plot_of_log
    .mc", "P-P_Plot_of_log.mc"))
189 #####
190 ### 正規逆ガウス分布を当てはめたときの P-P プロット (2018 会計度)
191 #####
192 fitted.log.mc.NIG %>%
193 plot(which = 4, plotTitles = c("Histogram_of_log.mc", "Log-Histogram_of_log.mc", "Q-Q_Plot_of_log
    .mc", "P-P_Plot_of_log.mc"))
194 #####
195 ### AIC, BIC 表: 対数株式時価総額への分布当てはめに関する情報量規準の比較
196 #####
197 AIC.log.mc <- AIC(fitted.log.mc.SN, fitted.log.mc.ST)
198 tab.AIC.log.mc <- rbind(AIC.log.mc, c(length(fitted.log.mc.NIG$param), AIC.NIG(fitted.log.mc.NIG
    )))
199 colnames(tab.AIC.log.mc) <- c("dim", "AIC")
200 rownames(tab.AIC.log.mc) <- c("SN", "ST", "NIG")
201 #
202 BIC.log.mc <- c(BIC.SN(fitted.log.mc.SN), BIC.ST(fitted.log.mc.ST), BIC.NIG(fitted.log.mc.NIG))
203 dim.log.mc <- c(fitted.log.mc.SN@size[3], fitted.log.mc.ST@size[3], length(fitted.log.mc.NIG$
    param))
204 tab.BIC.log.mc <- cbind(dim.log.mc, BIC.log.mc)
205 colnames(tab.BIC.log.mc) <- c("dim", "BIC")
206 rownames(tab.BIC.log.mc) <- c("SN", "ST", "NIG")
207 #
208 tab.AIC.BIC.log.mc <- cbind(tab.AIC.log.mc, BIC.log.mc)
209 colnames(tab.AIC.BIC.log.mc) <- c("dim", "AIC", "BIC")
210 rownames(tab.AIC.BIC.log.mc) <- c("SN", "ST", "NIG")
211 #####
212 ### AIC, BIC 表の LaTeX 出力
213 #####
214 print(
215 xtable(tab.AIC.BIC.log.mc,
216 digits = c(0,0,2,2),
217 floating = FALSE,
218 caption = c("AIC_BIC_表: 対数株式時価総額への分布当てはめに関する情報量規準の比較"),
219 label = "table.distribution.AIC.BIC"),
220 caption.placement = "top",
221 table.placement = "H")
222 #####
223 ### 先進国の対数株式時価総額のヒストグラムと非対称正規分布の当てはめ (2018 会計度)
224 #####
225 plot(fitted.log.mc.SN.G, which = 2)
226 #####
227 ### 先進国の対数株式時価総額へ非対称正規分布を当てはめたときの P-P プロット (2018 会計度)
228 #####
229 plot(fitted.log.mc.SN, which = 4)
230 #####
231 ### 新興国の対数株式時価総額のヒストグラムと非対称正規分布の当てはめ (2018 会計度)
232 #####
233 plot(fitted.log.mc.SN.E, which = 2)
234 #####
235 ### 新興国の対数株式時価総額へ非対称正規分布を当てはめたときの P-P プロット (2018 会計度)

```

```

236 #####
237 plot(fitted.log.mc.SN, which = 4)
238 #####
239 ### 先進国の対数株式時価総額のヒストグラムと非対称ティータ分布の当てはめ (2018 会計度)
240 #####
241 plot(fitted.log.mc.ST.G, which = 2, ylim = c(0, 0.4))
242 #####
243 ### 先進国の対数株式時価総額へ非対称ティータ分布を当てはめたときの P-P プロット (2018 会計度)
244 #####
245 plot(fitted.log.mc.ST.G, which=4)
246 #####
247 ### 新興国の対数株式時価総額のヒストグラムと非対称ティータ分布の当てはめ (2018 会計度)
248 #####
249 plot(fitted.log.mc.ST.E, which = 2, ylim = c(0, 0.4))
250 #####
251 ### 新興国の対数株式時価総額へ非対称ティータ分布を当てはめたときの P-P プロット (2018 会計度)
252 #####
253 plot(fitted.log.mc.ST.E, which=4)
254 #####
255 ### 先進国の対数株式時価総額のヒストグラムと正規逆ガウス分布の当てはめ (2018 会計度)
256 #####
257 fitted.log.mc.NIG.G %>%
258 plot(which = 1, plotTitles = c("Histogram_of_log.mc_(G)", "Log-Histogram_of_log.mc_(G)", "Q-Q_
  Plot_of_log.mc_(G)", "P-P_Plot_of_log.mc_(G)"))
259 #####
260 ### 先進国の対数株式時価総額へ正規逆ガウス分布を当てはめたときの P-P プロット (2018 会計度)
261 #####
262 fitted.log.mc.NIG.G %>%
263 plot(which = 4, plotTitles = c("Histogram_of_log.mc_(G)", "Log-Histogram_of_log.mc_(G)", "Q-Q_
  Plot_of_log.mc_(G)", "P-P_Plot_of_log.mc_(G)"))
264 #####
265 ### 新興国の対数株式時価総額のヒストグラムと正規逆ガウス分布の当てはめ (2018 会計度)
266 #####
267 fitted.log.mc.NIG.E %>%
268 plot(which = 1, plotTitles = c("Histogram_of_log.mc_(E)", "Log-Histogram_of_log.mc_(E)", "Q-Q_
  Plot_of_log.mc_(E)", "P-P_Plot_of_log.mc_(E)"))
269 #####
270 ### 新興国の対数株式時価総額へ正規逆ガウス分布を当てはめたときの P-P プロット (2018 会計度)
271 #####
272 fitted.log.mc.NIG.E %>%
273 plot(which = 4, plotTitles = c("Histogram_of_log.mc_(E)", "Log-Histogram_of_log.mc_(E)", "Q-Q_
  Plot_of_log.mc_(E)", "P-P_Plot_of_log.mc_(E)"))
274 #####
275 ### AIC, BIC 表: 対数株式時価総額 (国タイプ別) への分布当てはめに関する情報量規準の比較
276 #####
277 AIC.log.mc.G <- AIC(fitted.log.mc.SN.G, fitted.log.mc.ST.G)
278 tab.AIC.log.mc.G <- rbind(AIC.log.mc.G, c(length(fitted.log.mc.NIG.G$param), AIC.NIG(fitted.log.
  mc.NIG.G)))
279 colnames(tab.AIC.log.mc.G) <- c("dim", "AIC")
280 rownames(tab.AIC.log.mc.G) <- c("SN", "ST", "NIG")
281 #
282 BIC.log.mc.G <- c(BIC.SN(fitted.log.mc.SN.G), BIC.ST(fitted.log.mc.ST.G), BIC.NIG(fitted.log.mc.
  NIG.G))
283 dim.log.mc.G <- c(fitted.log.mc.SN.G@size[3], fitted.log.mc.ST.G@size[3], length(fitted.log.mc.
  NIG.G$param))
284 tab.BIC.log.mc.G <- cbind(dim.log.mc.G, BIC.log.mc.G)
285 colnames(tab.BIC.log.mc.G) <- c("dim", "BIC")
286 rownames(tab.BIC.log.mc.G) <- c("SN", "ST", "NIG")
287 #
288 AIC.log.mc.E <- AIC(fitted.log.mc.SN.E, fitted.log.mc.ST.E)
289 tab.AIC.log.mc.E <- rbind(AIC.log.mc.E, c(length(fitted.log.mc.NIG.E$param), AIC.NIG(fitted.log.

```

```

      mc.NIG.E))
290 colnames(tab.AIC.log.mc.E) <- c("dim", "AIC")
291 rownames(tab.AIC.log.mc.E) <- c("Skew-Normal", "Skew-T", "Normal_Inverse_Gaussian")
292 #
293 BIC.log.mc.E <- c(BIC.SN(fitted.log.mc.SN.E), BIC.ST(fitted.log.mc.ST.E), BIC.NIG(fitted.log.mc.
      NIG.E))
294 dim.log.mc.E <- c(fitted.log.mc.SN.E@size[3], fitted.log.mc.ST.E@size[3], length(fitted.log.mc.
      NIG.E$param))
295 tab.BIC.log.mc.E <- cbind(dim.log.mc.E, BIC.log.mc.E)
296 colnames(tab.BIC.log.mc.E) <- c("dim", "BIC")
297 rownames(tab.BIC.log.mc.E) <- c("Skew-Normal", "Skew-T", "Normal_Inverse_Gaussian")
298 #
299 tab.AIC.BIC.log.mc.G <- cbind(tab.AIC.log.mc.G, BIC.log.mc.G)
300 colnames(tab.AIC.BIC.log.mc.G) <- c("dim", "AIC", "BIC")
301 rownames(tab.AIC.BIC.log.mc.G) <- c("SN", "ST", "NIG")
302 #
303 tab.AIC.BIC.log.mc.E <- cbind(tab.AIC.log.mc.E, BIC.log.mc.E)
304 colnames(tab.AIC.BIC.log.mc.E) <- c("dim", "AIC", "BIC")
305 rownames(tab.AIC.BIC.log.mc.E) <- c("SN", "ST", "NIG")
306 #####
307 ### AIC, BIC 表 (先進国) の LaTeX 出力
308 #####
309 print(
310   xtable(tab.AIC.BIC.log.mc.G,
311     digits = c(0,0,2,2),
312     floating = FALSE,
313     caption = c("AIC, BIC_表: 先進国の企業の対数株式時価総額への分布当てはめに関する情報量規準の比較"
314       ),
315     label = "table.distribution.AIC.BIC.G"),
316   caption.placement = "top",
317   table.placement = "H")
318 #####
319 ### AIC, BIC 表 (新興国) の LaTeX 出力
320 #####
321 print(
322   xtable(tab.AIC.BIC.log.mc.E,
323     digits = c(0,0,2,2),
324     floating = FALSE,
325     caption = c("AIC, BIC_表: 新興国の企業の対数株式時価総額への分布当てはめに関する情報量規準の比較"
326       ),
327     label = "table.distribution.AIC.BIC.E"),
328   caption.placement = "top",
329   table.placement = "H")
330 #####
331 ### 混合比率の推定値の計算
332 #####
333 obs <- x %>% group_by(type) %>% summarise(obs = n()) %>% pull(obs)
334 n <- sum(obs)
335 nG <- obs[2]
336 nE <- obs[1]
337 nGE <- c(nG, nE)
338 varpi <- c(nG/n, nE/n)
339 #####
340 ### 有限混合分布当てはめに関する情報量規準を計算する関数
341 #####
342 IC.GE <- function(x)
343 {
344   require(tidyverse)
345   require(sn)
346   require(GeneralizedHyperbolic)
347   obs <- x %>% group_by(type) %>% summarise(obs = n()) %>% pull(obs)

```

```

346 n <- sum(obs)
347 nG <- obs[2]
348 nE <- obs[1]
349 nGE <- c(nG, nE)
350 varpi <- c(nG/n, nE/n)
351 #
352 SN <- x %>% selm(log.mc ~ 1, family = "SN", data = .)
353 ST <- x %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
354 NIG <- x %>% pull(log.mc) %>% nigFit()
355 #
356 SN.G <- x %>% filter(type == "G") %>% selm(log.mc ~ 1, family = "SN", data = .)
357 SN.E <- x %>% filter(type == "E") %>% selm(log.mc ~ 1, family = "SN", data = .)
358 ST.G <- x %>% filter(type == "G") %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
359 ST.E <- x %>% filter(type == "E") %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
360 NIG.G <- x %>% filter(type == "G") %>% pull(log.mc) %>% nigFit()
361 NIG.E <- x %>% filter(type == "E") %>% pull(log.mc) %>% nigFit()
362 #
363 SN.maxLL <- SN@logL
364 ST.maxLL <- ST@logL
365 NIG.maxLL <- NIG$maxLik
366 #
367 SN.G.maxLL <- SN.G@logL
368 SN.E.maxLL <- SN.E@logL
369 ST.G.maxLL <- ST.G@logL
370 ST.E.maxLL <- ST.E@logL
371 NIG.G.maxLL <- NIG.G$maxLik
372 NIG.E.maxLL <- NIG.E$maxLik
373 #
374 AIC.SN <- -2 * SN.maxLL + 2 * SN@size[3]
375 AIC.ST <- -2 * ST.maxLL + 2 * ST@size[3]
376 AIC.NIG <- -2 * NIG.maxLL + 2 * length(NIG$params)
377 #
378 BIC.SN <- -2 * SN.maxLL + SN@size[3]*log(n)
379 BIC.ST <- -2 * ST.maxLL + ST@size[3]*log(n)
380 BIC.NIG <- -2 * NIG.maxLL + length(NIG$params)*log(n)
381 #
382 AIC.SN.G <- -2 * SN.G.maxLL + 2 * SN.G@size[3]
383 AIC.SN.E <- -2 * SN.E.maxLL + 2 * SN.E@size[3]
384 AIC.ST.G <- -2 * ST.G.maxLL + 2 * ST.G@size[3]
385 AIC.ST.E <- -2 * ST.E.maxLL + 2 * ST.E@size[3]
386 AIC.NIG.G <- -2 * NIG.G.maxLL + 2 * length(NIG.G$params)
387 AIC.NIG.E <- -2 * NIG.E.maxLL + 2 * length(NIG.E$params)
388 #
389 BIC.SN.G <- -2 * SN.G.maxLL + SN.G@size[3]*log(nG)
390 BIC.SN.E <- -2 * SN.E.maxLL + SN.E@size[3]*log(nE)
391 BIC.ST.G <- -2 * ST.G.maxLL + ST.G@size[3]*log(nG)
392 BIC.ST.E <- -2 * ST.E.maxLL + ST.E@size[3]*log(nE)
393 BIC.NIG.G <- -2 * NIG.G.maxLL + length(NIG.G$params)*log(nG)
394 BIC.NIG.E <- -2 * NIG.E.maxLL + length(NIG.E$params)*log(nE)
395 #
396 tab.maxLL <- data.frame(dst.type = c("NIG.G", "NIG.E", "SN.G", "ST.E"), maxLL = c(NIG.G.maxLL,
  NIG.E.maxLL, SN.G.maxLL, ST.E.maxLL))
397 # AIC, BIC of NIG.G.NIG.E
398 AIC.NIG.G.NIG.E <- -2 * (sum(nGE * log(varpi)) + NIG.G.maxLL + NIG.E.maxLL) + 2 * (length(NIG.G
  $param) + length(NIG.E$params) + 1)
399 BIC.NIG.G.NIG.E <- -2 * (sum(nGE * log(varpi)) + NIG.G.maxLL + NIG.E.maxLL) + (length(NIG.G
  $param) + length(NIG.E$params) + 1) * log(n)
400 # AIC, BIC of NIG.G.ST.E
401 AIC.NIG.G.ST.E <- -2 * (sum(nGE * log(varpi)) + NIG.G.maxLL + ST.E.maxLL) + 2 * (length(NIG.G$
  param) + ST.E@size[3] + 1)
402 BIC.NIG.G.ST.E <- -2 * (sum(nGE * log(varpi)) + NIG.G.maxLL + ST.E.maxLL) + (length(NIG.G$

```



```

      param) + ST.E@size[3] + 1) * log(n)
403 # AIC, BIC of SN.G.NIG.E
404 AIC.SN.G.NIG.E <- -2 * (sum(nGE * log(varpi)) + SN.G.maxLL + NIG.E.maxLL) + 2 * (SN.G@size[3] +
      length(NIG.E$param) + 1)
405 BIC.SN.G.NIG.E <- -2 * (sum(nGE * log(varpi)) + SN.G.maxLL + NIG.E.maxLL) + (SN.G@size[3] +
      length(NIG.E$param) + 1) * log(n)
406 # AIC, BIC of SN.G.ST.E
407 AIC.SN.G.ST.E <- -2 * (sum(nGE * log(varpi)) + SN.G.maxLL + ST.E.maxLL) + 2 * (SN.G@size[3] +
      ST.E@size[3] + 1)
408 BIC.SN.G.ST.E <- -2 * (sum(nGE * log(varpi)) + SN.G.maxLL + ST.E.maxLL) + (SN.G@size[3] +
      ST.E@size[3] + 1) * log(n)
409 #
410 G <- c("NIG", "NIG", "SN", "SN")
411 E <- c("NIG", "ST", "NIG", "ST")
412 dimension <- c(length(NIG.G$param) + length(NIG.E$param) + 1,
413               length(NIG.G$param) + ST.E@size[3] + 1,
414               SN.G@size[3] + length(NIG.E$param) + 1,
415               SN.G@size[3] + ST.E@size[3] + 1)
416 AIC.vec <- c(AIC.NIG.G.NIG.E, AIC.NIG.G.ST.E, AIC.SN.G.NIG.E, AIC.SN.G.ST.E)
417 BIC.vec <- c(BIC.NIG.G.NIG.E, BIC.NIG.G.ST.E, BIC.SN.G.NIG.E, BIC.SN.G.ST.E)
418 tab.AIC.BIC.mix <- data.frame(G, E, dimension, AIC.vec, BIC.vec)
419 colnames(tab.AIC.BIC.mix) <- c("G", "E", "df", "AIC", "BIC")
420 #
421 tab.AIC.BIC <- data.frame(dim = c(SN@size[3], ST@size[3], length(NIG$param)),
422                          AIC = c(AIC.SN, AIC.ST, AIC.NIG),
423                          BIC = c(BIC.SN, BIC.ST, BIC.NIG))
424 colnames(tab.AIC.BIC) <- c("dim", "AIC", "BIC")
425 rownames(tab.AIC.BIC) <- c("SN", "ST", "NIG")
426 #
427 tab.AIC.BIC.G <- data.frame(dim = c(SN.G@size[3], ST.G@size[3], length(NIG.G$param)),
428                          AIC.G = c(AIC.SN.G, AIC.ST.G, AIC.NIG.G),
429                          BIC.G = c(BIC.SN.G, BIC.ST.G, BIC.NIG.G))
430 colnames(tab.AIC.BIC.G) <- c("dim", "AIC", "BIC")
431 rownames(tab.AIC.BIC.G) <- c("SN.G", "ST.G", "NIG.G")
432 #
433 tab.AIC.BIC.E <- data.frame(dim = c(SN.E@size[3], ST.E@size[3], length(NIG.E$param)),
434                          AIC.E = c(AIC.SN.E, AIC.ST.E, AIC.NIG.E),
435                          BIC.E = c(BIC.SN.E, BIC.ST.E, BIC.NIG.E))
436 colnames(tab.AIC.BIC.E) <- c("dim", "AIC", "BIC")
437 rownames(tab.AIC.BIC.E) <- c("SN.E", "ST.E", "NIG.E")
438 #
439 list(
440   nGE = nGE, varpi = varpi,
441   log.nGE.varpi = nGE*log(varpi),
442   tab.maxLL = tab.maxLL,
443   tab.AIC.BIC.mix = tab.AIC.BIC.mix,
444   tab.AIC.BIC = tab.AIC.BIC,
445   tab.AIC.BIC.G = tab.AIC.BIC.G,
446   tab.AIC.BIC.E = tab.AIC.BIC.E
447 )
448 }
449 tmp <- IC.GE(x)
450 #####
451 ### AIC, BIC 表: 対数株式時価総額への有限混合分布の当てはめに関する情報量規準
452 #####
453 print(
454   xtable(tmp$tab.AIC.BIC.mix,
455           digits = c(0,0,0,2,2),
456           floating = FALSE,
457           caption = c("AIC_ BIC_ 表: 対数株式時価総額への混合分布の当てはめに関する情報量規準の比較"),
458           label = "table.mixture.distribution.AIC.BIC"),

```

```

459 caption.placement = "top",
460 table.placement = "H")
461 #####
462 ## 有限混合分布を当てはめるための関数
463 #####
464 plot.modelfitting.GE <- function(x)
465 {
466   require(tidyverse)
467   require(sn)
468   require(GeneralizedHyperbolic)
469   obs <- x %>% group_by(type) %>% summarise(obs = n()) %>% pull(obs)
470   n <- sum(obs)
471   nG <- obs[2]
472   nE <- obs[1]
473   nGE <- c(nG, nE)
474   varpi <- c(nG/n, nE/n)
475   #
476   SN <- x %>% selm(log.mc ~ 1, family = "SN", data = .)
477   SN.G <- x %>% filter(type == "G") %>% selm(log.mc ~ 1, family = "SN", data = .)
478   coef.SN.G <- coef(SN.G, param.type = "DP")
479   SN.E <- x %>% filter(type == "E") %>% selm(log.mc ~ 1, family = "SN", data = .)
480   coef.SN.E <- coef(SN.E, param.type = "DP")
481   ST <- x %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
482   coef.ST <- coef(ST, param.type = "DP")
483   ST.G <- x %>% filter(type == "G") %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
484   coef.ST.G <- coef(ST.G, param.type = "DP")
485   ST.E <- x %>% filter(type == "E") %>% sn::selm(log.mc ~ 1, family = "ST", data = .)
486   coef.ST.E <- coef(ST.E, param.type = "DP")
487   NIG <- x %>% pull(log.mc) %>% nigFit()
488   NIG.G <- x %>% filter(type == "G") %>% pull(log.mc) %>% nigFit()
489   NIG.E <- x %>% filter(type == "E") %>% pull(log.mc) %>% nigFit()
490   #
491   pdf.mixture.NIG.G.NIG.E <- function(x, varpiG, varpiE, muG, deltaG, alphaG, betaG, muE, deltaE,
    alphaE, betaE)
492   {
493     varpiG * dnic(x, mu = muG, delta = deltaG, alpha = alphaG, beta = betaG) + varpiE * dnic(x,
    mu = muE, delta = deltaE, alpha = alphaE, beta = betaE)
494   }
495   #
496   pdf.mixture.NIG.G.ST.E <- function(x, varpiG, varpiE, muG, deltaG, alphaG, betaG, muE, xiE,
    omegaE, alphaE, nuE)
497   {
498     varpiG * dnic(x, mu = muG, delta = deltaG, alpha = alphaG, beta = betaG) + varpiE * dst(x, xi
    = xiE, omega = omegaE, alpha = alphaE, nu = nuE)
499   }
500   #
501   pdf.mixture.SN.G.NIG.E <- function(x, varpiG, varpiE, xiG, omegaG, alphaG, muE, deltaE, alphaE,
    betaE)
502   {
503     varpiG * dsn(x, xi = xiG, omega = omegaG, alpha = alphaG) + varpiE * dnic(x, mu = muE, delta
    = deltaE, alpha = alphaE, beta = betaE)
504   }
505   #
506   pdf.mixture.SN.G.ST.E <- function(x, varpiG, varpiE, xiG, omegaG, alphaG, xiE, omegaE, alphaE,
    nuE)
507   {
508     varpiG * dsn(x, xi = xiG, omega = omegaG, alpha = alphaG) + varpiE * dst(x, xi = xiE, omega =
    omegaE, alpha = alphaE, nu = nuE)
509   }
510   #
511   p.log.mc <- x %>% ggplot(aes(log.mc))

```

```

512 p <- p.log.mc +
513   geom_histogram(aes(y=.density..), alpha = 0.5) +
514   geom_rug() +
515   theme(legend.position="top") +
516   #geom_density(alpha = 0.8) +
517   coord_cartesian(ylim=c(0, 0.4)) +
518   stat_function(fun = dst,
519                 args = list(xi = coef.ST[1], omega = coef.ST[2], alpha = coef.ST[3], nu =
520                             coef.ST[4]),
521                 aes(linetype = "ST"))
522 ) +
523 stat_function(fun = pdf.mixture.NIG.G.NIG.E,
524               args = list(varpiG = varpi[1], varpiE = varpi[2],
525                             muG = coef(NIG.G)[1], deltaG = coef(NIG.G)[2], alphaG = coef(NIG.G
526                               ) [3], betaG = coef(NIG.G)[4],
527                             muE = coef(NIG.E)[1], deltaE = coef(NIG.E)[2], alphaE = coef(NIG.E
528                               ) [3], betaE = coef(NIG.E)[4]),
529               aes(linetype = "NIG.G.NIG.E"))
530 ) +
531 stat_function(fun = pdf.mixture.SN.G.NIG.E,
532               args = list(varpiG = varpi[1], varpiE = varpi[2],
533                             xiG = coef.SN.G[1], omegaG = coef.SN.G[2], alphaG = coef.SN.G[3],
534                             muE = coef(NIG.E)[1], deltaE = coef(NIG.E)[2], alphaE = coef(NIG.
535                               E)[3], betaE = coef(NIG.E)[4]),
536               aes(linetype = "SN.G.NIG.E"))
537 ) +
538 scale_color_manual("Distributions", values = c("blue", "green", "red"))
539 plot(p)
540 }
541 #####
542 ### 有限混合分布の当てはめ結果のプロット
543 #####
544 plot.modelfitting.GE(x)
545 #####
546 ### 情報量規準 (AIC, BIC) の経年変化のプロット
547 #####
548 IC.ts <- function(obj)
549 {
550   require(sn)
551   require(tidyverse)
552   require(lubridate)
553   require(GeneralizedHyperbolic)
554   #
555   n_fun <- function(res) as.integer(res@size[4])
556   #
557   AIC_SN_fun <- function(obj) AIC(obj)
558   AIC_ST_fun <- function(obj) AIC(obj)
559   AIC_NIG_fun <- function(obj) -2*obj$maxLik + 2*length(obj$param)
560   #
561   BIC_SN_fun <- function(obj) unname(-2*obj@logL + obj@size[3]*log(obj@size[4]))
562   BIC_ST_fun <- function(obj) unname(-2*obj@logL + obj@size[3]*log(obj@size[4]))
563   BIC_NIG_fun <- function(obj) -2*obj$maxLik + length(obj$param)*log(length(obj$obs))
564   #
565   mod_SN_fun <- function(df) selm(log.mc ~ 1, family = "SN", data = df)
566   mod_ST_fun <- function(df) selm(log.mc ~ 1, family = "ST", data = df)
567   mod_NIG_fun <- function(df) nigFit(df$log.mc)
568   #
569   res <- obj %>% group_by(year) %>%
570     arrange(year) %>%
571     nest() %>%
572     mutate(

```

```

569     SN_rst = map(data, mod_SN_fun),
570     ST_rst = map(data, mod_ST_fun),
571     NIG_rst = map(data, mod_NIG_fun)
572   ) %>%
573   transmute(year,
574             n_obs = map_int(SN_rst, n_fun),
575             AIC_SN = map_dbl(SN_rst, AIC_SN_fun),
576             AIC_ST = map_dbl(ST_rst, AIC_ST_fun),
577             AIC_NIG = map_dbl(NIG_rst, AIC_NIG_fun),
578             BIC_SN = map_dbl(SN_rst, BIC_SN_fun),
579             BIC_ST = map_dbl(ST_rst, BIC_ST_fun),
580             BIC_NIG = map_dbl(NIG_rst, BIC_NIG_fun)
581   )
582   data.frame(res)
583 }
584 suppressWarnings(invisible(capture.output(finSPT.sub.IC.ts <- IC.ts(finSPT.sub))))
585 suppressWarnings(invisible(capture.output(finSPT.sub.G.IC.ts <- IC.ts(finSPT.sub.G))))
586 suppressWarnings(invisible(capture.output(finSPT.sub.E.IC.ts <- IC.ts(finSPT.sub.E))))
587 plot.AIC.ts <- function(obj)
588 {
589   require(tidyverse)
590   require(gridExtra)
591   x <- obj %>% select(year, n_obs, AIC_SN, AIC_ST, AIC_NIG) %>%
592     pivot_longer(-year, names_to = "model", values_to = "value")
593   y <- obj %>% mutate(AIC_SN_n = AIC_SN/n_obs, AIC_ST_n = AIC_ST/n_obs, AIC_NIG_n = AIC_NIG/n_obs)
594     %>%
595     select(year, n_obs, AIC_SN_n, AIC_ST_n, AIC_NIG_n) %>%
596     pivot_longer(-year, names_to = "model", values_to = "value")
597   p1 <- x %>% filter(model == "n_obs") %>%
598     ggplot(aes(year, value, group = model, shape = model)) +
599     geom_line(linetype="dotted") + geom_point() + ggtitle("n") +
600     theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
601     element_rect(fill = NA, colour = NA))
602   p2 <- x %>% filter(model != "n_obs") %>%
603     ggplot(aes(year, value, group = model, color = model, shape = model)) +
604     geom_line(linetype="dotted") + geom_point() + ggtitle("AIC") +
605     theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
606     element_rect(fill = NA, colour = NA))
607   p3 <- y %>% filter(model != "n_obs") %>%
608     ggplot(aes(year, value, group = model, color = model, shape = model)) +
609     geom_line(linetype="dotted") + geom_point() + ggtitle("AIC/n") +
610     theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
611     element_rect(fill = NA, colour = NA))
612   grid.arrange(p1, p2, p3)
613 }
614 plot.BIC.ts <- function(obj)
615 {
616   require(tidyverse)
617   require(gridExtra)
618   x <- obj %>% select(year, n_obs, BIC_SN, BIC_ST, BIC_NIG) %>%
619     pivot_longer(-year, names_to = "model", values_to = "value")
620   y <- obj %>% mutate(BIC_SN_n = BIC_SN/n_obs, BIC_ST_n = BIC_ST/n_obs, BIC_NIG_n = BIC_NIG /n_obs)
621     %>%
622     select(year, n_obs, BIC_SN_n, BIC_ST_n, BIC_NIG_n) %>%
623     pivot_longer(-year, names_to = "model", values_to = "value")
624   p1 <- x %>% filter(model == "n_obs") %>%
625     ggplot(aes(year, value, group = model, shape = model)) +
626     geom_line(linetype="dotted") + geom_point() + ggtitle("n") +
627     theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
628     element_rect(fill = NA, colour = NA))
629   p2 <- x %>% filter(model != "n_obs") %>%

```

```

624 ggplot(aes(year, value, group = model, color = model, shape = model)) +
625 geom_line(linetype="dotted") + geom_point() + ggtitle("BIC") +
626 theme(legend.position= c(0.5, 1.1), legend.direction = "horizontal", legend.background =
  element_rect(fill = NA, colour = NA))
627 p3 <- y %>% filter(model != "n_obs") %>%
628 ggplot(aes(year, value, group = model, color = model, shape = model)) +
629 geom_line(linetype="dotted") + geom_point() + ggtitle("BIC/n") +
630 theme(legend.position= c(0.5, 1.1), legend.direction = "horizontal", legend.background =
  element_rect(fill = NA, colour = NA))
631 grid.arrange(p1, p2, p3)
632 }
633 # 非対称正規分布, 非対称ティール分布, 正規逆ガウス分布を当てはめたときの AIC の経年変化のプロット
634 plot.AIC.ts(finSPT.sub.IC.ts)
635 # 非対称正規分布, 非対称ティール分布, 正規逆ガウス分布を当てはめたときの BIC の経年変化のプロット
636 plot.BIC.ts(finSPT.sub.IC.ts)
637 # AIC, BIC 表: 対数株式時価総額への分布当てはめに関する情報量規準の経年変化
638 print(
639   xtable(
640     finSPT.sub.IC.ts,
641     digits = c(0, 0, 0, rep(2, 6)),
642     floating = FALSE,
643     caption = c("AIC, BIC 表: 対数株式時価総額への分布当てはめに関する情報量規準の経年変化"),
644     label = "table.year.variation.distribution.AIC.BIC"),
645     caption.placement = "top",
646     include.rownames=FALSE,
647     table.placement = "H")
648 # 非対称正規分布, 非対称ティール分布, 正規逆ガウス分布を先進国のデータに当てはめたときの AIC の経年
  変化のプロット
649 plot.AIC.ts(finSPT.sub.G.IC.ts)
650 # 非対称正規分布, 非対称ティール分布, 正規逆ガウス分布を先進国のデータに当てはめたときの BIC の経年
  変化のプロット
651 plot.BIC.ts(finSPT.sub.G.IC.ts)
652 # AIC, BIC 表: 対数株式時価総額 (先進国) への分布当てはめに関する情報量規準の経年変化
653 print(
654   xtable(
655     finSPT.sub.G.IC.ts,
656     digits = c(0, 0, 0, rep(2, 6)),
657     floating = FALSE,
658     caption = c("AIC, BIC 表: 先進国における企業の対数株式時価総額への分布当てはめに関する情報量
  規準の経年変化"),
659     label = "table.year.variation.distribution.AIC.BIC.G"),
660     caption.placement = "top",
661     include.rownames=FALSE,
662     table.placement = "H")
663 # 非対称正規分布, 非対称ティール分布, 正規逆ガウス分布を新興国のデータに当てはめたときの AIC の経年
  変化のプロット
664 plot.AIC.ts(finSPT.sub.E.IC.ts)
665 # 非対称正規分布, 非対称ティール分布, 正規逆ガウス分布を新興国のデータに当てはめたときの BIC の経年
  変化のプロット
666 plot.BIC.ts(finSPT.sub.E.IC.ts)
667 # AIC, BIC 表: 対数株式時価総額 (新興国) への分布当てはめに関する情報量規準の経年変化
668 print(
669   xtable(
670     finSPT.sub.E.IC.ts,
671     digits = c(0, 0, 0, rep(2, 6)),
672     floating = FALSE,
673     caption = c("AIC, BIC 表: 新興国における企業の対数株式時価総額への分布当てはめに関する情報量
  規準の経年変化"),
674     label = "table.year.variation.distribution.AIC.BIC.E"),
675     caption.placement = "top",
676     include.rownames=FALSE,

```

```

677 table.placement = "H")
678 #####
679 ### 有限混合分布当てはめに関する情報量規準の経年変化のプロット
680 #####
681 finSPT.sub.Mix.IC.ts <- finSPT.sub.G.IC.ts %>%
682   left_join(finSPT.sub.E.IC.ts, by = "year") %>%
683   rename(nG = n_obs.x, nE = n_obs.y) %>%
684   mutate(n = nG + nE, varpiG = nG/n, varpiE = nE/n) %>%
685   mutate(nGlogvarpiG = nG*log(varpiG), nElogvarpiE = nE*log(varpiE)) %>%
686   mutate(
687     AIC_NIG.G_NIG.E = AIC_NIG.x + AIC_NIG.y -2*(nGlogvarpiG + nElogvarpiE) + 2,
688     AIC_SN.G_NIG.E = AIC_SN.x + AIC_NIG.y -2*(nGlogvarpiG + nElogvarpiE) + 2,
689     BIC_NIG.G_NIG.E = BIC_NIG.x + BIC_NIG.y -2*(nGlogvarpiG + nElogvarpiE) + log(n) - 4*log(nG)
690       -4*log(nE) + 8*log(n),
691     BIC_SN.G_NIG.E = BIC_SN.x + BIC_NIG.y -2*(nGlogvarpiG + nElogvarpiE) + log(n) - 3*log(nG) -4
692       *log(nE) + 7*log(n)
693   ) %>% select(year, n, nG, nE, varpiG, varpiE, AIC_NIG.G_NIG.E:BIC_SN.G_NIG.E)
694 plot.AIC.Mix.ts <- function(obj)
695 {
696   require(tidyverse)
697   require(gridExtra)
698   x <- obj %>%
699     select(year, n, nG, nE, AIC_NIG.G_NIG.E, AIC_SN.G_NIG.E) %>%
700     pivot_longer(-year, names_to = "model", values_to = "value")
701   y <- obj %>%
702     mutate(AIC_NIG.G_NIG.E_n = AIC_NIG.G_NIG.E/n, AIC_SN.G_NIG.E_n = AIC_SN.G_NIG.E/n) %>%
703     select(year, n, AIC_NIG.G_NIG.E_n, AIC_SN.G_NIG.E_n) %>%
704     pivot_longer(-year, names_to = "model", values_to = "value")
705   p1 <- x %>% filter(model == "n" | model == "nG" | model == "nE") %>%
706     ggplot(aes(year, value, group = model, shape = model)) +
707     geom_line(linetype="dotted") + geom_point() + ggtitle("n") +
708     theme(legend.position= c(0.5, 1.1), legend.direction = "horizontal", legend.background =
709       element_rect(fill = NA, colour = NA))
710   p2 <- x %>% filter(model != "n", model != "nG", model != "nE") %>%
711     ggplot(aes(year, value, group = model, color = model, shape = model)) +
712     geom_line(linetype="dotted") + geom_point() + ggtitle("AIC") +
713     theme(legend.position= c(0.5, 1.1), legend.direction = "horizontal", legend.background =
714       element_rect(fill = NA, colour = NA))
715   p3 <- y %>% filter(model != "n") %>%
716     ggplot(aes(year, value, group = model, color = model, shape = model)) +
717     geom_line(linetype="dotted") + geom_point() + ggtitle("AIC/n") +
718     theme(legend.position= c(0.5, 1.1), legend.direction = "horizontal", legend.background =
719       element_rect(fill = NA, colour = NA))
720   grid.arrange(p1, p2, p3)
721 }
722 plot.BIC.Mix.ts <- function(obj)
723 {
724   require(tidyverse)
725   require(gridExtra)
726   x <- obj %>%
727     select(year, n, nG, nE, AIC_NIG.G_NIG.E, AIC_SN.G_NIG.E) %>%
728     pivot_longer(-year, names_to = "model", values_to = "value")
729   y <- obj %>%
730     mutate(BIC_NIG.G_NIG.E_n = BIC_NIG.G_NIG.E/n, BIC_SN.G_NIG.E_n = BIC_SN.G_NIG.E/n) %>%
731     select(year, n, BIC_NIG.G_NIG.E_n, BIC_SN.G_NIG.E_n) %>%
732     pivot_longer(-year, names_to = "model", values_to = "value")
733   p1 <- x %>% filter(model == "n" | model == "nG" | model == "nE") %>%
734     ggplot(aes(year, value, group = model, shape = model)) +
735     geom_line(linetype="dotted") + geom_point() + ggtitle("n") +
736     theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
737       element_rect(fill = NA, colour = NA))

```

```

732 p2 <- x %>% filter(model != "n", model != "nG", model != "nE") %>%
733 ggplot(aes(year, value, group = model, color = model, shape = model)) +
734 geom_line(linetype="dotted") + geom_point() + ggtitle("BIC") +
735 theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
  element_rect(fill = NA, colour = NA))
736 p3 <- y %>% filter(model != "n") %>%
737 ggplot(aes(year, value, group = model, color = model, shape = model)) +
738 geom_line(linetype="dotted") + geom_point() + ggtitle("BIC/n") +
739 theme(legend.position = c(0.5, 1.1), legend.direction = "horizontal", legend.background =
  element_rect(fill = NA, colour = NA))
740 grid.arrange(p1, p2, p3)
741 }
742
743 # 有限混合分布の当てはめに関する AIC の経年変化のプロット
744 plot.AIC.Mix.ts(finSPT.sub.Mix.IC.ts)
745 # 有限混合分布の当てはめに関する BIC の経年変化のプロット
746 plot.BIC.Mix.ts(finSPT.sub.Mix.IC.ts)
747 # 企業数と混合比率の推定値の経年変化
748 print(
749   xtable(
750     finSPT.sub.Mix.IC.ts[, seq(1, 6)],
751     digits = c(rep(0, 5), rep(2, 2)),
752     floating = FALSE,
753     caption = c("企業数と混合比率の推定値の経年変化: 企業数  $\hat{n}$ , 先進国企業数  $\hat{n}(G)$ , 新興国企業数  $\hat{n}(E)$ , 先進
      国の混合比率推定値  $\hat{varpi}(G)$ , 新興国の混合比率推定値  $\hat{varpi}(E)$ "),
754     label = "table.year.variation.observations"),
755     caption.placement = "top",
756     include.rownames=FALSE,
757     table.placement = "H")
758 # AIC, BIC 表: 対数株式時価総額への有限混合分布の当てはめに関する情報量規準の経年変化
759 print(
760   xtable(
761     finSPT.sub.Mix.IC.ts[, c(1, seq(7, 10))],
762     digits = c(rep(0, 2), rep(2, 4)),
763     floating = FALSE,
764     caption = c("AIC, BIC 表: 対数株式時価総額への有限混合分布の当てはめに関する情報量規準の経年
      変化"),
765     label = "table.year.variation.mixture.distribution.AIC.BIC"),
766     caption.placement = "top",
767     include.rownames=FALSE,
768     table.placement = "H")
769 #####
770 ### データの要約
771 #####
772 summary(finSPT.sub)
773 #####
774 ### 国タイプ情報
775 #####
776 library(countrycode)
777 country.type <- finSPT %>% select(country.y, type) %>%
778   arrange(country.y) %>% unique() %>%
779   arrange(type) %>%
780   mutate(country_name = countrycode(country.y, 'iso2c', 'country.name')) %>%
781   rename(country = country.y) %>%
782   select(country_name, country, type) %>% data.frame()
783 colnames(country.type) <- c("国名", "ISO2c", "国タイプ")
784 # 国タイプ情報の LaTeX 出力
785 print(
786   xtable(
787     country.type,
788     floating = FALSE,

```

```

789 caption = c("国タイプ情報"),
790 label = "table:country.type"),
791 caption.placement = "top",
792 table.placement="H",
793 size="\fontsize{9pt}{10pt}\selectfont"
794 )
795 #####
796 ### データの欠測情報の可視化
797 #####
798 library(VIM)
799 aggr(finSPT.sub, cex.axis = 0.7, col = "grey")
800 #####
801 ### 企業数情報
802 #####
803 suppressWarnings(invisible(capture.output(tmp <- count.obs())))
804 colnames(tmp) <- c("会計年度", "ESG 指標年度", "先進国企業数", "新興国企業数", "総企業数")
805 # 企業数情報の LaTeX 出力
806 print(
807   xtable(
808     tmp,
809     digits = c(0,0,0,0,0,0),
810     display=c("s",rep("f",5)),
811     floating = FALSE,
812     caption = c("年度別企業数"),
813     label = "table:number.firms"),
814     caption.placement = "top",table.placement="H",
815     size="\setlength{\tabcolsep}{2pt}"
816   )
817   #####
818   ### 企業数情報 (欠測値削除後)
819   #####
820   suppressWarnings(invisible(capture.output(tmp.na.omit <- count.obs.omit.NA())))
821   colnames(tmp.na.omit) <- c("会計年度", "ESG 指標年度", "先進国企業数", "新興国企業数", "総企業数")
822   # 企業数情報の LaTeX 出力
823   print(
824     xtable(
825       tmp.na.omit,
826       digits = c(0,0,0,0,0,0),
827       display=c("s",rep("f",5)),
828       floating = FALSE,
829       caption = c("欠測値除去後の年度別企業数"),
830       label = "table:number.firms.na.omit"),
831       caption.placement = "top",table.placement="H",
832       size="\setlength{\tabcolsep}{2pt}"
833     )
834     #####
835     ### 非対称正規分布の密度関数のプロット
836     #####
837     ggplot.sn <- function(x = seq(-10, 10, 0.01), xi = 0, omega = 1, alpha = 0, yup = 1)
838     {
839       require(sn)
840       require(ggplot2)
841       require(dplyr)
842       df <- data.frame(x, y = dsn(x, xi = xi, omega = omega, alpha = alpha))
843       p <- df %>% ggplot(aes(x, y)) + geom_line() + ylim(0, yup)
844       p + labs(title = paste("xi, _omega, _alpha =_", xi, ", ", omega, ", ", alpha, ")), y = "Deinsity")
845     }
846     # プロット
847     ggplot.sn(x = seq(-1, 3, 0.01), alpha = 5)
848     ggplot.sn(x = seq(-3, 1, 0.01), alpha = -5)
849     #####

```



```

850 ### 非対称テイラー分布の密度関数のプロット
851 #####
852 ggplot.st <- function(x = seq(-10, 10, 0.01), xi = 0, omega = 1, alpha = 0, nu = 0, yup = 1)
853 {
854   require(sn)
855   require(ggplot2)
856   require(dplyr)
857   df <- data.frame(x, y = sn::dst(x, xi = xi, omega = omega, alpha = alpha, nu = nu))
858   p <- df %>% ggplot(aes(x, y)) + geom_line() + ylim(0, yup)
859   p + labs(title = paste("(xi, _omega, _alpha, _nu) = (" , xi, ", ", omega, ", ", alpha, ", ", nu, ")"), y = "
      Deinsity")
860 }
861 # プロット
862 ggplot.st(x = seq(-3, 10, 0.01), alpha = 5, nu = 1)
863 ggplot.st(x = seq(-10, 3, 0.01), alpha = -5, nu = 1)

```

## 付録 G コンピュータ環境

本稿の執筆に際して主に利用したコンピュータ環境の情報を与える。

### ハードウェア環境

- MacBook Pro 2021:
  - Processor: Apple M1 Max
  - CPU Cores: 10
  - Main Memory: 64 GB
  - OS: macOS Monterey (12.5)
- Mac Studio 2022:
  - Processor: Apple M1 Ultra
  - CPU Cores: 20
  - Main Memory: 128 GB
  - OS: macOS Monterey (12.5)

### ソフトウェア環境

- R (R. Ihaka, R. Gentleman, R Core Team, <https://www.r-project.org/>)
- R Packages

- `countrycode` (V. Arel-Bundock, <https://vincentarelbundock.github.io/countrycode/>)
- `dplyr` (H. Wickham, <http://dplyr.tidyverse.org/>)
- `GeneralizedHyperbolic` (David Scott, <https://r-forge.r-project.org/projects/rmetrics/>)
- `GGally` (B. Schloerke, <http://ggobi.github.io/ggally/>)
- `ggplots2` (H. Wickham, [https://ggplot2.tidyverse.org](https://ggplot2.tidyverse.org/))
- `lubridate` (V. Spinu, G. Grolemund, H. Wickham, <https://lubridate.tidyverse.org>)
- `magrittr` (S. M. Bache, H. Wickham, and L. Henry, <https://magrittr.tidyverse.org>)
- `purrr` (L. Henry and H. Wickham, <https://purrr.tidyverse.org>)
- `readr` (H. Wickham et al., <https://github.com/tidyverse/readr>)
- `sn` (A. Azzalini, <http://azzalini.stat.unipd.it/SN/>)
- `VIM` (M. Templ, <https://github.com/statistikat/VIM>)
- `xtable` (D. B. Dahl, <http://xtable.r-forge.r-project.org/>)
- `RStudio` (`RStudio`, <https://www.rstudio.com/>)
- `Sweave` (F. Leisch, <https://leisch.userweb.mwn.de/Sweave/>)

R 関数 `sessionInfo` を実行することによって、本稿を執筆すること  
に利用した R に関する環境情報を以下に与える：

## sessionInfo による情報

- R version 4.2.1 (2022-06-23), aarch64-apple-darwin20
- Locale: ja\_JP.UTF-8/ja\_JP.UTF-8/ja\_JP.UTF-8/C/ja\_JP.UTF-8/ja\_JP.UTF-8
- Running under: macOS Monterey 12.5
- Matrix products: default
- BLAS:  
/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRblas.0.dylib
- LAPACK:  
/Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib
- Base packages: base, datasets, graphics, grDevices, grid, methods, stats, stats4, utils
- Other packages: colorspace 2.0-3, countrycode 1.4.0, dplyr 1.0.9, e1071 1.7-11, forcats 0.5.1, GeneralizedHyperbolic 0.8-4, GGally 2.1.2, ggplot2 3.3.6, gridExtra 2.3, lubridate 1.8.0, purrr 0.3.4, readr 2.1.2, sn 2.0.2, stringr 1.4.0, tibble 3.1.7, tidyr 1.2.0, tidyverse 1.3.2, VIM 6.1.1, xtable 1.8-4
- Loaded via a namespace (and not attached): abind 1.4-5, assertthat 0.2.1, backports 1.4.1, bit 4.0.4, bit64 4.0.5, boot 1.3-28, broom 1.0.0, car 3.1-0, carData 3.0-5, cellranger 1.1.0, class 7.3-20, cli 3.3.0, compiler 4.2.1, crayon 1.5.1, data.table 1.14.2, DBI 1.1.3, dbplyr 2.2.1, DEoptimR 1.0-11, digest 0.6.29, DistributionUtils 0.6-0, ellipsis 0.3.2, fansi 1.0.3, farver 2.1.1, fs 1.5.2, gargle 1.2.0, generics 0.1.3, glue 1.6.2, googledrive 2.0.0, googlesheets4 1.0.0, gtable 0.3.0, haven 2.5.0, hms 1.1.1, httr 1.4.3, jsonlite 1.8.0, labeling 0.4.2, laeken 0.5.2, lattice 0.20-45, lifecycle 1.0.1, lmtest 0.9-40, magrittr 2.0.3, MASS 7.3-57, Matrix 1.4-1, mnormt 2.1.0, modelr 0.1.8, munsell 0.5.0, nnet 7.3-17, numDeriv 2016.8-1.1, parallel 4.2.1, pillar 1.7.0, pkgconfig 2.0.3, plyr 1.8.7, proxy 0.4-27, R6 2.5.1, ranger 0.14.1, RColorBrewer 1.1-3, Rcpp 1.0.9, readxl 1.4.0, reprex 2.0.1, reshape 0.8.9, rlang 1.0.4, robustbase 0.95-0, rvest 1.0.2, scales 1.2.0, sp 1.5-0, stringi 1.7.8, tidyselect 1.1.2, tools 4.2.1, tzdb 0.3.0, utf8 1.2.2, vcd 1.4-10, vctrs 0.4.1, vroom 1.5.7, withr 2.5.0, xml2 1.3.3, zoo 1.8-10

なお、本稿では、非対称分布族をデータへ当てはめるために `sn` パッケージ<sup>26)</sup> を利用しており、一般化双曲型分布族の当てはめには `GeneralizedHyperbolic` パッケージを利用している。これらのパッケージを含む金融分野で利用される R 環境の包括的な解説は Pfaff (2016) が詳しい。

26) `sn` の構造とコマンドの概説は以下の URL が参考になる：<https://cran.r-project.org/web/packages/sn/vignettes/pkg-overview.html>