

安定分布における投資教育について

沼田 涼 *1

N145219k@st.u-gakugei.ac.jp

*1: 東京学芸大学教育学部人間社会科学課程総合社会システム専攻

◎Key Words 正規分布, ファットテール, 投資教育

1. はじめに

今日、統計分析の様々な手法やモデルは、一般線形モデル、一般線形化モデルという包括的方法で議論されている。このことにより、取り扱うデータごとに選択していた複雑なモデルを簡易に捉えることができる。また、こうして簡易に捉えることにより、データ分析の一般的信頼性の確保と分析の労力の軽減を実現させることが可能となる。

一般線形モデル、一般線形化モデルでは、理論的に正規分布の役割は非常に大きい。なぜならデータの正規性を仮定することは多様な分析・予測を可能にさせるなど、現代の統計学と分析一般において、様々な利点を持っているからである。ところが、実際のデータを観察すると、データの分布が安定分布、いわゆる正規分布でない場合が多い。正規性がないと判定することが妥当と思われるデータの中には「裾の厚い」データがある。これは「ファットテール」と呼ばれるものであり、近年非常に注目を集めている分布である。

正規分布の特徴の一つに、分布の裾が指数とともに減少する、ということがある。従前、株価データや外国為替データは、観察の結果、正規分布であると仮定して議論を展開することが多かった。またこの仮定によって、一般線形モデル等の枠組みを活用して議論を展開されていた。しかし、実際に、例えば株価データや金融時系列データでは、しばしばこの「裾の厚い」分布が観測されている。

本研究では、実際に外国為替データを用いることにより、投資教育における裾の厚い分布、いわゆる「ファットテール」について検討をし、今後の投資教育のあり方について考察する。

2. 正規分布

先に述べてきた安定分布の中でも最も一般的である正規分布について整理する。

2.1 正規分布とは

正規分布(normal distribution)とは、偶発的なデータの揺らぎによって生じる統計学で最も基本的な確率分布である。あるサンプル集団のばらつきが、その平均値を境として前後同じ程度にばらついている状態を示し、分布図で見ると、平均値を線対称軸とした左右対称の釣鐘型でなだらかな曲線を描く。つまり、平均値の周辺にサンプルが多く集まり、値が大小の左右の裾野に向かうとサンプル数が急激に減るということである。

2.2 正規分布の特徴

正規分布は自然界で起こる現象の多くがその分布に当てはまること、特に平均値に関する分布が当てはまることから、統計学では最も重要な分布となっている。

正規分布は平均 μ 、分散 σ^2 とした、ただ2つのパラメーターによってその形が決定される。一変量の確率変数 x が平均分散の正規分布に従う時、確率密度関数は以下の式で表される。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu=0$ $\sigma=1$ の場合を特に標準正規分布²といい、確率密度関数は以下の通り

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{x^2}{2}}$$

上記2つの式を見てもらえばわかるが、正規分布の確率密度関数を使うと「平均、分散」が求まれば確率密度がすぐに計算でき、ひいては確率分布も求めることが可能となる。

2.3 確率密度関数

連続型確率変数³ X に対して、 X が a 以上 b 以下となる確率が、積分を用いて

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

で与えられる時、 $f(x)$ を確率密度関数という。

確率=面積で捉えることにより連続型確率変数の様子を簡潔に表現することが可能となる。

確率密度関数の特徴として連続型確率変数 X の取りうる値の下限値を A 、上限値を B と置くと、 $f(x)$ が確率密度関数の時

$$\int_A^B f(x) dx$$

が成立する。確率は全て足し合わせると1になるからである。

2.4 中心極限定理

正規分布が統計学上特別な地位を持つのは中心極限定理が存在するためである。

中心極限定理の定義であるが

「平均 μ 、分散 σ^2 の母集団からの取り出されたサンプル n の標本(サンプル)の平均 \bar{X} の分布は、 n が大きければ、近似的に平均 μ 、分散 $\sigma^2/2$ の正規分布に従う。」とする。

この定理の要点として以下のことが挙げられる。

1)母集団が正規分布でなくとも十分性質の良い分布の場合は標本(サンプル)は正規分布に従うというのが重要であるということ。ただし、母集団が平均 μ 、分散の意味を失うような分布(例えばべき分布⁴)の場合にはこの定義は使えない。

2)分散は標準偏差の平方であるため、標本平均の標準偏差は母集団の標準偏差の $1/\sqrt{n}$ になる。つまり、サンプル数を増やすと標本平均 \bar{X} の母平均 μ からのズレがどんどんなくなっていく(精度が良くなる)。もちろん元の分布に平均や分散が存在しない場合は、中心極限定理は成り立たない。

3. 外国為替データの収集・整理

本節では、円ドルレート収益率データの収集を行い、実際にそれらのデータに正規性が見られるのか確認する。

長期間でデータの収集を行うとどうしても外部的な経済動向の影響を受けてしまうため、今回は短期間のデータを取り扱った。

具体的な期間として、2017年6月13日18時00分より5分毎に250回連続して集計を行った。

以下に収集した為替レートの変化を載せる(図1)これはx軸で回数経過(時間経過)を表し、y軸でx軸に対応する為替レートを示した分布グラフである。

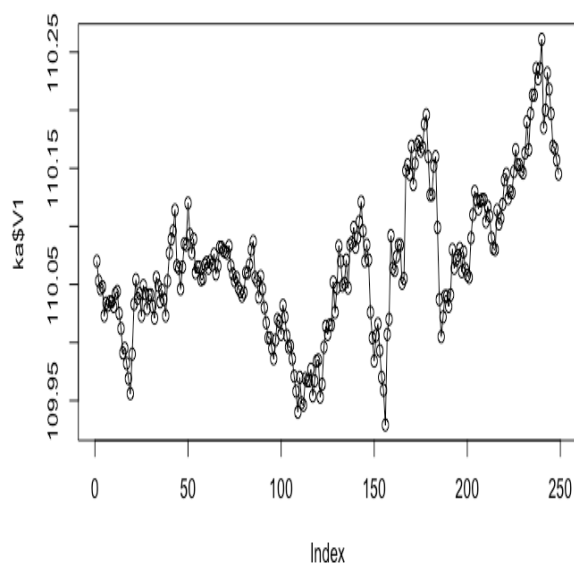


図1)x軸=時間、y軸=為替レート

また図1の為替レートの変化をただ見ただけでは取り扱っているデータが正規分布しているかどうか判断がつかないため、以下にQQプロット(図2)とヒストグラム(図3)を載せる。

QQプロット・ヒストグラムはそれぞれがデータの分布が正規分布しているかどうかを目視で判断するためのグラフである。

QQプロットとは、x軸に観測した累積パーセント、y軸に期待累積パーセントをもつグラフである。

図2における直線は正規分布を直線化したものであり、観測したデータがその直線に沿っていればそのデータは正規分布していることになる。

図2を見てもらうとx軸の値が1を超えたあたりから観測したデータが正規分布を表す直線からずれていることが確認できる。これは図3のヒストグラムのx軸110.10の値を超えたあたりから正規分布よりも緩やかな曲線になっていることを表している。また、図2のx軸が-1のあたりでは、観測したデータの分布が正規分布を表す直線よりも下にずれている。これは図3のヒストグラムのx軸109.95の値を超えたあたりから正規分布よりも急激な曲線を描いていることを表している。

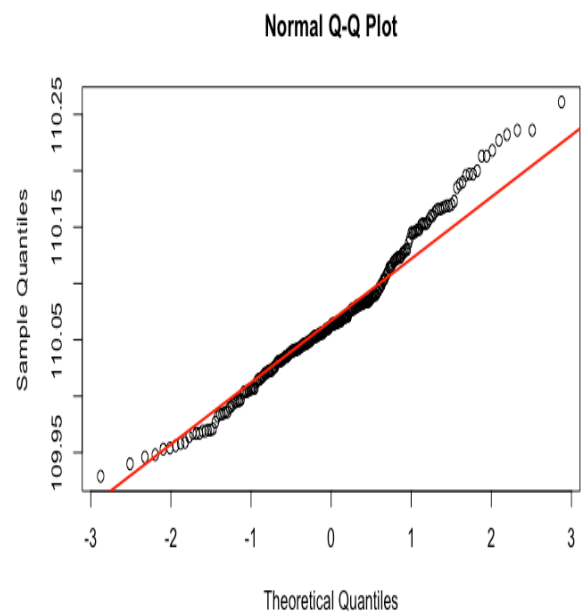


図2) QQプロット

図3は図1をヒストグラムにしたものであり、ヒストグラムはx軸で階級を、y軸で度数を表したグラフである。

図3の曲線はヒストグラムに対する確率分布である。

図2、図3を関連して考えることにより、観測したデータが正規分布しているか否かを判断できるようになっている。

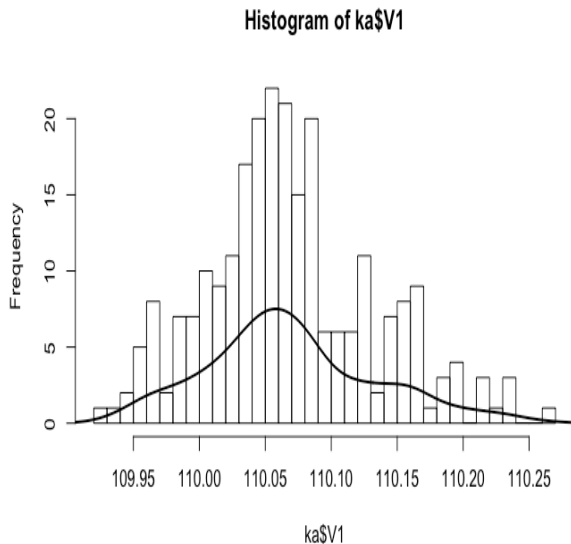


図3) ヒストグラム(30分割)

- ※ 過去為替データについては幅広い外国為替市場を取り扱っている「FXSTREET」
<http://www.fxstreet.jp/forex-tools/ratehistorytools/?tf=5min&period=250&pair=usdjpy>を利用し集計した。

4. 分布の分析

4.1 Shapiro-Wilk検定を用いた正規性判定

前章で得た円ドル為替レートの収益率について、正規性を確認していく。今回用いるのは、Shapiro-Wilk検定（以下SW検定）である。SW検定は、正規性の検定の一つで、データが正規分布しているかを判断するために用いる。これは帰無仮説⁵を「変数は正規分布に従う」とする仮説検定方法で、正規性検定の最も一般的なものである。

SW検定は同じく正規性検定に用いられる「Kolomogorow-Sminov検定⁶」に比べて少ないデータの検定に適しているため、SW検定を用いることにした。

検定に置いて、検定統計量に対応した確率p値が0.05未満であれば、帰無仮説が棄却され「データは正規分布に従っていない」ことがわかりp値が0.05以上であれば帰無仮説は棄却されないことになる。

今回は統計処理言語「R version3.3.2」を用いて判定する。

4.2 非正規性の確認

SW検定は先にも述べたが、得られたデータが正規性に従うものなのか否かを調べる検定方法である。すなわちデータが正規母集団に由来するという帰無仮説を棄却する。

以下に今回取り扱ったデータの検定結果を載せる。

```
####—シャピロ検定(5%有意)で正規性がないものを除去(正規性がないとはいえない銘柄を表示)
```

```
#株価データの読み込み
```

```
ka <- read.csv("kawasedata4.csv", header = F)
```

```
#文字列を数値化
```

```
as.numeric(ka$V1)
```

```
#時系列変動の図示
```

```
plot(ka$V1)
```

```
lines(ka$V1)
```

```
#正規性判定(シャピロウィルク検定)
```

```
shapiro.test(ka$V1)
```

```
#QQプロット
```

```
qqnorm(ka$V1)
```

```
qqline(ka$V1, lwd=2, col="red")
```

```
#ヒストグラムとして表示
```

```
hist(ka$V1, breaks = 20)
```

```
lines(density(ka$V1), col=1, lwd=2)
```

```
Shapiro-Wilk normality test
```

```
data: ka$V1
```

```
W = 0.97914, p-value = 0.001011
```

P値=0.001011となり、5%有意水準において棄却され、正規性でないことを確認できた。

5. 考察

5.1 従来のリスク評価方法について

予想していた通り、外国為替レートについても正規分布していないことが見て取れた。またサンプル数こそ少ないものの、図2、図3より正規分布していないだけでなく、裾の重い「ファットテール」も確認することができる。

こうした分布を正規分布として仮定し、議論を進めていくことはリスクを過小評価し危険性の伴うことにつながっている。

一方で、大数の法則の観点から、取り扱うサンプル数を限りなく増やした際に観測するデータが正規分布していくのかどうか、という点においてはまだまだ議論が必要である。しかし少なくとも、今回取り扱ったデータでは、短期における変動において非正規性を確認し、正規性を仮定とした投資ではリスクを捉えきることができないということが示せた。

リスクをより正確に捉えることのできる新しいリスク評価手法が求められる。

5.2 投資教育における今後の展望

上で述べてきた円ドル為替レートのように、証券価格全体は非線形現象であるとし、また、変動の多くは正規分布とされているが、実際のデータを再度検定にかけると非正規分布であることが確認できた。

従来の投資の捉え方においては正規分布であると仮定

して議論が進められることが基本となっていたが、そうではなく「正規分布ではない」と仮定し、リスクを過小評価することなく正確に捉えることが必要になってくる。

では実際に、投資セミナーであったり証券会社を訪れた投資をこれから始める人・投資をすでに始めている人に対し、どのように「投資」というものを教えていくのか。投資教育現場において教える側と受け手側の知識の乖離は大きい。そうした状況の中、裾の重い分布「ファットテール」というのは取るに足りない、考慮に入れなくともよい現象なのかもしれない。しかし、これらは投資に関わる人全てが知るべき知識であり、今後出てくるであろう新しいリスクの考え方に必要不可欠なものとなってくるのではないか。

「正規分布している」という仮定を棄却しリスクを正確に捉えることこそが、今後の投資教育の要になり、今後新しくなるであろうリスクの考え方の変化に柔軟に対応するための武器になる。

安定分布を仮定とする投資教育の今を見直す必要がある。

6. おわりに

本稿では、果たして為替変動に正規性が見られるのか実際のデータを取り扱い、検証を行った。現在投資というのはより重要性を増してきている。金融庁も「貯蓄から投資へ」とスローガンを掲げているように今後投資がますます活発なものになっていくのは間違いない。

投資に対する研究は進み、また様々な投資解析ソフトも発達してきている。しかし一方で投資教育においては欧米諸国に圧倒的に遅れをとっている現状がある。

本稿では、そうした現状を踏まえ、日本における投資の考えを整理した上で「非正規性」を捉えることの重要性を唱えることを一つの目的とし、一つの例として円ドル為替レートの正規性検定を行った

「実際の為替レートの正規性判定をし、非正規性を確認する」ということのみにとどまったが、その事実を活かしたリスク評価手法については言及できていない。また取り扱ったデータのサンプル数が少なく、まだまだ改善点があるように思える。

これらを今後の課題とし、さらなる実践的なリスク評価手法に取り組んでいきたい。

参考文献

- (1) 松原望『入門確率過程』p12-33 東京図書株式会社 (2011)
- (2) 松原望『統計の考え方』松原団法人放送大学教育振興会 (2003)
- (3) 高安秀樹『経済物理学の発見』光文社新書 (2013)
- (4) 新堀友太 「裾厚な分布を対象としたリスク強化の検討～外国為替データのt分布による考察を中心に～」東京学芸大学 (2015)

¹ ファットテールは正規分布では示すことができないような極端な変動を示す分布の一つ。大きな変動を示すリターン頻度が高くなり、リターン頻度分布の裾野が厚くなる。

² 正規分布の中で、確率変数を一定の式にし、確率密度関数を簡単な形にしたものこと。平均が0、標準偏差1になるという特徴がある。

³ 時間・温度。身長など取りうる値が無限大に存在する確率変数のこと。(⇔離散型確率変数)

⁴ ベキ分布は1890年代にイタリアのパレートが発見。極端な値をとるサンプルの数が正規分布より多く、そのため大きな値の方向に向かって曲線は長くなだらかに裾野を伸ばしていく。裾野は正規分布よりも広い(ファット・テール)。

⁵ 統計学上の仮説で、ある一つの変数が他の一つの変数、もしくは一群の変数と関係がないとする仮説あるいは二つ以上の母集団の間の差がないとする仮説これが成立するならば、得られた結果は偶然によって支配されたと予想される結果と違わないことになる。否定された場合には対立仮説の信頼度が高くなる

⁶ Kolmogorov-Smirnov検定は統計学における仮説検定の一種であり、有限個の標本に基づいて、二つの母集団の確率分布が異なるものであるかどうか、あるいは母集団の確率分布が帰無仮説で提示された分布と異なっているかどうかを調べるために用いられる。