論文の内容の要旨

論文題目：
“Goodness-of-Fit-Tests for Heavy Tailed Distributions” (“裾の厚い分布の適合度検定”)

氏名：
東京大学大学院経済学研究科
経済理論学専攻・統計学コース 37012
松井宗也

—博士学位論文要旨—

最初に本研究の統計学や経済学における位置づけを述べ，後に本文の要旨を各章ごとに関連付けながら述べる．

現代統計学において正規分布の果たす役割は大きい．ある統計モデルを考えたときにその確率要素に仮定されるのはまず正規分布である．正規性の仮定は便利で多くの利点を持つことが理由として挙げられる．しかし，正規性の仮定だけでは十分に説明できない統計データがあることは昔から知られてきた．近年注目を集めている裾の厚いデータもそうである．正規分布はその性質から裾が指数的に減少する場合以外はデータへの当てはまりがよくない．経済学では以前から金融時系列データが裾の厚い分布によってよく説明されることが分かっている．また保険の分野でも裾の厚さは注目を浴びている。さらに最近では World Wide Web に関する多くのデータは裾が厚いことが分かってきた．その他，裾が厚いとされるデータが観察される分野は，統計物理，信号処理，通信など多岐にわたる．データをモデル化する目的は，説明，予測，リスク管理など様々である．本論文では裾の厚いモデルを論じる．

裾の厚いデータをモデル化する方法はいくつかある。詳しくは本論文の前書きを参照されたい．ここでは最も簡単なものとして直接データに裾の厚い分布を当てはめる方法のみ挙げておく．我々が主に扱うのもこの方法である．第 1.1 節に，実際のデータに対し当てはまりが良いとされる分布を紹介する．ごく最近であるが t 分布に歪みを入れるなどの新しい分布の研究が盛んになりつつあり，それについても触れている．注意しなければならないのは，理論的な背景から創られた分布や分布自体が複雑なものもあり一概に簡単とは言えない点である．

今までに裾の厚いデータに対して多くの研究がなされてきた．新しいモデルや分布の構築，それに関連した推定量の考案などが挙げられる．ただ，こうした新しいモデルや分布の検証に関する研究はまだ少ない．具体的には実際のデータに照らし合わせ，モデルや分布の仮定が妥当であるか検証する研究である．統計モデルが現実のデータをよく説明するために考え出されるとしたら，実際によく説明されているかを検証することは重要である．モデルの構築，検証という繰り返しがよりよいモデルを生んでゆく．また説明できる限界を知ることも重要である．

この方向ではモデル選択や分布の適合度検定などの研究分野が挙げられる．モデル選択に関しては本論文の前書きを参照されたい．我々の取り扱う研究は裾の厚い分布の適合度検定に関してである．この研究は裾の厚いデータを直接分布によってモデル化する際，その当てはまり具合を調べる研究である．正しいと仮定される分布とデータから得られる経験分布との距離を何らかの方法で測り，その離れ具合を見て検定を行う．カイ 2 乗検定が最も有名であるが，他にコルモゴロフ・スミルノフ検定，アンダーソン・ダーリン検定などが知られている．

本論文扱う対象は安定分布という裾の厚い分布のクラスである．正規分布以外は 2 次のモーメントを持たないことで知られている．扱う理由は本文の前書きにも述べているが以下に簡単に述べる．「安定分布は一般中心極限定理の極限分布として得られる重要な分布で，正規分布の自然な拡張と考えることができる．理論的には古くからいろいろと応用が考えられている．加法過程の中でも時間のスケールに関して不変な確率過程を表現する分布族として知られている．」

ただし忘れてはならない欠点もある．2 次以上の有限モーメントを持つ分布がモデル化できない．多くの安定分布は密度関数が解析的に表現できないことなどである．前者に関しては第 1.1 節に述べる他の分布族を用いればよい．後者の欠点が大きい為，実際の取り扱いがやや難しい．

さて我々は安定分布に関する適合度検定を行なうわけであるが，解析的に表現できない密度関数もしくは分布関数をどう取り扱うかという問題が残る．解決方法としては分布関数と一対一の関係にある特性関数を用いる．正しいと仮定される分布の特性関数とデータから得られる経験特性関数の距離を見るのである．以下では本文の各章の内容を我々の目的である安定分布の適合度検定とのつながりに触れつつ述べる．

第 1 章（導入）は導入として基本的なことを述べる．第 1.1 節（裾の厚い分布）ではこれから応用が期待されるであろう裾の厚い分布族を簡単な性質を交え述べる．一般化ハイパボリック分布，一般化 t 分布，安定分布の 3 つである．安定分布以外の分布への適合度検定も応用として考えられる．特に後の章で取り扱う安定分布に関しては第 1.2 節（安定分布の特筆すべき性質）で Feller (1971) に沿って理論的な性質に触れる．安定分布が一般化中心極限定理の極限分布をなすということの意味を明確に述べた．第 1.3 節（適合度検定）では適合度検定の基本的なものとしてクラメール・ホンミーゼス統計量とアンダーソン・ダーリン統計量を Anderson and Darling (1952) に基づいてやや詳しく解説する．これ

らは仮定される分布とデータから得られる経験的な分布の2乗距離に重み関数を掛けて積分し導出される．第4，5章で見る経験特性関数を用いた適合度検定も2乗距離を測るタイプの検定であり共通する部分も多いのでここで予め説明しておく．

第2章（安定分布の密度とその微分）では安定分布の欠点であった密度関数の取り扱いを対称安定分布の場合に述べる．密度関数とパラメータに関する微分を数値的に求め，それを用いてフィッシャー情報量を計算した．方法としては密度関数のZolotarevによる積分表現と密度関数の無限級数展開を組み合わせ数値的な計算を行うものである．Nolan (1997) を改良したものとなっている．

第3章では安定分布の最尤推定量の性質について論じる．先行研究や第2章を見れば最尤法は数値的に可能なことが分かる．理論的にもパラメータ空間の境界を除いては漸近正規性も証明されている．ただし正規分布に近づくにつれて情報量が発散するためこの周辺での最尤法の挙動は明らかではなかった．Nagaev and Shkol'nik (1988) は対称な場合に正規分布近くでの指数パラメータ α の情報量の挙動を明らかにした．この章ではこれを拡張して一般安定分布の全てのパラメータの情報量の挙動を明らかにした．

第5，6章では前の章の結果を用いて最終目的である対称安定分布の適合度検定を行う．検定統計量としてはデータから推定されたパラメータを代入した真の特性関数とデータから経験的に得られた特性関数との2乗距離に重み関数をかけて積分したものである．指数パラメータ α を未知として扱うか既知として扱うかによって2通りの検定を考えた．未知として取り扱う場合は検定の対象が分布族になることに注意されたい。パラメータの推定量としては主にMLEを用いた．検定統計量の分布は通常モンテカルロ実験により求められるが本論文では他に数値的にその漸近分布を求めた。第1.4節でもあるように検定統計量の漸近分布はガウス過程を2乗して積分したもので与えられる．まずガウス過程の共分散構造を解析的に求め，それを固有値展開することで検定統計量の特性関数が求まる．そして特性関数を数値的に反転して漸近分布を求めた．特に反転に経路積分 Slepian (1957) を用いた点が新しい点である．対立仮説に対する検出力もモンテカルロ実験により導出した． (要旨ここまで)