

「Python3 ではじめるシステムトレード」の学びかたガイド

システムトレードという言葉は和製英語で正式には systematic trading といいます。予め定められた規則に従って株式、債券、為替などの流動性の高い(取引の活発な)金融商品を売買する方法だと考えればよいと思います。長い間、金融商品の価格を科学的に分析するといっても、十分なデータが得られずに来ました。そのために、イデオロギーや概念、語り継がれている売買手法が注目を浴びてきました。しかし、システムティック・トレーディングの世界ではデータを重視して、売買の方法を探していきます。したがって、本書がもっとも大事にしているのが

1. データ
2. モデル
3. 売買の規則

です。

データを客観的に分析するために、価格の動きのメカニズムを説明するモデルを考えます。また、投資目的を達成するために売買の方法を作り上げていきます。設計した売買が実現可能かどうかを判断するためには、金融市場の売買のメカニズムを知っている必要があります。そして、出来上がった取引手法の作り出す成果、またはその効果を客観的に判断するために、統計的手法を用います。本書では、頭の中で考えたり、紙に書いて方程式を解くというよりも、コンピューターのプログラムを使って、頭に描いたイメージを模擬売買することに

重点をおいています。その過程で、価格の動きをグラフに書いたり、模擬売買の結果をグラフに書いたり、統計的に分析したりします。

本書を読む目的

本書の最終的な目的は持続可能な取引戦略を手に入れることです。

本書を読むために必要な知識

本書を読むために必要な知識は、為替、債券、株式等の基礎知識、Python の基礎的な文法、データ解析ライブラリの使い方、統計学の知識などですが、これらの知識は必要に応じて本書の中で説明して行きます。為替、債券、株式、Python、ライブラリ、統計学などの漠然としたイメージがあればよいと思います。

金融市場の知識

為替、債券、株式などの金融商品の知識、取引所、私的市場の取引の規則に関する知識です。

Python3 の基礎的な文法

初心者の方でも心配はいりません。PCを使い慣れている方ではどなたでも読みこなせると思います。Python の文法については、リスト、タプル、辞書などの組み込みデータ構造、for ループ、in 関数、print 関数、def 文による関数作成の使い方を学んでいきます。クラスを自分で作る知識は必要なく、クラスを使う知識を習得していきます。

データ解析ライブラリの基本的な使い方

本書で用いるデータ解析ライブラリは numpy,statsmodel,matplotlib,pandas です。numpy,pandas,statsmodel は使いますが、精通している必要はありません。matplotlib はグラフを書くだけですので、すぐに使いこなせます。まずは、データを取得して、取引手法のバックテストができるように説明して行きます。

数学の知識

微積分や行列の知識は必要ありません。足し算、引き算、掛け算、割り算ができれば十分です。

統計学の知識

統計学の知識が必要かというよりも、本書でデータ分析に必要な統計的手法を学んでいきます。本書で最も重視しているのが線形回帰とランダムウォークです。ランダムウォークのもつ意味が分かれば、本書の目的の一部は達成されたこととなります。

本書の構成と難易度

難易度は左から順に金融の知識・Python の知識・ライブラリの知識・数学の知識・統計学の知識です。

難易度の程度はA:知識不用、B:基礎知識要、C:知識要

|第 1 章|はじめに|AAAAA|

|第 2 章|Python notebook をはじめよう ...|AAAAA|

|第 3 章|まずは Python から.....|AAAAA|

|第 4 章|pandas と matplotlib を使ってみよう..|AABBA|

|第 5 章|トレンドを理解する.....|AAAAA|

|第 6 章|トレンドをモデル化しよう……|AABBB|

|第 7 章|確率的トレンドとは何か?……|AABBB|

|第 8 章|さらにグラフィカルに調べよう……|AABAA|

|第 9 章|3Dサーフェイスで見える世界……|AABAA|

|第 10 章|モンテカルロで見える世界……|AAABC|

|第 11 章|季節性とマクロ変数を加える……|BABBC|

|第 12 章|取引戦略の第一歩……|BBBAA|

|第 13 章|歩み値の世界へようこそ……|BBBAA|

|第 14 章|歩み値から学べ……|BBBAA|

|第 15 章|歩み値にありて飛べ……|BBBAA|

|第 16 章|ランダムウォークを制覇する……|CBBA|

本書の読み方

速習コース

基本的な知識をなるべく早く身につけたい方は

1-2-3-4-5-6-7-9-12

を読んでみてください。

特訓コース

Python、ライブラリの基礎知識があるので、システムトレードの知識を得たい

人は

5-6-7-8-9-10-11-12-13

を読んでみてください。

HFTコース

ランダムウォーク、時間トレンド、自己回帰モデルの理解が確かで、高頻度取引を目指すひとは、

13-14-15-16

を読んでみてください。

ランダムウォークの理解を確認するには、確率的トレンド、ドリフト率、そして確定的トレンドの意味が理解できているかどうか、明確に区別できるかどうかで分かります。長い経験をもつプロの方でもこれらの意味があいまいな場合には本書の5章以降を丁寧に読む必要があるかもしれません。

本書を読むための環境構築

anaconda jupyter notebook が PC にインストールされている必要があります。本書のプログラムコードは Python 3.5 に対応していますが、[innovation1005/python3-for-system-trade/360](https://github.com/innovation1005/python3-for-system-trade/tree/master/360) から 3.6 に対応したコードが [innovation1005/python3-for-system-trade/270](https://github.com/innovation1005/python3-for-system-trade/tree/master/270) から 2.7 に対応したコードがダウンロードできます。

また、本書のプログラムコードが用いるデータですが、多くは yahoo finance または fred からダウンロードします。しかし、一部購入するしか手に入らないデータがあります。出版社のご厚意で、その一部のデータに関してダウンロード可能です。データの著作権上すべてのデータを配布出来ないことをご了承ください。データのダウンロードは「Python3 ではじめるシステムトレード」購入者に限り、panrolling 社のホームページから可能です。

Python 3.x と 2.7 について

本書のプログラムコードは、Window 10、Python 3.6 で動作確認がされています。また、Ubuntu16、Python 3.6 で動作確認がされています。Mac での大きな問題はいまのところ報告されていません。Python 2.7 に対応したプログラムコードは [innovation1005/python3-for-system-trade/270](https://github.com/innovation1005/python3-for-system-trade/270) からダウンロード可能です。

大きな違いは実行スピードです。明らかに Python 2.7 の方が実行速度は本書のコードに限っていえば早いです。その他に関しては大きな違いが見られません。print 文の使い方、pandas.DataFrame の使い方で若干の変更がありますが、その他には特に問題は見られませんでした。したがって、スピード重視の方は 2.7 を使うという手はあるかと思います。しかし、python2 の開発が終了していることを考えると、python3 がお勧めです。また、python3 でも気を付けてプログラムして行けば Python2 で大きな変更がなく動作することから、どうしても速度がほしいときだけ、2.7 で動かすという手もあると思います。

多く寄せられた質問

利回り、収益率、変化率についての質問

投資によりどれだけの収益が得られたかを表す用語として、利回りとか、収益率とか、変化率とか、金融の世界では混乱を招く表現が数多くあります。利回りとは、投資期間中に得られた収益を、その投資額で割った比率です。この収益には投資期間中に得られた利子、配当、資産価格の変化が含まれます。

利回り

たとえば銀行に 100 万円預金したときに、利子率が 5%で 5 年間投資すると、利子 25 万円が得られます。そうするとそのときに収益は 25 万円になります。年間収益は(25/5)万円=5 万円になります。これを投資金額の 100 万円で割ると 5%になります。これを利回りと呼び、このような利回りは単利といいます。つぎに途中で支払われる利子にも利子が支払われる場合があります。このような性質をもつ利子は、単利ではなく複利といいます。

$$1 \text{ 年後の元利合計 } 100 \times (1 + 0.05) = 105$$

$$2 \text{ 年後の元利合計 } 105 \times (1 + 0.05) = 110.25$$

$$3 \text{ 年後の元利合計 } 110.25 \times (1 + 0.05) = 115.7625$$

$$4 \text{ 年後の元利合計 } 115.7625 \times (1 + 0.05) = 121.55$$

$$5 \text{ 年後の元利合計 } 121.55 \times (1 + 0.05) = 127.63$$

になります。将来の受取額は

$$\text{預金元本} \times (1 + \text{利子率})^{\text{年数}}$$

になります。単利と複利で受取額が異なるので注意してください。また、年何回の利子を得るかでも計算は違ってきます。

つぎに利子が支払われない割引債の投資を考えてみましょう。

$$\text{単利では } (\text{償還価格} - \text{買付価格}) / \text{買付価格} \times 365 / \text{未経過日数} (1 \text{ 年未満}) \times 100$$

$$\text{複利では } (\text{償還価格} / \text{買付価格})^{(1 / \text{残存年数})} \times 100$$

になります。このような計算は、日本、米国、ヨーロッパで方式が異なるので注意しましょう。

リターン

さて、投資、金融工学の文献にはリターンという用語がよくでてきます。x を価格、価格の変化を dx とし、変化率を(dx)/x と定義すると、多くの場合、リターンは変化率という意味で使われています。たとえば、株式の売買で価格の変化により得た利益は、利息とか、配当とかは関係ありません。つまり、価格の変化の度合いだけによって、利益が確定しています。したがって、リターンは

$$\text{売付価格/買付価格}-1$$

で表現されます。これを年率に換算したい場合があります。その際には、

$$(\text{売付価格/買付価格})^{(252/\text{投資営業日数})}-1$$

が用いられる場合があります。営業日をもとに考える場合、1年間の長さは国によって異なり、大体 250-260 の間の数字が使われるようです。

対数リターン・対数収益率

長期の投資では対数リターンが用いられる場合もあります。対数リターンは

$$\ln(S(t))-\ln(S(t-1))$$

で計算されます。対数リターンの平均は

$$E(X)=(X_1+X_2+\dots+X_n)/n$$

で求められます。ここで $X_1=\ln(X(t))-\ln(X(t-1))$ です。X(t)はt時の価格です。

これを年率換算するには

$$\exp(E(X))^{252}$$

となります。E(X)は幾何平均で近似可能です。

ボラティリティの計算

金融市場でのボラティリティとは、年率換算した価格の変動性を示します。一般には日々の対数リターンの標準偏差を年率換算します。対数リターンは $\ln(S(t)) - \ln(S(t-1))$

で計算されます。分散はそれぞれの確率変数 (X_1, X_2, \dots, X_n) が独立であれば $V(X_1, X_2, \dots, X_n) = V(X_1) + V(X_2) + \dots + V(X_n)$

が成り立つので、年率換算した分散は日足の対数リターンから求めた分散を 252 倍、 $V(X) \times 252$ すれば求められます。

したがって、その標準偏差は日足の対数リターンの標準偏差に 252 の平方根を掛ければ、得られます。

対数リターンの代わりに幾何平均を用いて近似が可能です。

どのように将来の価格を予測するのですか？

将来の価格は予測するものではなく、作るものです。それは経済も一緒です。将来の経済を、予測できないほど変化にとんだ楽しいものにして行きましょう。そして、不確実な世界を楽しみましょう。

ランダムウォークなんて酔っぱらいの千鳥足に例えらえる価格の動きのことでしょ？

ランダムウォークの概念を、もっと正確にとらえるようにしましょう。それは将来の価格の予測と密接に関係があります。ランダムウォークでもトレンドは出ます。しかし、そのトレンドはいつ起こり、いつ終わるかが予測できないのです。したがって、売買のタイミングを図るためには使えません。また、ランダムウォークにしたがう価格は、たとえ上昇トレンドにあっても、一度暴落するとそれがもと

の水準にもどるかどうかは確かではありません。また、科学的にランダムウォークというときには、それは単なるモデルとして捉えられているので、現実の世界の価格の動きはもっと複雑だということを忘れないようにしましょう。

価格の対数をとるのはなぜですか？

(p.80-82 に価格の対数の記述がありますが、数式の誤植、説明不足の部分があり、読者の方にはご迷惑をおかけしました。お詫び申し上げます。)

価格の桁の調整

価格は上昇し続けると価格の桁が変わってしまいます。そのような価格上昇をグラフに描くと右肩上がりの指数関数的な上昇になります。そのようなグラフでは、直近の価格の動きは把握しやすいのですが、価格の水準の低いときの価格の動きはすべてつぶれてしまいます。そのようなグラフでは直感的に価格の動きをつかめません。

たとえば、価格が 1,10,100,1000,10000,100000,1000000 と動くとき、x軸に時間、y 軸に価格をとると、時間の経過とともに、指数関数的に上昇するグラフになります。

この価格を 10 の 0,1,2,3,4,5,6 乗と表現することができます。この整数の部分を y 軸にプロットすると直線のグラフになり、見やすくなります。

このような計算には対数を用います。対数には常用対数と自然対数があります。常用対数では $x=a^p$ を $\log_a(x)=p$ のようにあらわします。a を底と呼びます。自然対数では $a=e$ (ネイピア数)とします。 $x=e^p$ を $\ln(x)=p$ と書きます。ここで気をつけなければいけないのは、本書のプログラムの中では numpy の log

を用いています。しかし、この log は ln の意味です。間違えないようにしましょう。

さらに numpy で常用対数を使いたいときには、log10 を用います。

瞬時的変化率

x を価格とすると、 $d\ln(x)/dx=1/x$ と書けます。これを書き換えると

$$d\ln(x)=dx/x$$

と書けます。dx/x は変化率、またはリターンと呼ばれるものです。したがって、

時間の変化をゼロに近づけると微小時間の変化になります。したがって、価格

の対数をとってグラフに描くとその一点一点の動きは瞬時的な変化率ととらえ

ることができます。金融の世界では、対数 log ではなくて、自然対数の ln を使

います。

成長率

$$(dx/dt)/x=1/x(dx/dt)$$

は成長率を表しています。ここで、 $x=\ln x(t)$ とすると、

$$d \ln(x)/dt = 1/x (dx/dt)$$

となり、対数変換は成長率の計算に便利であることが分かります。たとえば、

$$y=x \cdot z$$

$$\ln(y) = \ln(x \cdot z) = \ln(x) + \ln(z)$$

$$1/y (dy/dt) = 1/x dx/dt + 1/z dz/dt$$

と書くことができます。y の瞬時的な成長率は x と z の瞬時的な成長率の和にな

ります。

弾力性

価格 x と、その価格に影響を与える別の要素 z の関係を調べるときに、それぞれの対数をとって線形のモデルとして分析する場合があります。

$$\ln(x_i) = \beta_1 + \beta_2 \cdot \ln(z_i) + u_i$$

この場合の β_2 は z に対する x の弾力性と考えられます。

対数リターン

金融工学の世界では価格の自然対数をとってその差を取ります。時刻 1 と 2 の価格を x_1, x_2 とすると $(x_2 - x_1) / x_1 = x_2 / x_1 - 1$ と書けます。 $x_2 / x_1 = x$ とおくと、

$$x_2 / x_1 = x + 1$$

と書けます。両辺の対数をとると

$$\ln(x_2) - \ln(x_1) = \ln(x + 1)$$

となります。 $\ln(x + 1) \doteq x$ と書けるので、対数リターンは変化率とほぼ等しくなります。

リスク管理の世界では

x_2 が x_1 の周りを循環的に動くとする、

$$x_2 = (1 + x)x_1 \quad \text{または} \quad x_2 = 1 / (1 + x)x_1$$

と書くことができます。両辺の対数をとって和をとると微小時間の世界では和はゼロとなります。したがって、リスクマネジメントの世界では対数リターンの平均値はゼロと考えます。

パネルデータってなんですか？

ある事象が他の事象に影響をおよぼすとき、その事象が影響を及ぼす原因として複数の因子があると考えられます。これらの因子の相互関係を分析する

際に、ある一時点にしぼって横断的分析する方法をクロス・セクション分析といいます。クロス・セクション分析では一時点のデータをクロス集計し、たとえば年齢層間の比較・収入層間の比較を通じて、年齢に応じた消費動向、収入に応じた消費動向を調べていきます。一方、この方法と対比する方法として時間の経過に対して分析する方法もあります。時代と共に消費の動向がどう変わるのかという問題です。そして、この両方の問題を扱うのがパネルデータ分析です。そして、クロス・セクションデータを時系列に並べたのがパネルデータです。

たとえば、売買高と始値と終値の関係を知りたいと思えば、まずは一時点に絞って全株式銘柄についてこれらの関係を調べてみるという方法があります。つぎに、一銘柄ずつ、売買高と始値と終値の関係を時間の経過にともなってダイナミックにどのように変化していくかを調べることもできます。そして、これらの2つの観点から調べることもできます。

本書を読み終えた後の参考文献

Python のプログラミング

Python で始めるプログラミング入門 コロナ社

詳細！ Python3 入門ノート ソーテック社

Python によるデータ分析入門 オライリージャパン(英語版では2版がでました)

Python 言語によるプログラミングイントロダクション第2版 近代科学社

統計学

数理統計学: データ解析の方法 東洋経済

数理統計学の考え方: 推測理論の基礎 岩波

統計学基礎: 日本統計学会編 東京図書

統計学: 日本統計学会編 東京図書

統計解析ハンドブック 朝倉書店

統計分布ハンドブック 朝倉書店

統計学の7原則 パンローリング社

Basic Econometrics by Gujarati

Theory and Problem of Probability and Statistics by Spiegel

AI・機械学習

ゼロから作る Deep Learning オライリージャパン

詳細ディープラーニング マイナビ

Python 機械学習プログラミング インプレス

強化学習 森北出版社

テクニカル分析

シュワッガーのテクニカル分析 パンローリング社

ターゲット流投資の魔術 徳間書店

私は株で2000000ドル儲けた パンローリング社

リスクマネジメント

金融リスクの理論 朝倉書店

入門経済物理学 PHP 研究所

オプションマーケットHBJ出版局

戦略的思考

タイミングの数理 最適停止問題 朝倉書店

戦略的思考とは何か: エール大学式「ゲーム理論」の発想法 TBSブリタニカ

アルゴリズム思考術 早川書房

金融市場

本当にわかる為替相場 日本実業出版社

日本人が知らなかった米国株 SE翔泳社

Nasdaq100 社の徹底研究 講談社

その他

ウォール街のランダムウォーカー 日本経済新聞出版社

ザ・タートル: 投資家たちの士官学校 日経BP

投資苑 パンローリング社

ゾーン: 「勝つ」相場心理学入門 パンローリング社

マーケットの魔術師 パンローリング社

続マーケットの魔術師 パンローリング社

新マーケットの魔術師 パンローリング社

マーケットの魔術師株式編 パンローリング社

リスク: 神々への反逆 上下 日本経済新聞社

その他 2

パナマ文書 角川

物理数学の直感的方法

物理学とはなんだろうか 上下 岩波新書

いかにして問題を解くか 丸善

熱力学の基礎 東京大学出版社

新版量子論の基礎 サイエンス社

ファインマン物理学 岩波書店

EXtreme Finanical Risk by Sornette

Critical Phenomena in Natural Sience by Sornette

Entropy, Large Deviations, and Statistical Mechanics by Ellis