

モデリング

日本アクチュアリー会

2005年4月

モデリング

目 次

序

第 1 章 回帰分析

1.1 回帰分析とは	1-1
1.2 用語の定義	1-1
1.3 最小二乗法（単回帰の場合）	1-3
1.4 決定係数（単回帰の場合）	1-6
1.5 重回帰	1-7
1.6 非線形回帰	1-14
1.7 確率分布の前提を置いた回帰モデルの分析	1-16
1.8 補足	1-26
1.9 参考文献	1-26

第 2 章 時系列解析

2.1 時系列解析とは	2-1
2.2 用語の定義	2-1
2.3 さまざまな時系列モデル	2-3
2.4 時系列モデルの同定・推定・診断・予測	2-18
2.5 参考文献	2-27
図表	2-28

第 3 章 確率過程

3.1 確率過程とは	3-1
3.2 マルコフ連鎖	3-1
3.3 ポアソン過程	3-5
3.4 ブラウン運動	3-8
3.5 モデリングの金融工学への適用（試験範囲外）	3-11
3.6 練習問題	3-17
3.7 練習問題解答	3-18
3.8 参考文献	3-21

第4章 シミュレーション

4.1	乱数の生成法	4-1
4.2	分散減少法	4-31
4.3	検定（試験範囲外）	4-55
4.4	一様乱数の生成（試験範囲外）	4-77
4.5	参考文献	4-108

第5章 線形計画法

5.1	定式化・例と本章の問題の範囲	5-1
5.2	標準形の線形計画問題	5-4
5.3	行列表現	5-6
5.4	頂点と拡大した制約条件行列との関係	5-9
5.5	シンプレックス基準	5-13
5.6	まとめ	5-21
5.7	練習問題	5-21
5.8	参考文献	5-23

序

国際アクチュアリー会の共通シラバスとして「モデリング」が導入されております。我が国での正会員試験にあって、この要素は多分に現行の数学、生保数理、損保数理に含まれておりますが、当会としてI A Aのシラバスに則った対応を行っていることを明確に示すこと、および「モデリング」として別途とらえる範囲を明確にすることを意図して、今回モデリングのテキストを新たに作成しました。

このテキストはアクチュアリー試験のテキストとして供することを目途としています。内容的には、5つの分野から代表的項目を選択いたしました。将来的に項目の増減が発生する可能性はありますが、ある程度一般性のある題材を選択いたしました。

ご覧のとおり選択した5分野はいずれもそれだけで一冊の名著となる内容ですが、これを約200ページに収めたことにより、やや至らぬ記述になっている部分があることは否めません。また、記述に思わぬ誤りが発生していることも危惧されることから、今後アクチュアリー試験の受験者も含めた皆様のご指摘、ご叱責を乞うばかりです。

また、執筆の労をお取りいただいた方々のお名前を下記の通り銘記させていただきお礼にかえたいと存じます。なお、本テキストの責任は各章とも日本アクチュアリー会にあります。また、テキストの内容自体が日本アクチュアリー会の公式見解を示すものではありません。

最後に、本テキストの会合に毎回のように出席され貴重なご意見を頂戴しました若林健一、重原正明の両氏にお礼申し上げますとともに、テキスト部会の歴代の部会長でありました湯浅味代士氏、石井一真氏に本テキストを上梓することができなかったことをお伝えして本序文とさせていただきます。

教育委員会 テキスト部会長
高久 伸一
2005年4月

各章の執筆者（敬称略）

第1章	回帰分析	高澤 俊幸
第2章	時系列解析	高澤 俊幸
第3章	確率過程	勝野健太郎
第4章	シミュレーション	服部 真
第5章	線形計画法	山内 恒人

補正稿の出版について

これまで正誤表で対応していた本文中の誤りについて補正しました。内容に変更は加えておりません。

教育委員会 テキスト部会長
重原正明
2008年7月

第 1 章 回帰分析

1.1 回帰分析とは

回帰分析とは、複数種類のデータの分析をする際に、種類別にデータを分けて分析するのではなく、データ種類間を含めて分析するための手法（一般的に多変量解析と言われているもの）の 1 つであり、あるデータを他のデータの関数によって説明するものである。

この章では確率論や統計学の基礎的な知識を前提として回帰分析の概要について説明する。
なお、一部省略している証明等については章末に示した文献等を参照されたい。

1.2 用語の定義

2 種類のデータ x, y について n 個の観測値 (x_i, y_i) ($i = 1, 2, \dots, n$) が与えられたとき、 x と y の関係を表す最も基本的な指標は相関係数 r_{xy} であり、これは下式によって定義される。

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1.1)$$

(\bar{x} は x_1, x_2, \dots, x_n の平均、 \bar{y} は y_1, y_2, \dots, y_n の平均)

良く知られているように、相関係数 r_{xy} は次の性質を持っている。

性質 1.1 $-1 \leq r_{xy} \leq 1$

性質 1.2 $r_{xy} = 1 \Leftrightarrow (x_1, y_1), \dots, (x_n, y_n)$ が (x, y) 平面上の右上がりの直線上にある

$r_{xy} = -1 \Leftrightarrow (x_1, y_1), \dots, (x_n, y_n)$ が (x, y) 平面上の右下がりの直線上にある

性質 1.3 r_{xy} が 1 に近い \Leftrightarrow 右上がりの直線的な関係が強い（「正の相関が強い」という）

r_{xy} が -1 に近い \Leftrightarrow 右下がりの直線的な関係が強い（「負の相関が強い」という）

性質 1.4 観測データの単位を変えても（正の定数を乗じても）相関係数は変わらない

観測データに定数を加えても相関係数は変わらない

問題 1.1 相関係数 r_{xy} が性質 1.1、性質 1.2 および性質 1.4 を満たすことを確認せよ。

解答 1.1 性質 1.4 を満たすことは定義式より明らかであるので、 $x_i - \bar{x}$ および $y_i - \bar{y}$ を改めて x_i, y_i として、 $\bar{x} = \bar{y} = 0$ という仮定の下で相関係数 r_{xy} が性質 1.1 および性質 1.2 を満たすことを示せば良い。

t を実数とし、 z_i を $z_i = y_i - tx_i$ ($i = 1, 2, \dots, n$) として定義すると、その二乗和は、

$$0 \leq \sum_{i=1}^n z_i^2 = \sum_{i=1}^n (y_i - tx_i)^2 = \sum_{i=1}^n y_i^2 - 2t \sum_{i=1}^n x_i y_i + t^2 \sum_{i=1}^n x_i^2 \quad (1.2)$$

であり、任意の t に対して右辺の t の二次式は非負値であるので、判別式の値は 0 以下となるはずである。

よって、

$$\left(\sum_{i=1}^n x_i y_i\right)^2 - \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right) \leq 0 \Leftrightarrow -1 \leq \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}} \leq 1 \quad (1.3)$$

となり、性質 1.1 が成り立つ。

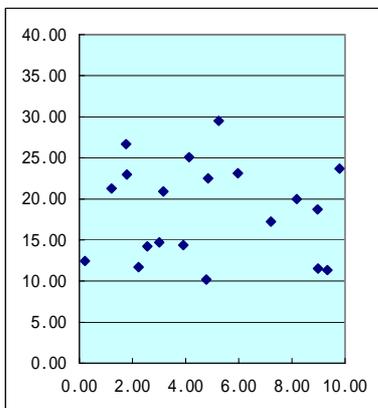
また、相関係数 r_{xy} が ± 1 となる場合、つまり判別式の値が 0 となる場合は、(1.2)式がある t_0 に対して 0 となる場合であるので、 $y_i - t_0 x_i = 0$ ($i = 1, 2, \dots, n$) でなければならない。逆に、 $y_i - t_0 x_i = 0$ ($i = 1, 2, \dots, n$) であれば相関係数 r_{xy} が ± 1 となることは明らかであろう。よって、相関係数 r_{xy} が ± 1 となることと、 (x_i, y_i) が (x, y) 平面上の直線上に乗ることは同値であることが分かる。

r_{xy} の符号によって直線が右上がりとなるか右下がりとなるかについては、 r_{xy} の符号が $\sum_{i=1}^n x_i y_i$ の符号に一致すること、さらに $\sum_{i=1}^n x_i y_i$ の符号は t_0 の符号に一致することから明らかであろう。

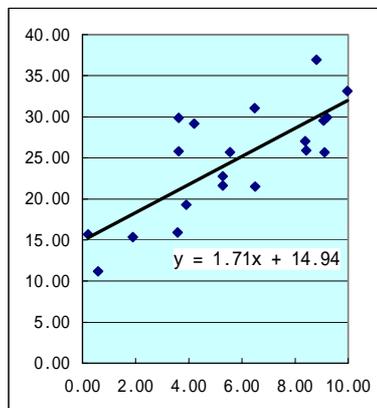
以下に相関係数の例を示しておく。

この例から、相関係数が x と y の直線的な関係の強さを表していることを確認することができるであろう。

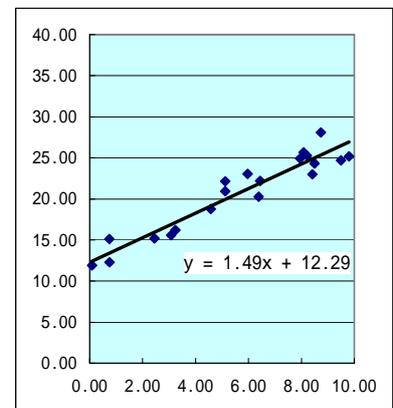
【 散布図と相関係数 (例) 】



相関係数 = -0.04



相関係数 = +0.76



相関係数 = +0.95

問題 1.2 (x, y) のデータが下表のとおり与えられている時、 x と y の相関係数を算出せよ。

x	1.6	1.7	3.4	3.3	5.3
y	2.4	4.8	15.2	12.9	15.8

解答 1.2 相関係数の定義式(1.1)に従って計算すれば良い。計算結果は 0.90 となる。

なお、定義式を式変形して得られる下式を用いる方が便利なこともある。もちろん算出結果は同じ値となる。

$$r_{xy} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right)}} \quad (1.4)$$

相関係数が1（もしくは-1）に近く、正の相関（負の相関）が認められるとき、一方の変数を他方の変数の一次関数として表すことが考えられる。

つまり、 y を x の関数として考えるならば、 $y = \alpha + \beta x$ というモデルとなるが、このモデルを（線形）回帰モデル、このモデルを用いた分析を回帰分析という。また、パラメータ α, β を回帰係数という。

上記のように y を x の関数として考えるとき、 x を説明変数（または独立変数） y を被説明変数（または従属変数）といい、 y を x によって説明することを「 y を x に回帰する」という。

ここでは説明変数が1つの場合を考えているが、観測データが (x_{1i}, x_{2i}, y_i) ($i = 1, 2, \dots, n$) として3種類あり、 $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ というモデル（ y を x_1 と x_2 に回帰するモデル）を考えることもできる。このような説明変数が複数ある場合を重回帰と呼び、これに対して説明変数が1つである場合を単回帰と呼ぶ。

α, β, \dots は観測値を用いて推定することになるが、推定された $\hat{\alpha}, \hat{\beta}, \dots$ を用いた関係式を回帰式（もしくは回帰直線）という。

注 1.1 相関係数 r_{xy} は x と y の直線的関係を測る尺度であり、この値から非線形の関係を読みとることはできない。

例えば、 $y = x^2$ という関係にある5つの観測値 $(-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4)$ について考えてみると、データを (x, y) 平面にプロットして見ても明らかに何らかの関係があることが見て取れるが、相関係数はゼロとなる。

注 1.2 観測値を時系列で把握する場合、時間の経過に伴う上昇トレンドによって相関係数 r_{xy} が1（もしくは-1）に近い値となってしまうことがある。このような場合には、まず、説明変数と被説明変数をそれぞれデータ時点 t に回帰し、観測値と説明変数が t である回帰直線との差を改めて (x_i, y_i) と置いて相関係数を算出することによって、時間の経過に伴う影響を除いた相関係数を算出することができる。（観測値にトレンドを与えていると考えられる他の要因（この例ではデータ時点 t ）に説明変数と被説明変数をそれぞれ回帰し、その影響を除いて算出された相関係数を偏相関係数という。）

1.3 最小二乗法（単回帰の場合）

ここでは単回帰に限定して、観測値 (x_i, y_i) ($i = 1, 2, \dots, n$) を用いた回帰係数の推定について

考える。

この場合は、 n 個の観測値に対して、 $y_i = \alpha + \beta x_i$ ($i = 1, \dots, n$) という回帰モデルを当てはめ、回帰係数 α, β の推定を行うことになるが、観測値が 1 つの直線上に完全に乗ることはあり得ないので、回帰式の誤差である $y_i - (\alpha + \beta x_i)$ を小さくする α, β を推定値とすることになる。

最小二乗法では、誤差の二乗和である下式 Q を最小にする α, β を推定値とする。

$$Q = \sum_{i=1}^n \{y_i - (\alpha + \beta x_i)\}^2 \quad (1.5)$$

非負値の二次式であるので、これを最小化する α, β は、 α と β で偏微分した式の値をゼロとするはずである。よって、次の連立方程式の解が求める推定値 $\hat{\alpha}, \hat{\beta}$ となる。

$$\begin{cases} \frac{\partial Q}{\partial \alpha} = -2 \sum_{i=1}^n \{y_i - (\alpha + \beta x_i)\} = 0 \\ \frac{\partial Q}{\partial \beta} = -2 \sum_{i=1}^n x_i \{y_i - (\alpha + \beta x_i)\} = 0 \end{cases} \quad (1.6)$$

\Downarrow

$$\begin{cases} \sum_{i=1}^n \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\} = 0 \\ \sum_{i=1}^n x_i \{y_i - (\hat{\alpha} + \hat{\beta} x_i)\} = 0 \end{cases} \quad (1.7)$$

\Downarrow

$$\begin{cases} \sum_{i=1}^n y_i = \hat{\alpha} n + \hat{\beta} \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = \hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 \end{cases} \quad (1.8)$$

$\hat{\alpha}, \hat{\beta}$ に関する連立方程式(1.8)を正規方程式という。

正規方程式(1.8)において $\hat{\alpha}$ を消去し、 $\hat{\beta}$ について解くことによって得られる数式を(1.4)式と同様の式変形を逆行を行うことによって次の結果が得られる。これが最小二乗法による回帰係数の推定値である。

$$\begin{cases} \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \\ \hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} = \frac{r_{xy} s_y}{s_x} \end{cases} \quad (1.9)$$

ここで、 s_x, s_y, s_{xy} は下式によって定義される値であり、この記号を用いれば、相関係数は $r_{xy} = s_{xy} / s_x s_y$ となる。(1.9)式はこの関係を用いて相関係数 r_{xy} を用いて表現したものである。

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad s_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (1.10)$$

上記の方程式から、次の回帰直線の性質を確認することができる。

まず、方程式(1.8)の第一式の両辺を n で割れば、 $\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$ となるので、

性質 1.5 回帰直線は観測データの平均 (\bar{x}, \bar{y}) を通る

が分かる。

次に、回帰式に x_i を代入して得られる値を \hat{y}_i (これを y_i の内挿値という) とし、 y_i と \hat{y}_i の差を $e_i = y_i - \hat{y}_i$ (これを残差という) として、方程式(1.7)に代入すれば、

性質 1.6 $\sum_{i=1}^n e_i = 0$

性質 1.7 $\sum_{i=1}^n x_i e_i = 0$

が分かる。

問題 1.3 問題 1.2 のデータに対して $y = \alpha + \beta x$ というモデルを考える。係数 α, β を最小二乗法によって推定せよ。

解答 1.3 問題 1.2 で相関係数を求めているので、定義式に従って s_x, s_y を求め、(1.9)式を用いて $\hat{\alpha}, \hat{\beta}$ を計算すれば良い。

結果のみ示すと、 $s_x = 1.35, s_y = 5.54$ であり、 $\hat{\alpha} = -1.00, \hat{\beta} = 3.67$ となる。

なお、 s_x, s_y の計算は定義式(1.10)を式変形して得られる下式を用いる方が便利なのこともある。もちろん算出結果は同じ値となる。

$$s_x = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n}}, \quad s_y = \sqrt{\frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n}} \quad (1.11)$$

問題 1.4 (t, x, y) のデータが下表のとおり与えられている時、次の問に答えよ。((x, y) データは問題 1.2 と同じものである。)

t	1.0	2.0	3.0	4.0	5.0
x	1.6	1.7	3.4	3.3	5.3
y	2.4	4.8	15.2	12.9	15.8

- (1) t と x の相関係数を算出せよ。また、 x を t に回帰したモデル $x = \alpha + \beta t$ を考え、係数 α, β を最小二乗法によって推定せよ。
- (2) t と y の相関係数を算出せよ。また、 y を t に回帰したモデル $y = \alpha + \beta t$ を考え、係数 α, β を最小二乗法によって推定せよ。

- (3) 問題(1)および問題(2)で求めた回帰式を用いた各 t_i に対する x, y の値 (内挿値) \hat{x}_i, \hat{y}_i と実データ x_i, y_i との差 $x_i - \hat{x}_i, y_i - \hat{y}_i$ を x'_i, y'_i とし、 x', y' の相関係数を算出せよ。
- また、この結果と問題 1.2 の結果を比較し、差がどのような理由から生じたものであるか考察せよ。

解答 1.4 最小二乗法による推定値の算出および相関係数の算出は前述のとおりであるので結果のみ示しておく。

$$(1) \quad r_{tx} = 0.94, \hat{\alpha} = +0.36, \hat{\beta} = +0.90$$

$$(2) \quad r_{ty} = 0.89, \hat{\alpha} = -0.25, \hat{\beta} = +3.49$$

$$(3) \quad r_{x'y'} = 0.38$$

x', y' は x, y から時間によるトレンドを除いたものであり、 x', y' の相関係数 0.38 は、時間によるトレンドが含まれている x, y の相関係数 0.90 (解答 1.2 参照) よりも小さい値となっている。

これは、(1)および(2)から確認できるとおり、 x, y ともに時間の経過による右上がりのトレンドを持っており、 x, y にはこの 2 変数以外の要因である時間による増加トレンドの影響が含まれることによって、相関関係が強くなってしまっているためと考えられる。

1.4 決定係数 (単回帰の場合)

次に、最小二乗法によって推定した回帰式が、観測値にどれくらい良く当てはまっているのかを測る尺度について考える。

観測値 y_i と回帰直線 (y_i の内挿値 \hat{y}_i) との差が小さければ「良く当てはまっている」と考えられるので、残差 $e_i = y_i - \hat{y}_i$ の二乗和によって評価することが考えられるが、この値は観測値の数が増えれば大きくなるため、観測値の数によって見方を変えなければならない。

この欠点を取り除き、観測値の数によらない尺度としたものが、決定係数と言われるものであり、次のような考えに基づいて定義されている。

観測値を (x_i, y_i) ($i = 1, 2, \dots, n$) とし、 y_i の平均 \bar{y} からの乖離の二乗和 $\sum_{i=1}^n (y_i - \bar{y})^2$ (これを全変動という) を考えると、次のようになる。

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n \{e_i + (\hat{y}_i - \bar{y})\}^2 \\ &= \sum_{i=1}^n e_i^2 + 2\sum_{i=1}^n e_i(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n e_i^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \end{aligned} \quad (1.12)$$

(第 3 式の第 2 項がゼロとなることは、 \hat{y}_i の定義 ($\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$) と、上記の性質 1.6 および性質 1.7 から導くことができる。)

上式の右辺最終式第1項を残差変動、第2項を回帰変動といい、この言葉を用いれば上式は、

$$\text{全変動} = \text{残差変動} + \text{回帰変動}$$

として表すことができる。

残差変動が小さければ「良く当てはまっている」ということになるが、上で述べたように観測値の数によって水準が変わってしまうため、観測値の数によらない適合度の尺度である決定係数 R^2 を下式によって定義する。

$$\text{決定係数 } R^2 = 1 - \frac{\text{残差変動}}{\text{全変動}} = \frac{\text{回帰変動}}{\text{全変動}} \quad (1.13)$$

全変動、残差変動および回帰変動はすべて正の値であるから、上式より決定係数 R^2 は 0 から 1 までの数値であり、1 に近いほど「回帰直線がデータに良く当てはまっている」と考えることができる。

決定係数は観測値と回帰直線の適合度の尺度であるため、観測値の直線的な関係を表す尺度である相関係数と関係がある。

全変動は $\sum_{i=1}^n (y_i - \bar{y})^2 = ns_y^2$ であり、回帰変動は定義式を変形すると、

$$\begin{aligned} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 &= \sum_{i=1}^n \{(\hat{\alpha} + \hat{\beta}x_i) - (\hat{\alpha} + \hat{\beta}\bar{x})\}^2 \\ &= \hat{\beta}^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \hat{\beta}^2 ns_x^2 \\ &= (r_{xy}s_y / s_x)^2 ns_x^2 \\ &= r_{xy}^2 s_y^2 n \end{aligned} \quad (1.14)$$

と表せるので、これらを用いれば決定係数 R^2 は、

$$\text{決定係数 } R^2 = r_{xy}^2 \quad (1.15)$$

となることが分かる。

また、この式および相関係数の性質 1.4 から、観測値の単位を変えても観測値に定数を加えても、決定係数の値は変わらないことが分かる。

問題 1.5 問題 1.2 のデータに対して $y = \alpha + \beta x$ というモデルを考え、最小二乗法によって係数 α, β を推定するとき、このモデルの決定係数を求めよ。

解答 1.5 問題 1.2 で相関係数を求めているので、(1.15)式を用いて計算すれば良い。算出結果は $R^2 = 0.80$ となる。

1.5 重回帰

1.5.1 回帰係数の推定

ここまでは単回帰の場合について考えてきたが、説明変数が2個以上の場合も単回帰と同様に考えることができる。

ここでは、観測値を $(x_{1i}, x_{2i}, \dots, x_{ki}, y_i)$ ($i = 1, 2, \dots, n$) として、

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1.16)$$

という説明変数が k 個の回帰モデルについて考える。

最小二乗法による回帰係数は、誤差の二乗和である下式を最小にするものとして推定される。

$$Q = \sum_{i=1}^n \{y_i - (\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})\}^2 \quad (1.17)$$

計算過程は省略するが、上式 Q を最小にする推定値 $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_k$ は下式によって与えられる。(この結果を導くためには線形代数の知識が必要となる。証明については参考文献3を参照されたい。)

$$\text{最小二乗法による推定値} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (X^T X)^{-1} X^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (1.18)$$

$$\text{ここで、} X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix} \quad (1.19)$$

単回帰のときのように、各パラメータについて明示的な式を示すことはできないが、(1.18)式右辺の行列計算をした結果が最小二乗法による推定値である。

問題 1.6 問題 1.3 の単回帰による係数 α, β の推定値を(1.18)式を用いて求め、問題 1.3 で導いた結果と同じ値となることを確認せよ。

解答 1.6 (1.18)式における各行列は次のようになる。

$$X = \begin{pmatrix} 1 & 1.6 \\ 1 & 1.7 \\ 1 & 3.4 \\ 1 & 3.3 \\ 1 & 5.3 \end{pmatrix}, \quad Y = \begin{pmatrix} 2.4 \\ 4.8 \\ 15.2 \\ 12.9 \\ 15.8 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.6 & 1.7 & 3.4 & 3.3 & 5.3 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 5 & 15.3 \\ 15.3 & 56.0 \end{pmatrix}, \quad (X^T X)^{-1} = \begin{pmatrix} 1.22 & -0.33 \\ -0.33 & 0.11 \end{pmatrix}$$

これらを用いて(1.18)式に従って計算をすれば、 $\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} -1.00 \\ +3.67 \end{pmatrix}$ が導かれる。

さらに、重回帰による回帰式は単回帰の時と同様に次の性質を満たす。

性質 1.8 回帰直線は観測データの平均 $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k, \bar{y})$ を通る

性質 1.9 $\sum_{i=1}^n e_i = 0$

性質 1.10 $\sum_{i=1}^n x_{1i}e_i = 0, \sum_{i=1}^n x_{2i}e_i = 0, \dots, \sum_{i=1}^n x_{ki}e_i = 0$

決定係数に関しても、単回帰の場合と同じ定義式(1.13)で同様の議論ができる。(決定係数 R^2 は0から1までの数値であり、1に近いほど「回帰式がデータに良く当てはまっている」となる。)

以上のように、重回帰についても単回帰と同様の議論ができる部分が多いが、重回帰では単回帰の場合にはなかった考え方がいくつかある。

ここでは、自由度修正決定係数、多重共線性およびダミー変数の3つについて説明する。

1.5.2 自由度修正決定係数

決定係数の定義から明らかなように、ある回帰式に対して説明変数を追加すると、その説明変数が妥当なものであるか否かにかかわらず、決定係数は必ず増加してしまう。このような問題を回避するために、次の自由度修正決定係数 \bar{R}^2 という指標を考えることがある。

$$\text{自由度修正決定係数 } \bar{R}^2 = 1 - \frac{\text{残差変動}/(n-k-1)}{\text{総変動}/(n-1)} \quad (1.20)$$

ここで、 n は観測値の数であり、 k は説明変数の数である。

この値は、説明変数を増やしても、一定水準以上の残差変動の減少がなければ大きくなりやうになっている。

1.5.3 多重共線性

重回帰では複数の説明変数を用いて被説明変数を表すモデルとなるが、説明変数間に相関がある場合、推定された回帰式が直感的に理解しにくい結果となることがある。(例えば、明らかに被説明変数と正の相関があると考えられる説明変数の回帰係数がマイナスになる、など。)数学的に考えれば、説明変数間に相関がある場合は回帰係数の推定値を求める方程式の解が誤差の影響によってぶれやすくなり、完全な相関がある場合(説明変数間に一次従属の関係がある場合)は回帰係数が一意に定まらなくなるが、これが上に述べたような結果を導くものとなる。

例 1.1 ある保険法人代理店 10 店の資本金、従業員数および収入保険料が下表のとおり与えられているものとする。

代理店	資本金 (単位：百万円)	従業員数 (単位：人)	収入保険料 (単位：百万円)
1	115	12	800
2	97	10	720
3	81	8	400
4	87	9	370
5	72	7	310
6	39	4	250
7	53	5	230
8	48	5	200
9	23	2	140
10	28	3	80

直感的に、「資本金が増えれば収入保険料が増える」、「従業員数が増えれば収入保険料が増える」と予想され、実際に、収入保険料を資本金、従業員数それぞれに単回帰すると、

$$\text{収入保険料(百万円)} = -106.49 + 7.10 \times \text{資本金(百万円)}$$

$$\text{収入保険料(百万円)} = -91.59 + 67.94 \times \text{従業員数(人)}$$

となって、直感に適った結果（回帰係数 β の推定値は正の値）が導かれる。

しかし、収入保険料を資本金と従業員数に重回帰すると、

$$\text{収入保険料(百万円)} = -79.63 - 4.54 \times \text{資本金(百万円)} + 110.97 \times \text{従業員数(人)}$$

となり、資本金の回帰係数はマイナスであるから、「資本金が増えると収入保険料は減る」という結果になってしまう。

この例では資本金と従業員数の相関が極めて強いため、このような歪んだ結果が導かれてしまっている。

説明変数間の相関が強い場合を説明変数間に多重共線性があるという。このような不安定さを含ませないように、なるべく多重共線性がないと考えられる説明変数を選ぶことが重要である。

推定された回帰式が直感的に疑わしい場合は、多重共線性を確認するために、説明変数間の相関係数を求め、相関の高い説明変数を除いた回帰式で推定し直すなどする必要がある。

1.5.4 ダミー変数

データをいくつかのグループに分けて考える必要があるとき、また、特殊な要因によって一部のデータが大きく歪んでいるときなど、次の考え方をを用いることによって、これらを考慮した回帰式を求めることができる。

例 1.2 月次で得られるデータ (x_i, y_i) ($i = n$ 年 4 月, n 年 5 月, \dots) を用いて、 $y = \alpha + \beta x$ という回帰式を考えたいが、季節的な影響によって、四半期単位で傾向が変わっていると考えられるものとする。

このとき、ダミー変数 d_1, d_2, d_3 を、

$$d_{1i} = \begin{cases} 1 & (i = \text{第1四半期に属する月}) \\ 0 & (i \neq \text{第1四半期に属する月}) \end{cases}$$

$$d_{2i} = \begin{cases} 1 & (i = \text{第2四半期に属する月}) \\ 0 & (i \neq \text{第2四半期に属する月}) \end{cases}$$

$$d_{3i} = \begin{cases} 1 & (i = \text{第3四半期に属する月}) \\ 0 & (i \neq \text{第3四半期に属する月}) \end{cases}$$

として定義し、このダミー変数を用いて、

$$y = \alpha + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 \quad (1.21)$$

という説明変数が 4 つである重回帰式を考える。

この回帰式によって推定を行えば、

$$\text{第 1 四半期} \Rightarrow y = (\hat{\alpha} + \hat{\beta}_2) + \hat{\beta}_1 x$$

$$\text{第 2 四半期} \Rightarrow y = (\hat{\alpha} + \hat{\beta}_3) + \hat{\beta}_1 x$$

$$\text{第 3 四半期} \Rightarrow y = (\hat{\alpha} + \hat{\beta}_4) + \hat{\beta}_1 x$$

$$\text{第 4 四半期} \Rightarrow y = \hat{\alpha} + \hat{\beta}_1 x$$

となり、定数項のみを変えた四半期毎の回帰式を求めることができる。

このような目的で用いられるダミー変数を定数項ダミーといい、特に季節要素を考慮するために用いられる定数項ダミーは季節ダミーとして知られている。

ここでは、 x の変化に対する y の変化量 β は、各四半期で同じであるという前提となるが、定数項 α を各四半期で固定して β を四半期毎に変えるようなダミー変数の使い方もある。この場合は、上で定義した d_1, d_2, d_3 を用いて、説明変数 (ダミー変数) として $d_1 x, d_2 x, d_3 x$ を追加した、

$$y = \alpha + \beta_1 x + \beta_2 (d_1 x) + \beta_3 (d_2 x) + \beta_4 (d_3 x) \quad (1.22)$$

という重回帰式を考える。

この回帰式によって推定を行えば、

$$\text{第 1 四半期} \Rightarrow y = \hat{\alpha} + (\hat{\beta}_1 + \hat{\beta}_2) x$$

$$\text{第 2 四半期} \Rightarrow y = \hat{\alpha} + (\hat{\beta}_1 + \hat{\beta}_3) x$$

$$\text{第 3 四半期} \Rightarrow y = \hat{\alpha} + (\hat{\beta}_1 + \hat{\beta}_4) x$$

$$\text{第 4 四半期} \Rightarrow y = \hat{\alpha} + \hat{\beta}_1 x$$

となり、回帰係数 $\hat{\beta}$ を四半期毎に変えた回帰式が得られる。

このような目的で用いられる d_1x, d_2x, d_3x のようなダミー変数を係数ダミーという。

例 1.3 データ $(x_i, y_i)(i = 1, 2, \dots, n)$ を用いて $y = \alpha + \beta x$ という回帰式を考えたいが、 $i = i_0$ のデータが特殊な要因によって大きく歪んでいるものとする。

このとき、例 1.3 で説明した定数項ダミーを用いて回帰式を次のように置くことによって、このデータの歪みを考慮した回帰式を推定することができる。考え方は季節ダミーと同じである。

$$y = \alpha + \beta_1 x + \beta_2 d$$

$$d_i = \begin{cases} 1 & (i = i_0) \\ 0 & (i \neq i_0) \end{cases}$$

ここでは2つの例についてのみ考えたが、説明変数間の相乗効果を考える場合、ある時点に境に傾向が変わっていると考えられる時系列データを扱う場合などについても、定数項ダミーと係数ダミーを組み合わせることによって、それらの影響を考慮した回帰式をつくることができる。

ただし、説明変数を増やせば決定係数は大きくなるので、ダミー変数を追加すれば、それが妥当なものであるか否かにかかわらず、決定係数は必ず増加してしまう。ダミー変数の設定に関しても、説明変数の追加と同様に、それが妥当なものであるか否かについては別途考察が必要である。

問題 1.7 (x, y) のデータが下表のとおり与えられている。定数項ダミーを用いて奇数時点と偶数時点で定数項 α を変えた回帰式 $y = \alpha + \beta x$ を推定せよ。

t	1	2	3	4	5	6
x	0.9	1.8	2.7	4.0	5.0	5.6
y	4.6	13.6	10.9	20.5	17.6	25.9

解答 1.7 ダミー変数 d を次のように定義する。

$$d_i = \begin{cases} 1 & (i = \text{奇数}) \\ 0 & (i = \text{偶数}) \end{cases}$$

このダミー変数を用い、回帰式を $y = \alpha + \beta_1 x + \beta_2 d$ としてパラメータ α, β_1, β_2 を推定すれば良い。重回帰であるため、行列によって表現された(1.18)式を用いて計算する。

まず、(1.18)式で用いる行列を計算すると次のようになる。

$$X = \begin{pmatrix} 1 & 0.9 & 1 \\ 1 & 1.8 & 0 \\ 1 & 2.7 & 1 \\ 1 & 4.0 & 0 \\ 1 & 5.0 & 1 \\ 1 & 5.6 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 4.6 \\ 13.6 \\ 10.9 \\ 20.5 \\ 17.6 \\ 25.9 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.9 & 1.8 & 2.7 & 4.0 & 5.0 & 5.6 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 6.0 & 20.0 & 3.0 \\ 20.0 & 83.7 & 8.6 \\ 3.0 & 8.6 & 3.0 \end{pmatrix}, \quad (X^T X)^{-1} = \begin{pmatrix} 1.25 & -0.24 & -0.56 \\ -0.24 & 0.06 & 0.06 \\ -0.56 & 0.06 & 0.72 \end{pmatrix}$$

これらを用いて(1.18)式に従って計算をすれば、 $\begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 7.87 \\ 3.19 \\ -5.99 \end{pmatrix}$ が導かれる。

よって、奇数時点の回帰式は $y = \hat{\alpha} + \hat{\beta}_1 x + \hat{\beta}_2 = 1.88 + 3.19x$ 、偶数時点の回帰式は $y = \hat{\alpha} + \hat{\beta}_1 x = 7.87 + 3.19x$ となる。

問題 1.8 (x, y) のデータが下表のとおり与えられている。係数ダミーを用いて奇数時点と偶数時点で係数 β を変えた回帰式 $y = \alpha + \beta x$ を推定せよ。

t	1	2	3	4	5	6
x	1.0	2.0	3.0	3.6	5.2	5.7
y	5.2	11.7	11.6	22.9	17.5	32.8

解答 1.8 変数 d を次のように定義する。

$$d_i = \begin{cases} 1 & (i = \text{奇数}) \\ 0 & (i = \text{偶数}) \end{cases}$$

この変数 d を用いたダミー変数（係数ダミー）を dx とし、これを含めた回帰式 $y = \alpha + \beta_1 x + \beta_2(dx)$ のパラメータ α, β_1, β_2 を推定すれば良い。問題 1.7 と同様に、行列によって表現された(1.18)式を用いて計算する。

まず、(1.18)式で用いる行列を計算すると次のようになる。

$$X = \begin{pmatrix} 1 & 1.0 & 1.0 \\ 1 & 2.0 & 0 \\ 1 & 3.0 & 3.0 \\ 1 & 3.6 & 0 \\ 1 & 5.2 & 5.2 \\ 1 & 5.7 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 5.2 \\ 11.7 \\ 11.6 \\ 22.9 \\ 17.5 \\ 32.8 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1.0 & 2.0 & 3.0 & 3.6 & 5.2 & 5.7 \\ 1.0 & 0 & 3.0 & 0 & 5.2 & 0 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 6.00 & 20.50 & 9.20 \\ 20.50 & 86.49 & 37.04 \\ 9.20 & 37.04 & 37.04 \end{pmatrix}, \quad (X^T X)^{-1} = \begin{pmatrix} 0.88 & -0.20 & -0.02 \\ -0.20 & 0.07 & -0.02 \\ -0.02 & -0.02 & 0.05 \end{pmatrix}$$

これらを用いて(1.18)式に従って計算をすれば、 $\begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1.99 \\ 5.47 \\ -2.42 \end{pmatrix}$ が導かれる。

よって、奇数時点の回帰式は $y = \hat{\alpha} + (\hat{\beta}_1 + \hat{\beta}_2)x = 1.99 + 3.04x$ 、偶数時点の回帰式は $y = \hat{\alpha} + \hat{\beta}_1 x = 1.99 + 5.47x$ となる。

1.6 非線形回帰

通常「回帰分析」という場合は線形回帰分析を指すことが多く、ここまでも線形を前提として説明をしてきたが、実務への応用を考えると、あるタイプの非線形関数であてはめるべきケースが数多くある。非線形関数で表される回帰式すべてについて、ここまでの議論に乗せて考えることはできないが、一部の非線形関数については、変数を適当に変換することによって線形回帰に帰着させることができ、線形回帰分析の議論を適用することができる。

ここでは、線形回帰に変換できる非線形回帰のなかで、応用上よく用いられるものについていくつか紹介しておく。

1.6.1 対数線形モデル $y = \alpha x^\beta$

この関数は微分方程式 $\frac{dy}{y} = \beta \cdot \frac{dx}{x}$ の解として得られるものである。この微分方程式から分かるように、例えば、説明変数の値が1%増加しても、被説明変数の値はその一定割合（例えば0.5%）しか増加しないと考えられる場合に用いられるモデルである。

この回帰モデルは、変数を $y' = \log y, x' = \log x$ とすることによって、 $y' = \log \alpha + \beta x'$ という線形モデルに変換することができる。

1.6.2 指数関数モデル $y = \alpha e^{\beta x}$

この関数は微分方程式 $\frac{dy}{y} = \beta \cdot dx (= \beta x \cdot \frac{dx}{x})$ の解として得られるものであり、被説明変数の増加率が、説明変数の増加率および説明変数の値に比例して大きくなる場合に用いられるモデルである。

この回帰モデルは、変数を $y' = \log y$ とすることによって、 $y' = \log \alpha + \beta x$ という線形モデルに変換することができる。

1.6.3 ロジスティック関数モデル $y = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} (\beta > 0)$

この関数は微分方程式 $\frac{dy}{dx} = \beta y(1-y)$ ($0 < y < 1$) の解として得られるものである。単調増加関数であり、 y が小さいとき傾きが小さく、 y が $1/2$ に近づくに従って傾きが大きくなり、 y が $1/2$ を超えて 1 に近づくと再び傾きが小さくなり傾き 0 に収束する。

生物の繁殖などを分析する際によく用いられるモデルであり、さらに、2項回帰モデルでの回帰式 (1.6.4 参照) にも用いられる。

この回帰モデルは、変数を $y' = \log\left(\frac{y}{1-y}\right)$ とすることによって、 $y' = \alpha + \beta x$ という線形モデルに変換することができる。

1.6.4 2項回帰モデル

ある事象の発生確率 y が何らかの説明変数 x に依存して決まる回帰モデルを考える。

線形回帰モデルでは回帰式の値が $[0,1]$ に収まらないため、取りうる値が $[0,1]$ に収まるような回帰式を考える必要がある。回帰式の値を $[0,1]$ に収めるために、ある確率分布の分布関数 F を用いて $y = F(\alpha + \beta x)$ というモデルとすることが考えられるが、この分布関数としてロジスティック分布 (分布関数がロジスティック関数 $F(x) = e^x / (1 + e^x)$ である確率分布) を用いるものをロジット・モデル、標準正規分布の分布関数を用いるものをプロビット・モデルという。ロジット・モデルは 1.6.3 ロジスティック関数モデル と全く同じものである。

ロジット・モデル、プロビット・モデルともに $y' = F^{-1}(y)$ とすることによって線形モデル $y' = \alpha + \beta x$ に変換することができる。

問題 1.9 (x, y) のデータが下表のとおり与えられている。(y は発生率や付帯率などの「率」であり、0% から 100% までの間の値を取るものとする。)

x	1.4	1.8	2.9	4.3	5.3
y	10.4%	24.6%	52.4%	91.3%	92.6%

この時、ロジットモデルを用いた回帰式 $y = F(\alpha + \beta x)$ を求めよ。

解答 1.9 ロジットモデルでは $y = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} (\beta > 0)$ というモデルを考える。この α, β を推定

するためには、 $y' = \log\left(\frac{y}{1-y}\right)$ として y を変換したデータ y' に対して線形回帰を

行えば良い。

(x, y') のデータは下表のようになるので、このデータを用いて (1.9) 式に従って計算すれば $\hat{\alpha}, \hat{\beta}$ が得られる。

x	1.4	1.8	2.9	4.3	5.3
y'	-2.15	-1.12	0.10	2.35	2.53

結果のみ示すと、 $\hat{\alpha} = -3.53, \hat{\beta} = 1.23$ となるので、求めるロジットモデルによる
 回帰式は、 $y = \frac{e^{-3.53+1.23x}}{1+e^{-3.53+1.23x}}$ となる。

1.7 確率分布の前提を置いた回帰モデルの分析

ここまでは回帰モデルに確率分布の前提を全く置かず、データをどのように扱って回帰係数の推定を行うのか、また、その回帰直線の適合度をどのような指標で測るのかについて説明してきた。

ここでは回帰モデルに対して確率分布を仮定して、回帰係数の推定量およびそれを用いて算出される予測量についての確率的・統計的な分析手法の概要を説明する。単回帰のみについて説明するが、重回帰についても基本的に同じように考えることができる。

1.7.1 推定量の分布

説明変数を x_i ($i = 1, \dots, n$)、誤差項を確率変数 ε_i ($i = 1, \dots, n$) として、被説明変数 Y_i ($i = 1, \dots, n$) は、 x_i, ε_i およびパラメータ α, β を用いた下式によって表されるものとする。

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \dots, n) \quad (1.23)$$

説明変数 x_i は確率変数ではない与えられた値として扱うが、誤差項 ε_i と被説明変数 Y_i は確率変数として考える。この前提に誤解があると以降の議論で混乱してしまうため、何を確率変数と考えているかについては十分理解しておく必要がある。この式から分かるように、誤差項 ε_i の分布が被説明変数 Y_i の分布を決定することになる。

さらに、誤差項 ε_i の分布として次の前提を置く。

前提 誤差項 ε_i は互いに独立に正規分布 $N(0, \sigma^2)$ に従う。

この前提の妥当性は個々のケースによるが、誤差が様々な独立な要因の積み上げによって生じているものと考えれば、中心極限定理より正規分布を仮定することには一定の合理性がある。

正規分布に従う確率変数 X に対して $aX + b$ (a, b は定数) も正規分布に従うので、上記回帰式と誤差項に対する前提から、被説明変数 Y_i の分布が直ちに導かれる。

被説明変数 Y_i は互いに独立に正規分布 $N(\alpha + \beta x_i, \sigma^2)$ に従う。

以降では、観測値を正規分布 $N(\alpha + \beta x_i, \sigma^2)$ からの標本と考えて、最小二乗法による回帰係数 α, β の推定量 $\hat{\alpha}, \hat{\beta}$ (最小二乗推定量という。これらは Y_i を用いて表されるため確率変数である。) がどのような分布に従い、この回帰式を用いて予測を行う場合の予測値の変動幅がどのように評価されるのか等について考える。

さらに、確率変数 Y_i の分布のパラメータは α, β, σ^2 の3つであり、誤差項 ε_i の分散 σ^2 も未知であるという前提で考えるため、 σ^2 の推定等についても考えることになる。

最小二乗推定量 $\hat{\beta}$ の式を変形すると、

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} - \frac{\bar{Y} \sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sum_{i=1}^n c_i Y_i \left(\text{ここで、} c_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \\ &= \beta + \sum_{i=1}^n c_i \varepsilon_i\end{aligned}\tag{1.24}$$

となる。(式変形では、 $\sum_{i=1}^n c_i = 0, \sum_{i=1}^n c_i x_i = 1$ を用いている。この2式が成り立つことは各自確認されたい。)

推定量 $\hat{\beta}$ は誤差項 ε_i の線形和であるから、平均と分散がそれぞれ、

$$E(\hat{\beta}) = E\left(\beta + \sum_{i=1}^n c_i \varepsilon_i\right) = \beta + \sum_{i=1}^n c_i E(\varepsilon_i) = \beta\tag{1.25}$$

$$V(\hat{\beta}) = V\left(\beta + \sum_{i=1}^n c_i \varepsilon_i\right) = \sum_{i=1}^n c_i^2 V(\varepsilon_i) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\tag{1.26}$$

である正規分布に従うことが分かる。

次に定数項 α について考える。最小二乗推定量 $\hat{\alpha}$ は、

$$\begin{aligned}\hat{\alpha} &= \bar{Y} - \hat{\beta}\bar{x} \\ &= \bar{Y} - \left(\sum_{i=1}^n c_i Y_i\right)\bar{x} \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}c_i\right) Y_i\end{aligned}$$

$$= \alpha + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}c_i \right) \varepsilon_i \quad (1.27)$$

となり、 $\hat{\beta}$ と同様に誤差項 ε_i の線形和であることが分かる。(式変形では、 $\sum_{i=1}^n (1/n - \bar{x}c_i) = 1$, $\sum_{i=1}^n (1/n - \bar{x}c_i)x_i = 0$ を用いている。この 2 式が成り立つことは各自確認されたい。)

よって、推定量 $\hat{\alpha}$ は平均と分散がそれぞれ、

$$E(\hat{\alpha}) = E\left(\alpha + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}c_i \right) \varepsilon_i \right) = \alpha + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}c_i \right) E(\varepsilon_i) = \alpha \quad (1.28)$$

$$V(\hat{\alpha}) = V\left(\alpha + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}c_i \right) \varepsilon_i \right) = \sum_{i=1}^n \left(\frac{1}{n} - \bar{x}c_i \right)^2 V(\varepsilon_i) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \quad (1.29)$$

である正規分布に従うことが分かる。

回帰式を用いて予測を行う際に $\hat{\alpha}$ と $\hat{\beta}$ の共分散 $Cov(\hat{\alpha}, \hat{\beta})$ も必要になるので、ここで導いておく。

$$\begin{aligned} Cov(\hat{\alpha}, \hat{\beta}) &= Cov\left(\alpha + \sum_{i=1}^n \left(\frac{1}{n} - c_i \bar{x} \right) \varepsilon_i, \beta + \sum_{i=1}^n c_i \varepsilon_i \right) \\ &= Cov\left(\sum_{i=1}^n \left(\frac{1}{n} - c_i \bar{x} \right) \varepsilon_i, \sum_{i=1}^n c_i \varepsilon_i \right) \\ &= E\left(\left(\sum_{i=1}^n \left(\frac{1}{n} - c_i \bar{x} \right) \varepsilon_i \right) \left(\sum_{i=1}^n c_i \varepsilon_i \right) \right) \\ &= E\left(\sum_{i=1}^n \left(\frac{1}{n} - c_i \bar{x} \right) c_i \varepsilon_i^2 \right) \\ &= \sum_{i=1}^n \left(\frac{1}{n} - c_i \bar{x} \right) c_i E(\varepsilon_i^2) \\ &= \sigma^2 \sum_{i=1}^n \left(\frac{1}{n} - c_i \bar{x} \right) c_i \\ &= \sigma^2 \left(\sum_{i=1}^n \frac{c_i}{n} - \bar{x} \sum_{i=1}^n c_i \right) \\ &= \frac{-\bar{x} \sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned} \quad (1.30)$$

誤差項の分散 σ^2 の推定量 $\hat{\sigma}^2$ としては、通常、回帰式の残差 $e_i = Y_i - (\hat{\alpha} + \hat{\beta}x_i)$ を用いた、

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2 \quad (1.31)$$

が用いられる。 $n-2$ で除しているのは不偏推定量とするためである。

証明は省略するが、推定量 $\hat{\sigma}^2$ は最小二乗推定量 $\hat{\alpha}, \hat{\beta}$ それぞれと独立であり、

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2) \quad (1.32)$$

であることが分かっている。

ここで導いた結果をまとめておく。

- ・ 最小二乗推定量 $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$ は正規分布 $N\left(\alpha, \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)\right)$ に従う。
- ・ 最小二乗推定量 $\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ は正規分布 $N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$ に従う。
- ・ 誤差分散の推定量 $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2$ は最小二乗推定量 $\hat{\alpha}, \hat{\beta}$ と独立であり、 $\frac{(n-2)\hat{\sigma}^2}{\sigma^2}$ はカイ二乗分布 $\chi^2(n-2)$ に従う。

これらの分布を用いて、推定された回帰係数および回帰式を用いた推定値（予測値）について、確率的な変動を考慮した分析を行うことができる。

以降で確率分布を用いて行われる分析をいくつか簡単に紹介する。統計学の用語がいくつか出てくるが、数学のテキスト等を参照されたい。

1.7.2 区間推定

最小二乗法によるパラメータの推定は点推定であり、どの程度ぶれている可能性があるかについては分からないが、誤差項に確率分布の前提を置くことによって、最小二乗推定量の分布が決まるため、この分布を用いて α, β の信頼区間（区間推定）を求めることができる。未知のパラメータ σ^2 が含まれるため、 $\hat{\alpha}, \hat{\beta}$ と独立である σ^2 の推定量 $\hat{\sigma}^2$ を使い、 t 分布から信頼区間を求める。

X が標準正規分布 $N(0,1)$ に従い、 Y が X と独立にカイ二乗分布 $\chi^2(n)$ に従うとき、 $X/\sqrt{Y/n}$ は t 分布 $t(n)$ に従う。これを用いれば、

$$\begin{aligned}
T &= \left(\frac{\hat{\alpha} - \alpha}{\sqrt{V(\hat{\alpha})}} \right) / \sqrt{\left(\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \right) / (n-2)} \\
&= \frac{\hat{\alpha} - \alpha}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}} \sim t(n-2)
\end{aligned} \tag{1.33}$$

であることが分かるから、次の手順で信頼係数 $1 - \varepsilon$ の信頼区間をつくることができる。

【回帰係数 α の信頼区間】

手順1 t 分布 $t(n-2)$ の両側 ε 点($t(n-2)$ に従う T に対して $P(|T| \leq t) = 1 - \varepsilon$ を満たす t の値。上側 $\varepsilon/2$ 点と同じ値である。)を算出
 t 分布の数表を用いる

$$\begin{aligned}
\text{手順2 } |T| \leq t &\Leftrightarrow -t \leq \frac{\hat{\alpha} - \alpha}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}} \leq t \\
&\Leftrightarrow \hat{\alpha} - t \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)} \leq \alpha \leq \hat{\alpha} + t \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}
\end{aligned} \tag{1.34}$$

であるので、標本値および手順1で求めた t の値から算出される上記不等式の両側で挟まれた区間が求める α の信頼区間

回帰係数 β の信頼区間も同様の手順で求めることができる。

問題 1.10 β の信頼区間((1.34)式に相当する不等式)を求めよ。

解答 1.10 α と同様に考えれば良い。

$X \sim N(0,1)$ 、 $Y \sim \chi^2(n)$ (X と Y は独立)であるとき、 $X/\sqrt{Y/n} \sim t(n)$ であるので、これを用いれば、

$$T = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t(n-2) \tag{1.35}$$

が導かれる。

よって、 t を t 分布の両側 ε 点として、

$$|T| \leq t \Leftrightarrow -t \leq \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \leq t$$

$$\Leftrightarrow \hat{\beta} - t \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \leq \beta \leq \hat{\beta} + t \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (1.36)$$

となる。

問題 1.11 問題 1.2 のデータに対して単回帰モデルを考えると、回帰係数 α, β の信頼係数 90% の信頼区間を求めよ。

なお、自由度 3 の t 分布の両側 10% 点(上側 5% 点)を 2.353 として計算すること。

解答 1.11 上で導いた(1.34)式、(1.36)式に従って計算すれば良い。

推定値 $\hat{\alpha}, \hat{\beta}$ は問題 1.3 で求めた $\hat{\alpha} = -1.00, \hat{\beta} = 3.67$ である。

推定値 $\hat{\sigma}^2$ は(1.31)式に従って計算すれば良いが、残差変動 $\sum_{i=1}^n e_i^2$ を計算する必要がある。残差変動 $\sum_{i=1}^n e_i^2$ は直接的に $\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$ を計算しても求めることができるが、(1.13)式および(1.15)式より、

$$\text{残差変動} \sum_{i=1}^n e_i^2 = (1 - \text{決定係数} R^2) \times \text{全変動} = (1 - r_{xy}^2) ns_y^2 \quad (1.37)$$

であるので、問題 1.2 で導いた r_{xy} および問題 1.3 で導いた s_y^2 を用いて、

$$\sum_{i=1}^n e_i^2 = (1 - 0.90^2) \times 5 \times 5.54^2 = 30.38$$

として計算することができる。よって、 $\hat{\sigma}^2 = 30.38 / (5 - 2) = 10.13$ となる。

(1.34)式および (1.36)式に含まれる $\sum_{i=1}^n (x_i - \bar{x})^2$ の値も問題 1.3 で導いた $s_x^2 = 1.35$ を用いて、

$$\sum_{i=1}^n (x_i - \bar{x})^2 = ns_x^2 = 5 \times 1.35^2 = 9.17$$

として求まる。

$\bar{x} = 3.06$ であるので、よって、求める信頼区間はそれぞれ次のようになる。

$$-1.00 - 2.353 \sqrt{10.13 \left(\frac{1}{5} + \frac{3.06^2}{9.17} \right)} \leq \alpha \leq -1.00 + 2.353 \sqrt{10.13 \left(\frac{1}{5} + \frac{3.06^2}{9.17} \right)}$$

$$\Leftrightarrow -9.27 \leq \alpha \leq +7.28$$

$$3.67 - 2.353 \sqrt{\frac{10.13}{9.17}} \leq \beta \leq 3.67 + 2.353 \sqrt{\frac{10.13}{9.17}}$$

$$\Leftrightarrow +1.19 \leq \beta \leq +6.14$$

1.7.3 検定

$\hat{\alpha}, \hat{\beta}$ の分布を用いて仮説検定を行うこともできる。この場合も、未知のパラメータ σ^2 に推定量 $\hat{\sigma}^2$ を用いた t 分布による検定 (t 検定) となる。

与えられた観測値から回帰係数 β の値が β_0 であると考えられるか否かの検定、つまり、帰無仮説を $H_0: \beta = \beta_0$ 、対立仮説を $H_1: \beta \neq \beta_0$ とする検定問題を考える。

帰無仮説 $H_0: \beta = \beta_0$ の下で、問題 1.10 で導いた結果より、

$$T = \frac{\hat{\beta} - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t(n-2) \quad (1.38)$$

となるので、次の手順で有意水準 ε の検定を行うことができる。

【 $H_0: \beta = \beta_0$ $H_1: \beta \neq \beta_0$ の検定】

手順 1 t 分布 $t(n-2)$ の両側 ε 点を算出

手順 2 標本値から算出される上記 T の絶対値が手順 1 の t 値より大きいならば H_0 を棄却、そうでない場合は H_0 を採択

なお、対立仮説 H_1 が $\beta > \beta_0$ もしくは $\beta < \beta_0$ である片側検定の場合は次のようになる。

【 $H_0: \beta = \beta_0$ $H_1: \beta > \beta_0$ の検定】

手順 1 t 分布 $t(n-2)$ の上側 ε 点 ($t(n-2)$ に従う T に対して $P(T \leq t) = 1 - \varepsilon$ を満たす t の値) を算出

手順 2 標本値から算出される上記 T の値が手順 1 の t 値より大きいならば H_0 を棄却、そうでない場合は H_0 を採択

【 $H_0: \beta = \beta_0$ $H_1: \beta < \beta_0$ の検定】

手順 1 t 分布 $t(n-2)$ の上側 ε 点を算出

手順 2 標本値から算出される上記 T の値が、手順 1 の t 値のマイナス符号の値より小さいならば H_0 を棄却、そうでない場合は H_0 を採択

問題 1.12 (x, y) のデータが下表のとおり与えられており、このデータから計算された値が次のとおり与えられている。

x 1.0 1.7 2.9 4.5 5.2 6.0 7.5 8.4 9.2 10.4

y 3.9 6.1 8.5 9.6 12.1 15.5 16.1 16.6 18.7 22.8

$\bar{x} = 5.680, \bar{y} = 12.990, s_x = 3.043, s_y = 5.660, r_{xy} = 0.985, \hat{\alpha} = 2.587, \hat{\beta} = 1.832$

この時、有意水準 5% で次の検定を行え。

なお、自由度 8 の t 分布 $t(8)$ の両側 5% 点 (上側 2.5% 点) を 2.306、上側 5% 点を 1.860 として計算すること。

$$(1) H_0 : \beta = 1.50 \quad H_1 : \beta > 1.50$$

$$(2) H_0 : \alpha = 2.00 \quad H_1 : \alpha \neq 2.00$$

解答 1.12 (1) (1.38)式から T 値を求めて上側 5% 点を用いて評価すれば良い。

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2 = \frac{(1-r_{xy}^2)ns_y^2}{n-2} = \frac{(1-0.985^2) \times 10 \times 5.660^2}{8} = 1.192$$

$$T = \frac{\hat{\beta} - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\hat{\beta} - \beta_0}{\sqrt{\frac{\hat{\sigma}^2}{ns_x^2}}} = \frac{1.832 - 1.50}{\sqrt{\frac{1.192}{10 \times 3.043^2}}} = 2.926$$

よって、 T 値が上側 5% 点 1.860 よりも大きいので H_0 は棄却される。

(2) 同様に T 値を求めて両側 5.0% 点を用いて評価すれば良い。

$$\begin{aligned} \frac{\hat{\alpha} - \alpha_0}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}} &= \frac{\hat{\alpha} - \alpha_0}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{ns_x^2} \right)}} \\ &= \frac{2.587 - 2.00}{\sqrt{1.192 \left(\frac{1}{10} + \frac{5.680^2}{10 \times 3.043^2} \right)}} \\ &= 0.803 \end{aligned}$$

よって、 T 値の絶対値が両側 5% 点 2.306 よりも小さいので H_0 は棄却されない。

(H_0 が採択される。)

1.7.4 点予測および区間予測

説明変数 x_{n+1} が与えられたときの被説明変数 Y_{n+1} の予測量は、回帰式を用いて、

$$\hat{Y}_{n+1} = \hat{\alpha} + \hat{\beta}x_{n+1} \quad (1.39)$$

として与えられる。 $\hat{\alpha}, \hat{\beta}$ とともに $Y_i (i=1, \dots, n)$ の線形和であるから、 \hat{Y}_{n+1} も Y_i の線形和であり、かつ Y_i は互いに独立であるので、 \hat{Y}_{n+1} は正規分布に従うことが分かる。

次に、予測誤差 $Y_{n+1} - \hat{Y}_{n+1}$ について考えると、

$$Y_{n+1} - \hat{Y}_{n+1} = (\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_{n+1} + \varepsilon_{n+1} \quad (1.40)$$

となるから、予測誤差も正規分布に従い、その平均および分散は、

$$\begin{aligned} E(Y_{n+1} - \hat{Y}_{n+1}) &= E\{(\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_{n+1} + \varepsilon_{n+1}\} \\ &= E(\alpha - \hat{\alpha}) + x_{n+1}E(\beta - \hat{\beta}) + E(\varepsilon_{n+1}) \end{aligned}$$

$$= 0 \tag{1.41}$$

$$\begin{aligned}
V(Y_{n+1} - \hat{Y}_{n+1}) &= V\{(\alpha + \beta x_{n+1} + \varepsilon_{n+1}) - (\hat{\alpha} + \hat{\beta} x_{n+1})\} \\
&= V(\hat{\alpha} + \hat{\beta} x_{n+1}) + V(\varepsilon_{n+1}) \\
&= V(\hat{\alpha}) + V(\hat{\beta} x_{n+1}) + 2\text{Cov}(\hat{\alpha}, \hat{\beta} x_{n+1}) + \sigma^2 \\
&= V(\hat{\alpha}) + x_{n+1}^2 V(\hat{\beta}) + 2x_{n+1} \text{Cov}(\hat{\alpha}, \hat{\beta}) + \sigma^2 \\
&= \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\} + x_{n+1}^2 \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + 2x_{n+1} \frac{-\bar{x}\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \sigma^2 \\
&= \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{\bar{x}^2 + x_{n+1}^2 - 2x_{n+1}\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\} \\
&= \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\} \tag{1.42}
\end{aligned}$$

となる。

分散は常に σ^2 よりも大きな値になるが、これは Y_{n+1} の分散（誤差項の分散に一致するので σ^2 ）に加えて予測量 \hat{Y}_{n+1} の分散が含まれるためである。 n が十分大きい場合は（例外的な場合を除いて）回帰係数の推定量が真の値に十分近くなるため、誤差項の分散 σ^2 に収束する式になっている。

この予測誤差の分布を用いれば、 Y_{n+1} の区間予測を行うことができる。考え方は α, β の区間推定と同様であり、未知のパラメータ σ^2 に推定量 $\hat{\sigma}^2$ を用いた t 分布から求める。

問題 1.13 問題 1.12 のデータに対して推定された回帰式を用いて $x_{n+1} = 13.0$ に対する点予測および信頼係数 5% の区間予測を求めよ。

なお、平均などの計算値は問題 1.12 で示されているものを用い、自由度 8 の t 分布 $t(8)$ の両側 5% 点は 2.306 とすること。

解答 1.13 点予測は推定した回帰式に $x_{n+1} = 13.0$ を代入すれば良いので、

$$\hat{Y}_{n+1} = \hat{\alpha} + \hat{\beta} x_{n+1} = 2.587 + 1.832 \times 13.0 = 26.403$$

となる。

区間予測については、上で導いた結果より、

$$Y_{n+1} - \hat{Y}_{n+1} \sim N \left(0, \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\} \right) \tag{1.43}$$

であり、 $\hat{\sigma}^2$ は $\hat{\alpha}, \hat{\beta}, \varepsilon_{n+1}$ と独立、つまり $Y_{n+1} - \hat{Y}_{n+1}$ と独立であるので、

$$T = \left(\frac{Y_{n+1} - \hat{Y}_{n+1}}{\sqrt{V(Y_{n+1} - \hat{Y}_{n+1})}} \right) / \sqrt{\left(\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \right) / (n-2)}$$

$$= \frac{Y_{n+1} - \hat{Y}_{n+1}}{\sqrt{\hat{\sigma}^2 \left\{ 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}}} \sim t(n-2)$$

(1.44)

であることが分かるので、この分布を用いれば良い。

求める信頼係数 5% の区間予測は、

$$-2.306 \leq \frac{Y_{n+1} - \hat{Y}_{n+1}}{\sqrt{\hat{\sigma}^2 \left\{ 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}}} \leq +2.306$$

$$\Leftrightarrow \hat{Y}_{n+1} - 2.306 \times \sqrt{\hat{\sigma}^2 \left\{ 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}} \leq Y_{n+1} \leq \hat{Y}_{n+1} + 2.306 \times \sqrt{\hat{\sigma}^2 \left\{ 1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}} \quad (1.45)$$

に対して、問題 1.12 に示されている計算値および問題 1.12 で計算した $\hat{\sigma}^2$ を代入すれば良い。結果は、 $23.14 \leq Y_{n+1} \leq 29.67$ (信頼区間は [23.14, 29.67]) となる。

なお、重回帰の場合についても、単回帰と同様に次の結果が導かれる。(この結果を導くためには線形代数の知識が必要となる。証明については参考文献 3 を参照されたい。)

- $Y_i (i = 1, \dots, n)$ は互いに独立であり、 $N(\alpha + \beta_1 x_{1i} + \dots + \beta_k x_{ki}, \sigma^2)$ に従う。
 - 最小二乗推定量 $\hat{\alpha}$ は正規分布 $N(\alpha, \sigma_0^2)$ に従う。
- ここで、 $\sigma_0^2 = (\sigma^2 (X^T X)^{-1})$ の (1,1) 要素) であり、 X は (1.19) 式で定義した行列である。

・最小二乗推定量 $\hat{\beta}_1$ は正規分布 $N(\beta_1, \sigma_1^2)$ に従う。

ここで、 $\sigma_1^2 = (\sigma^2(X^T X)^{-1})$ の(2,2)要素)である。

⋮

・最小二乗推定量 $\hat{\beta}_k$ は正規分布 $N(\beta_k, \sigma_k^2)$ に従う。

ここで、 $\sigma_k^2 = (\sigma^2(X^T X)^{-1})$ の($k+1, k+1$)要素)である。

・誤差分散の推定量 $\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n-k-1}$ は最小二乗推定量 $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_k$ と独立であり、 $\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2}$

はカイ二乗分布 $\chi^2(n-k-1)$ に従う。

これらを用いて重回帰についても区間推定、検定、点予測および区間予測等を行うことができる。

さらに、重回帰の場合は、個々の回帰係数に関する区間推定等に加えて、説明変数の従属に関する検定、ダミー変数の有意性に関する検定なども行うことができる。

詳細については章末に示している参考文献等を参照されたい。

1.8 補足

この章では一部を除いて「前提としている回帰モデルは正しい」という立場で説明をしてきた。しかし、実際にデータ分析を行う場合は、回帰係数の推定や予測値の算出と同様に、回帰モデルの妥当性検証が重要な作業になる。

この検証に関しても様々な統計的手法が考えられているが、最も簡単かつ直感的な検証方法は残差のプロットである。厳密には残差を用いた統計量による検定となるが、単に残差をプロットするだけでも前提の誤り(歪み)が分かるケースが多いので、憶えておいていただきたい。

1.9 参考文献

1. 涌井良幸・涌井貞美「図解でわかる回帰分析」岩波書店(2002)

確率論や統計学の基礎知識をほとんど前提としない入門書である。

回帰分析の基本的な考え方が平易に分かりやすく説明されており、この一冊で全体像を掴むことができる。

2. 田中勝人「計量経済学」岩波書店(1998)

計量経済学の書籍には回帰分析について分かりやすく説明されているものが多くあり、この本もその一冊。確率論や統計学の初歩的な知識のみを前提に説明されているため、比較的読みやすく、また、ここでは全く述べなかった誤差項の確率分布に対する条件を緩めたモデルなどに

についても詳細に説明されている。

なお、回帰分析以外に時系列解析なども含まれている。

3. 佐和隆光「回帰分析」朝倉書店（1979）

線形代数をベースに回帰分析について説明されており、理論としての十分な説明に加えて、実際問題への応用を見据えた議論も多く含まれている。

難易度が高く、一定の数学的予備知識が必要である。

4. 縄田和満「Excel による回帰分析入門」（2002）

表計算ソフトの Excel を用いて、実際のデータを扱いながら回帰分析を解説している。

コンパクトにまとめられた解説に加えて、Excel によるデータ分析手続も概観することができる。Excel を使って回帰分析の実務への応用を考えている方にとって有益な 1 冊である。

第2章 時系列解析

2.1 時系列解析とは

時系列解析とは、時刻 t をパラメータとして持つ確率変数の列 $\{Y_t\}$ (すなわち、確率過程 $\{Y_t\}$) の実現値である $\{y_t\}$ (時系列データという) を用いて、将来の Y_t がそれよりも前の Y_t および確率的な誤差によって定まるモデル (時系列モデルという) にあてはめ、その時系列の構造の分析や将来予測を行うための手法である。

この章では時刻 t が離散的 (つまり、 $t = 0, \pm 1, \pm 2, \dots$) である時系列について、基礎的な概念およびいくつかの時系列モデルを説明した後、実際に時系列データを用いて特定の時系列モデルにあてはめ、予測を行う際の手順について概要を説明する。

2.2 用語の定義

第1章の回帰分析は「あるデータを他のデータの関数によって説明するもの」であったが、時系列解析ではその時系列データの過去の実現値を手掛かりとする。基本的に時系列データは1時刻で1つの実現値しか得られず、一般的な時系列モデルを対象として考えてしまうとモデルを特定するための情報が足りないため、いくつかの仮定を置いて時系列モデルの範囲を限定することになる。

時系列解析による分析を行う際は、通常、確率過程 $\{Y_t\}$ に対して定常性と言われる次の条件が満たされるものと仮定する。

定義 2.1 確率過程 $\{Y_t\}$ は定常 (定常性を持つ)

条件 1 $E(Y_t) = \mu$ (Y_t の期待値は時刻 t に依存しない定数)

条件 2 $Cov(Y_t, Y_{t-h}) = \gamma_h$ (Y_t の分散は時刻 t に依存しない定数であり、 Y_t と Y_{t-h} の共分散は時刻 t に依存せず時差 h のみに依存する)

このような定常性の前提を置いてしまうと、例えば、実際のデータに多い一定の増加トレンドを含む時系列を考える場合は、期待値が一定とはならないため条件1を満たさず、分析対象外となってしまいうように見える。しかし、そのような時系列についても、階差 $\Delta Y_t = Y_t - Y_{t-1}$ (もしくは2次の階差 $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$) をとることによって定常性の条件1を満たす時系列に変換できる場合がある。

与えられた時系列を階差等によって変換した結果が定常性を持つと考えられる場合も分析対象となるため、対象となる時系列の範囲は十分に広いものである。

問題 2.1 指数関数的な増加トレンド e^{at+b} を持つ時系列データについては、データをどのように変換すれば定常性の条件1を満たす時系列データに変換できると考えられるか。

解答 2.1 対数変換した後、階差をとれば良い。

次に時系列解析で用いる基礎的な概念をいくつか定義しておく。

定義 2.2 $\gamma_h = Cov(Y_t, Y_{t-h})$ を時差 h の自己共分散という。

定義 2.3 $\rho_h = \frac{\gamma_h}{\gamma_0} \left(= \frac{Cov(Y_t, Y_{t-h})}{V(Y_t)} \right)$ を時差 h の自己相関という。

(h の関数と考えた自己相関 ρ_h をコレログラムともいう。)

$V(Y_t) = V(Y_{t-h}) = \gamma_0$ であることより自己相関 ρ_h は Y_t と Y_{t-h} の相関係数であるので、 $-1 \leq \rho_h \leq 1$ を満たす。

確率過程 $\{Y_t\}$ の実現値である時系列データ $\{y_t\}_{t=1}^n$ から計算されるこれらの値の推定値が、時系列モデルあてはめの際の有力な情報になる。

定義 2.4 $\hat{\gamma}_h = \frac{\sum_{t=h+1}^n (y_t - \bar{y})(y_{t-h} - \bar{y})}{n}$ を時差 h の標本自己共分散という。

定義 2.5 $\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} \left(= \frac{\sum_{t=h+1}^n (y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \right)$ を時差 h の標本自己相関という。

(h の関数と考えた標本自己相関 $\hat{\rho}_h$ を標本コレログラムともいう。)

問題 2.2 時系列データ $\{y_t\}_{t=1}^{10}$ が下表のとおり与えられているとき、標本自己相関 $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$ を求めよ。

t	1	2	3	4	5	6	7	8	9	10
y_t	0.79	1.73	2.29	2.96	3.25	3.58	3.66	3.69	3.83	3.93

解答 2.2 定義 2.5 に従って計算すれば良い。

$n = 10, \bar{y} = 2.97, \hat{\gamma}_0 = 0.99, \hat{\gamma}_1 = 0.61, \hat{\gamma}_2 = 0.32, \hat{\gamma}_3 = 0.06$ であるので、

$\hat{\rho}_1 = 0.62, \hat{\rho}_2 = 0.33, \hat{\rho}_3 = 0.06$ となる。

次にラグ作用素について定義しておく。

定義 2.6 時刻 t をパラメータとして持つ確率変数の時刻を 1 つ前に戻す作用素をラグ作用素といい、 L で表す。

つまり、 $LY_t = Y_{t-1}$ であり、 $L^2 Y_t = L(LY_t) = LY_{t-1} = Y_{t-2}, \dots$ と表記する。

この章では時系列モデルの一表現として用いるだけであるが、このラグ作用素 L は通常の数

と同じように演算を行うことができ、時系列解析における議論をスムーズに展開するために必要不可欠な概念となっている。

2.3 さまざまな時系列モデル

ここでは、基本的な4つの時系列モデルについて説明する。時系列解析ではこれら4つのモデルの理解が出発点となる。

以降で考える確率過程 $\{Y_t\}$ は、特に断らない限り、すべて定常(つまり、期待値および分散は t に依存しない定数 $E(Y_t) = \mu, V(Y_t) = \gamma_0$)であり、自己共分散は時差 h のみに依存する定数 $Cov(Y_t, Y_{t-h}) = \gamma_h$)であるものと仮定する。

2.3.1 自己回帰モデル $AR(p)$

定義 2.7 確率過程 $\{Y_t\}$ が下式を満たすとき、 $\{Y_t\}$ は p 次の自己回帰モデル (Auto-regressive Model) に従うといい、 $\{Y_t\} \sim AR(p)$ と表す。

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (2.1)$$

ここで、パラメータ $\phi_0, \phi_1, \dots, \phi_p$ は定数であり、誤差項である ε_t は $Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ と独立に平均0、分散 σ^2 である t に依存しない同一分布に従う確率変数である。

例えば、ある資産の収益率が一期前の当該資産の収益率水準と確率的なブレの和によって説明できると考えるのであれば、その時系列モデルは $AR(1)$ になる。

p は時刻 t の値がそれ以前のいくつの履歴(および誤差項)によって定まるかを示す次数であり、 p の値が大きいほど複雑なモデルになる。

自己回帰モデルが満たす(2.1)式をラグ作用素 L を用いて表せば、

$$\begin{aligned} Y_t &= \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \\ &= \phi_0 + \phi_1 L Y_t + \phi_2 L^2 Y_t + \cdots + \phi_p L^p Y_t + \varepsilon_t \\ &= \phi_0 + (\phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p) Y_t + \varepsilon_t \\ \{1 - (\phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p)\} Y_t &= \phi_0 + \varepsilon_t \end{aligned}$$

となるから、(2.1)式は、

$$\phi(L) Y_t = \phi_0 + \varepsilon_t \quad (2.2)$$

$$\text{ここで、} \phi(L) = 1 - (\phi_1 L + \phi_2 L^2 + \cdots + \phi_p L^p)$$

と表現することができる。

(2.1)式(もしくは(2.2)式)を満たす確率過程 $\{Y_t\}$ すべてが必ずしも定常性を満たすとは限らないが、次の条件を満たせば、定常性を持つことが分かっている。

命題 2.1 自己回帰モデル $AR(p)$ の特性方程式 $\phi(x) = 1 - (\phi_1 x + \phi_2 x^2 + \dots + \phi_p x^p) = 0$ の解の絶対値がすべて1より大きいとき、 $AR(p)$ は定常性を持つ。

この条件の意味するところを $p=1$ の場合について考えてみると次のようになる。

定常性を仮定せず、 $p=1$ の(2.1)式のみを用いれば、 $Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$ に対して $Y_{t-1} = \phi_0 + \phi_1 Y_{t-2} + \varepsilon_{t-1}$, $Y_{t-2} = \phi_0 + \phi_1 Y_{t-3} + \varepsilon_{t-2}$, \dots を代入することによって、

$$\begin{aligned} Y_t &= \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t \\ &= \phi_0(1 + \phi_1) + \phi_1^2 Y_{t-2} + (\varepsilon_t + \phi_1 \varepsilon_{t-1}) \\ &= \phi_0(1 + \phi_1 + \phi_1^2) + \phi_1^3 Y_{t-3} + (\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2}) \\ &\quad \vdots \\ &= \phi_0(1 + \phi_1 + \phi_1^2 + \dots + \phi_1^{n-1}) + \phi_1^n Y_{t-n} + (\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots + \phi_1^{n-1} \varepsilon_{t-n}) \end{aligned}$$

が得られる。

$p=1$ の場合の特性方程式の解は $1/\phi_1$ であり、この絶対値が1より大きいということは、 $-1 < \phi_1 < 1$ であることを意味する。 $-1 < \phi_1 < 1$ であれば、 n が大きくなる時、上式第2項はゼロに収束するので、 ε_t に対する仮定から Y_t の平均および分散は時刻 t に依存しない一定値(それぞれ公比 ϕ_1, ϕ_1^2 の等比級数の和)となり、さらに、自己共分散が時刻 t に依存しないことも導かれる。逆に、特性方程式の解の絶対値が1以下である場合、つまり $\phi_1 \leq -1$ もしくは $1 \leq \phi_1$ の場合は、遠い過去の影響が増幅して積み重なり、平均も分散も発散するであろうことが上式から読み取れる。

$p \geq 2$ の場合は行列に関する知識が必要となるため詳細な説明は省略するが、基本的な考え方は同じであり、特性方程式の解の絶対値が1より大きければ、平均、分散、自己共分散は時刻 t に依存しない値となり、特性方程式の解の絶対値が1以下であれば、遠い過去の影響が増幅して積み重なり、平均、分散は発散することになる。

次に、時系列データ $\{y_t\}_{t=1}^n$ を $AR(p)$ にあてはめる際に、データから計算される標本自己共分散 $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \dots$ や標本自己相関 $\hat{\rho}_1, \hat{\rho}_2, \dots$ を用いてパラメータ $\phi_0, \phi_1, \dots, \phi_p$ を推定する必要があるため、これらの関係を示しておく。

(2.1)式およびその両辺の期待値を取った式を2つ並べ、2式の対応する項を差し引きすると、定常性および ε_t に対する仮定より、

$$\begin{aligned} Y_t &= \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \\ \mu &= \phi_0 + \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu \\ Y_t - \mu &= \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + \dots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t \end{aligned} \tag{2.3}$$

となる。(第2式より $\mu = \phi_0 / (1 - \phi_1 - \phi_2 - \dots - \phi_p)$ である。)

ここで、(2.3)式の両辺に $(Y_t - \mu), (Y_{t-1} - \mu), \dots, (Y_{t-p} - \mu), \dots$ を乗じて期待値を取れば、定常性の仮定および ε_t と $Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ が独立であることより、

【平均】

$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$ の両辺の期待値を取ることにより、 $\mu = \phi_0 + \phi_1 \mu$ が得られるので、これを μ について解けば、

$$\mu = \phi_0 / (1 - \phi_1) \tag{2.11}$$

となる。

【分散・自己共分散】

(2.4)式および(2.5)式より、 γ_0, γ_1 の連立方程式が得られるので、これを解けば、

$$\begin{cases} \gamma_0 = \phi_1 \gamma_1 + \sigma^2 \\ \gamma_1 = \phi_1 \gamma_0 \end{cases} \quad \begin{cases} \gamma_0 = \sigma^2 / (1 - \phi_1^2) \\ \gamma_1 = \phi_1 \sigma^2 / (1 - \phi_1^2) \end{cases}$$

となる。

さらに、(2.8)以降の式を用いれば、

$$\begin{aligned} \gamma_2 &= \phi_1 \gamma_1 = \phi_1^2 \sigma^2 / (1 - \phi_1^2) \\ \gamma_3 &= \phi_1 \gamma_2 = \phi_1^3 \sigma^2 / (1 - \phi_1^2) \\ &\vdots \\ \gamma_h &= \phi_1 \gamma_h = \phi_1^h \sigma^2 / (1 - \phi_1^2) \\ &\vdots \end{aligned}$$

として、任意の h に対する自己共分散を求めることができる。

なお、ここでは $t = 0, 1, 2, \dots$ として考えているが、負の t に対しては定常性の条件より $\gamma_h = \gamma_{-h}$ となるので、以上をまとめると、

$$\gamma_h = \phi_1^{|h|} \sigma^2 / (1 - \phi_1^2) \quad h = 0, \pm 1, \pm 2, \dots \tag{2.12}$$

となる。

【自己相関】

上で求めた分散および自己共分散を用いれば、

$$\rho_h = \gamma_h / \gamma_0 = \phi_1^{|h|} \quad h = 0, \pm 1, \pm 2, \dots \tag{2.13}$$

となる。

定常性の仮定の下では $-1 < \phi_1 < 1$ であるから、コレログラムは指数関数的にゼロに収束する形状となる。

なお、ユール=ウォーカー方程式を用いれば、分散および自己共分散を用いることなく、この結果をただちに導くことができるので各自確認されたい。

問題 2.3 AR(1) モデル $Y_t = 2.5 + 0.5Y_{t-1} + \varepsilon_t$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.6$) について、 $\{Y_t\}$ の平均 μ 、分散 γ_0 、自己共分散 γ_1, γ_2 および自己相関 ρ_1, ρ_2 を求めよ。

解答 2.3 (2.11)式、(2.12)式および(2.13)式を用いて計算すれば良い。

$$\begin{aligned} \mu &= \phi_0 / (1 - \phi_1) = 2.5 / (1 - 0.5) = 5.00 \\ \gamma_h &= \phi_1^{|h|} \sigma^2 / (1 - \phi_1^2) = 0.5^{|h|} \times 0.6 / (1 - 0.5^2) = 0.80 \times 0.5^{|h|} \end{aligned}$$

$$\begin{aligned} &\Rightarrow \gamma_0 = 0.80, \gamma_1 = 0.40, \gamma_2 = 0.20 \\ \rho_h &= \phi_1^{|h|} = 0.5^{|h|} \\ &\Rightarrow \rho_1 = 0.50, \rho_2 = 0.25 \end{aligned}$$

問題 2.4 AR(1) モデル $Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$ の平均 μ 、分散 γ_0 および自己共分散 γ_1 が下記のとおり与えられているとき、パラメータ ϕ_0, ϕ_1 および ε_t の分散 σ^2 を求めよ。

$$\mu = 4.0, \quad \gamma_0 = 0.6, \quad \gamma_1 = 0.3$$

解答 2.4 自己相関の定義より $\rho_1 = \gamma_1 / \gamma_0 = 0.3 / 0.6 = 0.50$ であり、(2.13)式より $\rho_1 = \phi_1$ であるので、 $\phi_1 = 0.50$ である。

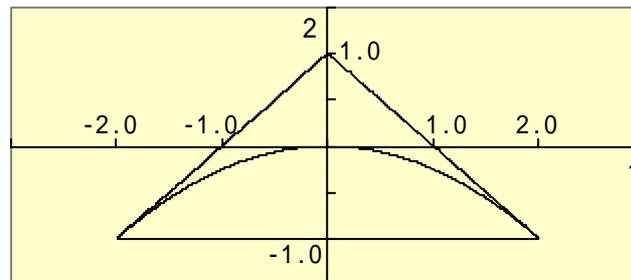
また、(2.11)式より $\phi_0 = \mu(1 - \phi_1)$ であるので、 $\phi_0 = 4.0 \times (1 - 0.5) = 2.00$ となる。分散 σ^2 については、(2.4)式より $\sigma^2 = \gamma_0 - \gamma_1 \phi_1$ という関係式が成り立つため、 $\sigma^2 = 0.6 - 0.3 \times 0.50 = 0.45$ として導かれる。

$$AR(2) (Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t)$$

【定常性の条件】

特性方程式 $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 = 0$ が実数解を持つ場合と虚数解を持つ場合に分け、それぞれについてさらに場合分けをして解の絶対値が1より大きい条件を (ϕ_1, ϕ_2) 平面上に図示すると次のようになる。(三角形の内部が求める領域であり、放物線で区切られた上側が実数解、下側が虚数解に対応する領域となっている。)

図 2.1



【平均】

$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ の期待値を取ることで、 $\mu = \phi_0 + \phi_1 \mu + \phi_2 \mu$ が得られるので、これを μ について解けば、

$$\mu = \phi_0 / (1 - \phi_1 - \phi_2) \quad (2.14)$$

となる。

【分散・自己共分散】

(2.4)式、(2.5)式および(2.6)式より、 $\gamma_0, \gamma_1, \gamma_2$ の連立方程式が得られるので、これを解けば、

$$\begin{cases} \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2 \\ \gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1 \\ \gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0 \end{cases} \quad \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 1 & -\phi_1 & -\phi_2 \\ -\phi_1 & 1-\phi_2 & 0 \\ -\phi_2 & -\phi_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \end{pmatrix} \quad (2.15)$$

となる。

γ_0 のみ計算結果を示すと次のようになる。

$$\gamma_0 = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)\{(1-\phi_2)^2 - \phi_1^2\}} \quad (2.16)$$

さらに、(2.8)以降の式を用いれば、

$$\begin{aligned} \gamma_3 &= \phi_1 \gamma_2 + \phi_2 \gamma_1 \\ \gamma_4 &= \phi_1 \gamma_3 + \phi_2 \gamma_2 \\ &\vdots \\ \gamma_h &= \phi_1 \gamma_{h-1} + \phi_2 \gamma_{h-2} \\ &\vdots \end{aligned}$$

として、任意の h に対する自己共分散を求めることができる。(負の t に対しては定常性の条件より $\gamma_h = \gamma_{-h}$ となるので、これを用いて求めることができる。)

【自己相関】

上で求めた分散および自己共分散を用いて、定義式 $\rho_h = \gamma_h / \gamma_0$ から求めることができる。

なお、ユール=ウォーカー方程式を用いれば、上記の結果を用いなくても、

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_1 \\ \rho_2 = \phi_1 \rho_1 + \phi_2 \end{cases} \quad \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} 1-\phi_2 & 0 \\ -\phi_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad (2.17)$$

$$\begin{cases} \rho_1 = \phi_1 / (1-\phi_2) \\ \rho_2 = \phi_1^2 / (1-\phi_2) + \phi_2 \end{cases}$$

として自己相関を直接求めることができる。

さらに、(2.8)以降の式の両辺を γ_0 で割って導かれる $\rho_h = \phi_1 \rho_{h-1} + \phi_2 \rho_{h-2}$ ($h=3,4,5,\dots$)を用いて順次 ρ_h を求めることができる。(負の t に対しては $\rho_h = \rho_{-h}$ を用いて求める。)

この場合も、定常性の仮定の下で、指数関数的にゼロに収束するコレログラムとなる。

問題 2.5 $AR(2)$ が定常性を満たすための条件が (ϕ_1, ϕ_2) 平面上の図 2.1 三角形内部で表されることを確認せよ。

解答 2.5 各自確認されたい。

問題 2.6 AR(2) モデル $Y_t = 2.0 + 0.5Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.6$) について、定常性を満たすことを確認し、 $\{Y_t\}$ の平均 μ 、分散 γ_0 、自己共分散 $\gamma_1, \gamma_2, \gamma_3$ および自己相関 ρ_1, ρ_2, ρ_3 を求めよ。

解答 2.6 定常性は $(\phi_1, \phi_2) = (0.5, 0.1)$ が図 2.1 の三角形の内部に入ることから確認できる。次に、(2.14)式、(2.16)式および(2.17)式等を用いれば、

$$\mu = \phi_0 / (1 - \phi_1 - \phi_2) = 2.0 / (1 - 0.5 - 0.1) = 5.0$$

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)\{(1 - \phi_2)^2 - \phi_1^2\}} = \frac{(1 - 0.1) \times 0.6}{(1 + 0.1)\{(1 - 0.1)^2 - 0.5^2\}} = 0.88$$

$$\rho_1 = \phi_1 / (1 - \phi_2) = 0.5 / (1 - 0.1) = 0.56$$

$$\rho_2 = \phi_1^2 / (1 - \phi_2) + \phi_2 = 0.5^2 / (1 - 0.1) + 0.1 = 0.38$$

$$\rho_3 = \phi_1\rho_2 + \phi_2\rho_1 = 0.5 \times 0.38 + 0.1 \times 0.56 = 0.24$$

が導かれる。

自己共分散 $\gamma_1, \gamma_2, \gamma_3$ については、 $\rho_h = \gamma_h / \gamma_0$ から求めれば良い。

$$\gamma_1 = \rho_1\gamma_0 = 0.56 \times 0.88 = 0.49$$

$$\gamma_2 = \rho_2\gamma_0 = 0.38 \times 0.88 = 0.33$$

$$\gamma_3 = \rho_3\gamma_0 = 0.24 \times 0.88 = 0.21$$

なお、 γ_1, γ_2 に関しては(2.15)式を用いて求めることもできる。

問題 2.7 AR(2) モデル $Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ の平均 μ 、分散 γ_0 および自己共分散 γ_1, γ_2 が下記のとおり与えられているとき、パラメータ ϕ_0, ϕ_1, ϕ_2 および ε_t の分散 σ^2 を求めよ。

$$\mu = 10.0, \quad \gamma_0 = 0.5, \quad \gamma_1 = 0.3 \quad \gamma_2 = 0.2$$

解答 2.7 自己相関の定義より $\rho_1 = \gamma_1 / \gamma_0 = 0.3 / 0.5 = 0.60$ 、 $\rho_2 = \gamma_2 / \gamma_0 = 0.2 / 0.5 = 0.40$ であるので、ユール=ウォーカー方程式は次のようになる。

$$\begin{cases} 0.60 = \phi_1 + 0.60\phi_2 \\ 0.40 = 0.60\phi_1 + \phi_2 \end{cases}$$

この方程式を ϕ_1, ϕ_2 について解けば、 $\phi_1 = 0.56, \phi_2 = 0.06$ が得られる。

また、(2.14)式より $\phi_0 = \mu(1 - \phi_1 - \phi_2)$ であるので、

$$\phi_0 = \mu(1 - \phi_1 - \phi_2) = 10.0 \times (1 - 0.56 - 0.06) = 3.75$$

となる。

分散 σ^2 については(2.4)式より $\sigma^2 = \gamma_0 - \gamma_1\phi_1 - \gamma_2\phi_2$ という関係式が成り立つため、

$$\sigma^2 = \gamma_0 - \gamma_1\phi_1 - \gamma_2\phi_2 = 0.5 - 0.3 \times 0.56 - 0.2 \times 0.06 = 0.32$$

として導かれる。

一般の p の場合は、定常性の条件（特定方程式の解の絶対値が1より大きい）を領域表現することはできないが、平均は $\mu = \phi_0 / (1 - \phi_1 - \phi_2 - \dots - \phi_p)$ であり、分散 γ_0 、自己共分散 γ_h および自己相関 ρ_h は、 $p = 2$ の場合と同様に行列を用いて表現することができる。

コレログラムの形状についても、 $p = 1, 2$ の場合と同様に指数関数的にゼロに収束するものとなる。

実際にデータを分析する際にはコレログラムなどの形状から定常性や次数に関する一定の判断をすることになるため、次数やパラメータによって形状がどのように変わるのかを知っておくことが重要である。

自己回帰モデルの実現値やコレログラムなどの例を章末に付けているので、そのグラフからイメージを掴んでいただきたい。（図 2.2 が $AR(1)$ の例、図 2.3 が $AR(2)$ の例である。）

2.3.2 移動平均モデル $MA(q)$

定義 2.8 確率過程 $\{Y_t\}$ が下式を満たすとき、 $\{Y_t\}$ は q 次の移動平均モデル (Moving Average Model) に従うといい、 $\{Y_t\} \sim MA(q)$ と表す。

$$Y_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.18)$$

ここで、パラメータ $\theta_0, \theta_1, \theta_2, \dots, \theta_q$ は定数であり、誤差項である ε_t は $Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ と独立に平均0、分散 σ^2 の t に依存しない同一分布に従う確率変数である。

$MA(q)$ は $AR(p)$ よりも具体的なイメージが掴みにくいモデルであるが、任意の定常な確率過程 $\{Y_t\}$ は決定的確率過程（過去の時系列データから将来の予測を完全に行うことができる過程）と $MA(\infty)$ の和として表現できることが分かっているので、表現力のあるモデルであると言える。

ラグ作用素 L を用いれば、(2.18)式は、

$$Y_t = \theta_0 + \theta(L)\varepsilon_t \quad (2.19)$$

$$\text{ここで、} \theta(L) = 1 - (\theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)$$

と表現することができる。

自己回帰モデル $AR(p)$ と異なり、移動平均モデル $MA(q)$ は必ず定常性を満たす。

誤差項 ε_t の分布が t によらないことから明らかであるが、実際に確認してみると、平均は (2.18)式の両辺の期待値をとれば、 $E(Y_t) = \theta_0$ (t に依存しない定数) となり、分散および自己共分散は、

$$\begin{aligned} \gamma_0 &= E((Y_t - \theta_0)^2) \\ &= E((\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})^2) \\ &= E(\varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2 + \dots + \theta_q^2 \varepsilon_{t-q}^2) \end{aligned}$$

$$\begin{aligned}
&= E(\varepsilon_t^2) + \theta_1^2 E(\varepsilon_{t-1}^2) + \cdots + \theta_q^2 E(\varepsilon_{t-q}^2) \\
&= \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)
\end{aligned} \tag{2.20}$$

$$\begin{aligned}
\gamma_1 &= E((Y_t - \theta_0)(Y_{t-1} - \theta_0)) \\
&= E((\varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q-1})) \\
&= E(-\theta_1\varepsilon_{t-1}^2 + \theta_1\theta_2\varepsilon_{t-2}^2 + \cdots + \theta_{q-1}\theta_q\varepsilon_{t-q}^2) \\
&= -\theta_1E(\varepsilon_{t-1}^2) + \theta_1\theta_2E(\varepsilon_{t-2}^2) + \cdots + \theta_{q-1}\theta_qE(\varepsilon_{t-q}^2) \\
&= \sigma^2(-\theta_1 + \theta_1\theta_2 + \cdots + \theta_{q-1}\theta_q)
\end{aligned} \tag{2.21}$$

$$\gamma_2 = E((Y_t - \theta_0)(Y_{t-2} - \theta_0)) = \cdots = \sigma^2(-\theta_2 + \theta_1\theta_3 + \cdots + \theta_{q-2}\theta_q) \tag{2.22}$$

⋮

$$\gamma_{q-1} = E((Y_t - \theta_0)(Y_{t-q+1} - \theta_0)) = \cdots = \sigma^2(-\theta_{q-1} + \theta_1\theta_q) \tag{2.23}$$

$$\gamma_q = E((Y_t - \theta_0)(Y_{t-q} - \theta_0)) = \cdots = \sigma^2(-\theta_q) \tag{2.24}$$

$$\gamma_{q+1} = E((Y_t - \theta_0)(Y_{t-q-1} - \theta_0)) = \cdots = 0 \tag{2.25}$$

$$\gamma_{q+2} = E((Y_t - \theta_0)(Y_{t-q-2} - \theta_0)) = \cdots = 0 \tag{2.26}$$

⋮

として時刻 t に依存しないことが確認できる。

この結果から自己相関を計算すると次のようになる。

$$\rho_1 = (-\theta_1 + \theta_1\theta_2 + \cdots + \theta_{q-2}\theta_{q-1} + \theta_{q-1}\theta_q)/(1 + \theta_1^2 + \cdots + \theta_q^2) \tag{2.27}$$

$$\rho_2 = (-\theta_2 + \theta_1\theta_3 + \cdots + \theta_{q-2}\theta_q)/(1 + \theta_1^2 + \cdots + \theta_q^2) \tag{2.28}$$

⋮

$$\rho_q = (-\theta_q)/(1 + \theta_1^2 + \cdots + \theta_q^2) \tag{2.29}$$

$$\rho_{q+1} = 0 \tag{2.30}$$

$$\rho_{q+2} = 0 \tag{2.31}$$

⋮

自己回帰モデルのコレログラムは指数関数的にゼロに収束する形状であったが、移動平均モデルでは $h = q + 1$ 以降ゼロとなるのが大きな特徴である。

移動平均モデルについてもコレログラムなどの例を章末に付けているので、次数およびパラメータによって形状がどのように変わるのか確認いただきたい。(図 2.4 が $MA(1)$ の例、図 2.5 が $MA(2)$ の例である。)

問題 2.8 $MA(1)$ モデル $Y_t = 2.0 + \varepsilon_t - 0.4\varepsilon_{t-1}$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.8$) について、 $\{Y_t\}$ の平均 μ 、分散 γ_0 、自己共分散 γ_1, γ_2 および自己相関 ρ_1, ρ_2 を求めよ。

解答 2.8 $Y_t = 2.0 + \varepsilon_t - 0.4\varepsilon_{t-1}$ の期待値を取れば平均 $\mu = 2.00$ が直ちに導かれる。

その他については(2.20)以降の式を用いて計算すれば良い。

$$\gamma_0 = \sigma^2(1 + \theta_1^2) = 0.8 \times (1 + 0.4^2) = 0.93$$

$$\gamma_1 = \sigma^2(-\theta_1) = 0.8 \times (-0.4) = -0.32$$

$$\gamma_2 = 0$$

$$\rho_1 = \gamma_1 / \gamma_0 = -0.32 / 0.93 = -0.34$$

$$\rho_2 = 0$$

問題 2.9 MA(2) モデル $Y_t = 2.0 + \varepsilon_t - 0.6\varepsilon_{t-1} - 0.2\varepsilon_{t-2}$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.6$) について、 $\{Y_t\}$ の平均 μ 、分散 γ_0 、自己共分散 $\gamma_1, \gamma_2, \gamma_3$ および自己相関 ρ_1, ρ_2, ρ_3 を求めよ。

解答 2.9 $Y_t = 2.0 + \varepsilon_t - 0.6\varepsilon_{t-1} - 0.2\varepsilon_{t-2}$ の期待値を取れば平均 $\mu = 2.00$ が直ちに導かれる。その他については(2.20)以降の式を用いて計算すれば良い。

$$\gamma_0 = \sigma^2(1 + \theta_1^2 + \theta_2^2) = 0.6 \times (1 + 0.6^2 + 0.2^2) = 0.84$$

$$\gamma_1 = \sigma^2(-\theta_1 + \theta_1\theta_2) = 0.6 \times (-0.6 + 0.6 \times 0.2) = -0.29$$

$$\gamma_2 = \sigma^2(-\theta_2) = 0.6 \times (-0.2) = -0.12$$

$$\gamma_3 = 0$$

$$\rho_1 = \gamma_1 / \gamma_0 = -0.29 / 0.84 = -0.34$$

$$\rho_2 = \gamma_2 / \gamma_0 = -0.12 / 0.84 = -0.14$$

$$\rho_3 = 0$$

ここで改めて自己回帰モデル $AR(p)$ について考えてみると、

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

に対して、

$$Y_{t-1} = \phi_0 + \phi_1 Y_{t-2} + \phi_2 Y_{t-3} + \cdots + \phi_p Y_{t-p-1} + \varepsilon_{t-1}$$

$$Y_{t-2} = \phi_0 + \phi_1 Y_{t-3} + \phi_2 Y_{t-4} + \cdots + \phi_p Y_{t-p-2} + \varepsilon_{t-2}$$

⋮

を順次代入することにより、

$$Y_t = (\text{定数}) + \varepsilon_t + \phi_1 \varepsilon_{t-1} + (\phi_1^2 + \phi_2) \varepsilon_{t-2} + \cdots$$

という表現が得られる。

$AR(p)$ が定常である場合 (特性方程式 $\phi(x) = 0$ の解の絶対値が1より大きい場合) はこの無限級数表現が可能であり、 $AR(p)$ は $MA(\infty)$ によって表現できることが分かっている。

$AR(p)$ の $MA(\infty)$ 表現を $Y_t = \xi_0 + \varepsilon_t + \xi_1 \varepsilon_{t-1} + \cdots$ とし、 $Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$ に対して Y_{t-1}, \dots, Y_{t-p} の $MA(\infty)$ 表現をそれぞれ代入することによって、

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

$$= \phi_0 + \varepsilon_t + \phi_1 (\xi_0 + \varepsilon_{t-1} + \xi_1 \varepsilon_{t-2} + \xi_2 \varepsilon_{t-3} + \cdots)$$

$$+ \phi_2 (\xi_0 + \varepsilon_{t-2} + \xi_1 \varepsilon_{t-3} + \xi_2 \varepsilon_{t-4} + \cdots)$$

⋮

$$+ \phi_p (\xi_0 + \varepsilon_{t-p} + \xi_1 \varepsilon_{t-p-1} + \xi_2 \varepsilon_{t-p-2} + \cdots)$$

が得られる。この式と Y_t の $MA(\infty)$ 表現の定数項および $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ の係数を比較することによって、 $AR(p)$ の $MA(\infty)$ 表現での係数 ξ_0, ξ_1, \dots を次のように求められることが分かる。

(表現を分かりやすくするために、便宜的に $\phi_{p+1} = \phi_{p+2} = \phi_{p+3} = \dots = 0$ としておく。)

$$\begin{cases} \xi_0 = \phi_0 / (1 - \phi_1 - \dots - \phi_p) \\ \xi_1 = \phi_1 \\ \xi_2 = \phi_1 \xi_1 + \phi_2 \\ \xi_3 = \phi_1 \xi_2 + \phi_2 \xi_1 + \phi_3 \\ \vdots \end{cases} \quad (2.32)$$

この表現を用いれば、時刻 t までの実績を用いて時刻 t 以降の予測をする場合の予測量の分散を容易に求めることができる。

問題 2.10 $AR(2)$ モデル $Y_t = 2.0 + 0.5Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.6$) を、 $MA(\infty)$ 表現 ($Y_t = \xi_0 + \varepsilon_t + \xi_1 \varepsilon_{t-1} + \dots$) した時の ξ_0, ξ_1, ξ_2 を求めよ。

さらに、この表現を用いて、時刻 t までの時系列データが与えられている時の Y_{t+3} の分散 (つまり、条件付分散 $V(Y_{t+3} | Y_t, Y_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots)$) を求めよ。

解答 2.10 (2.32)式を用いれば、

$$\xi_0 = \phi_0 / (1 - \phi_1 - \phi_2) = 2.0 / (1 - 0.5 - 0.1) = 5.00$$

$$\xi_1 = \phi_1 = 0.50$$

$$\xi_2 = \phi_1 \xi_1 + \phi_2 = 0.5 \times 0.50 + 0.1 = 0.35$$

となる。

条件付分散はこの ξ_1, ξ_2 を用いて、

$$\begin{aligned} V(Y_{t+3} | Y_t, Y_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots) &= \sigma^2 + \xi_1^2 \sigma^2 + \xi_2^2 \sigma^2 \\ &= 0.6 + 0.50^2 \times 0.6 + 0.35^2 \times 0.6 \\ &= 0.82 \end{aligned}$$

となる。

逆に、 $MA(q)$ を $AR(\infty)$ で表現できるか否か (これを反転可能性という) については、次の条件の下で可能であることが分かっている。

定常な $AR(p)$ が $MA(\infty)$ によって表現できることを既知とすれば、 $AR(p)$ と $MA(q)$ が Y_t と ε_t に関して対称な表現式となっていることより、この命題は直ちに導かれる。

命題 2.2 $MA(q)$ の特性方程式 $\theta(x) = 1 - (\theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q) = 0$ の解の絶対値がすべて1より大きい時、 $MA(q)$ は反転可能である。

次に識別可能性について説明しておく。

自己回帰モデルでは平均 μ および自己共分散 γ_h (自己相関 ρ_h) が与えられれば、既に述べたように $\phi_0, \phi_1, \dots, \phi_p$ および σ^2 が定まり、(2.1)式を特定することができる。しかし、移動平均

モデルではこれが成り立たない。

具体的な例として次の2つの移動平均モデルMA(1)を考える。

$$Y_t = 1 + \varepsilon_t - 0.5\varepsilon_{t-1}, V(\varepsilon_t) = \sigma^2$$

$$Y'_t = 1 + \varepsilon'_t - 2.0\varepsilon'_{t-1}, V(\varepsilon'_t) = 0.25\sigma^2$$

平均はともに1であり、分散および自己共分散も上で導いた式を用いて、

$$\begin{cases} \gamma_0 = \sigma^2(1+0.5^2) = 1.25\sigma^2 \\ \gamma_1 = \sigma^2(-0.5) = -0.5\sigma^2 \\ \gamma_2 = 0 \\ \vdots \end{cases} \quad \begin{cases} \gamma'_0 = 0.25\sigma^2(1+2.0^2) = 1.25\sigma^2 \\ \gamma'_1 = 0.25\sigma^2(-2.0) = -0.5\sigma^2 \\ \gamma'_2 = 0 \\ \vdots \end{cases}$$

として完全に一致する。

つまり、移動平均モデルに関しては時系列データから平均、分散および自己共分散を推定できても、モデルを1つに特定できないことが分かる。

平均、分散および自己共分散が与えられたときにモデルが一意に定まることを識別可能であるという。上記の例から移動平均モデルは一般に識別可能ではないことが分かるが、次の条件を置くことによって識別可能となることが分かっている。

命題 2.3 MA(q) の特性方程式 $\theta(x) = 1 - (\theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q) = 0$ の解の絶対値がすべて1以上である時、MA(q) は識別可能である。

つまり、反転可能性の条件を若干緩め、解の絶対値が1である場合も含めたものが識別可能性の条件であり、反転可能であれば識別可能、ということになる。

問題 2.11 識別可能であるMA(1)モデル $Y_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ の平均 μ 、分散 γ_0 および自己共分散 γ_1 が下記のとおり与えられている時、パラメータ θ_0, θ_1 および ε_t の分散 σ^2 を求めよ。

$$\mu = 2.0, \quad \gamma_0 = 5.2, \quad \gamma_1 = 1.0$$

解答 2.11 $\mu = \theta_0$ であるので、 $\theta_0 = 2.0$ である。

θ_1 および σ^2 は、(2.20)式および(2.21)式から得られる方程式、

$$\begin{cases} \gamma_0 = 5.2 = \sigma^2(1 + \theta_1^2) \\ \gamma_1 = 1.0 = \sigma^2(-\theta_1) \end{cases}$$

を解けば良いが、この方程式は θ_1 の二次方程式であるため、これを解くと $(\theta_1, \sigma^2) = (-0.20, 5.00), (-5.00, 0.20)$ として2つの解が得られる。ここで、識別可能性の仮定より、MA(1)の特性方程式 $\theta(x) = 1 - \theta_1 x = 0$ の解の絶対値が1以上という条件を用いれば、 $|1/\theta_1| \geq 1$ 、つまり、 $(\theta_1, \sigma^2) = (-0.20, 5.00)$ であることが分

かる。

時系列データを移動平均モデル $MA(q)$ にあてはめる場合は、通常、反転可能性の条件が満たされているものと仮定し、識別可能性を前提として議論を進める。

2.3.3 自己回帰移動平均モデル $ARMA(p, q)$

定義 2.9 確率過程 $\{Y_t\}$ が自己回帰モデルと移動平均モデルを組み合わせた下式を満たすとき、 $\{Y_t\}$ は次数 p, q の自己回帰移動平均モデル (Autoregressive Moving Average Model) に従うといい、 $\{Y_t\} \sim ARMA(p, q)$ と表す。

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad (2.33)$$

ここで、パラメータ $c, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_p$ は定数であり、誤差項である ε_t は $Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ と独立に平均 0、分散 σ^2 である t に依存しない同一分布に従う確率変数である。

この定義より、自己回帰移動平均モデルは自己回帰モデルおよび移動平均モデルをその特殊な場合として含むモデルであり、 $AR(p) = ARMA(p, 0)$ 、 $MA(q) = ARMA(0, q)$ であることが分かる。

ラグ作用素 L を用いれば、(2.33)式は、

$$\phi(L)Y_t = c + \theta(L)\varepsilon_t \quad (2.34)$$

$$\text{ここで、 } \phi(L) = 1 - (\phi_1 L + \cdots + \phi_p L^p), \quad \theta(L) = 1 - (\theta_1 L + \cdots + \theta_q L^q)$$

と表現することができる。

自己回帰移動平均モデル $ARMA(p, q)$ の定常性、反転可能性および識別可能性の条件は次のようになる。

命題 2.4 $\phi(x) = 0$ の解の絶対値がすべて 1 より大きい時、 $ARMA(p, q)$ は定常性を持つ。

命題 2.5 $\theta(x) = 0$ の解の絶対値がすべて 1 より大きい時、 $ARMA(p, q)$ は反転可能である。

命題 2.6 $\theta(x) = 0$ の解の絶対値がすべて 1 以上であり、かつ $\phi(x) = 0$ と共通解を持たない時、 $ARMA(p, q)$ は識別可能である。

定常である $ARMA(p, q)$ は $MA(\infty)$ で表現することができ、反転可能である $ARMA(p, q)$ は $AR(\infty)$ で表現することができる。

$ARMA(p, q)$ が $MA(\infty)$ 表現可能であるとき、 $MA(\infty)$ による表現を $Y_t = \xi_0 + \varepsilon_t + \xi_1 \varepsilon_{t-1} + \cdots$ とすれば、 $\xi_0, \xi_1, \xi_2, \dots$ は $ARMA(p, q)$ の係数 $c, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_p$ を用いて次のように表される。この表現も $AR(p)$ から $MA(\infty)$ 表現を導いた方法と同様の方法で導かれる。(表現を分か

りやすくするために、便宜的に $\phi_{p+1} = \phi_{p+2} = \dots = 0$, $\theta_{q+1} = \theta_{q+2} = \dots = 0$ としておく。))

$$\begin{cases} \xi_0 = c/(1 - \phi_1 - \dots - \phi_p) \\ \xi_1 = \phi_1 - \theta_1 \\ \xi_2 = \phi_1 \xi_1 + \phi_2 - \theta_2 \\ \xi_3 = \phi_1 \xi_2 + \phi_2 \xi_1 + \phi_3 - \theta_3 \\ \vdots \end{cases} \quad (2.35)$$

問題 2.12 定常な ARMA(1,1) モデル、

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (2.36)$$

のパラメータ ϕ_1, θ_1 および $\sigma^2 (= V(\varepsilon_t))$ を用いて自己相関 $\rho_h (h = 1, 2, \dots)$ を表せ。

解答 2.12 まず、このモデルの平均を求めておく。(2.36)式の両辺の期待値を取れば $\mu = \phi_1 \mu$ であり、定常性の仮定より $\phi(x) = 1 - \phi_1 x = 0$ の解は 1 より大、つまり $\phi_1 < 1$ であるので、 $\mu = 0$ が分かる。

次に、(2.36)式の $\phi_1 Y_{t-1}$ を左辺に移項した $Y_t - \phi_1 Y_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ について、両辺の分散 (期待値が 0 であるので、二乗の期待値に等しい) を計算すると、

$$\begin{aligned} E((Y_t - \phi_1 Y_{t-1})^2) &= E((\varepsilon_t - \theta_1 \varepsilon_{t-1})^2) \\ \Leftrightarrow E(Y_t^2) - 2\phi_1 E(Y_t Y_{t-1}) + \phi_1^2 E(Y_{t-1}^2) &= E(\varepsilon_t^2) - 2\theta_1 E(\varepsilon_t \varepsilon_{t-1}) + \theta_1^2 E(\varepsilon_{t-1}^2) \\ \Leftrightarrow \gamma_0 - 2\phi_1 \gamma_1 + \phi_1^2 \gamma_0 &= \sigma^2 + \theta_1^2 \sigma^2 \\ \Leftrightarrow \frac{\sigma^2}{\gamma_0} &= \frac{1 - 2\phi_1 \rho_1 + \phi_1^2}{1 + \theta_1^2} \end{aligned} \quad (2.37)$$

となる。

また、(2.36)式の t を $t+1$ とした $Y_{t+1} = \phi_1 Y_t + \varepsilon_{t+1} - \theta_1 \varepsilon_t$ の両辺に Y_t を乗じて期待値を取ると、

$$\begin{aligned} E(Y_t Y_{t+1}) &= E(\phi_1 Y_t^2 + Y_t \varepsilon_{t+1} - \theta_1 Y_t \varepsilon_t) \\ \Leftrightarrow \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma^2 \\ \Leftrightarrow \rho_1 &= \phi_1 - \frac{\theta_1 \sigma^2}{\gamma_0} \end{aligned} \quad (2.38)$$

となるので、この右辺に(2.37)式を代入すれば、

$$\rho_1 = \phi_1 - \frac{\theta_1 (1 - 2\phi_1 \rho_1 + \phi_1^2)}{1 + \theta_1^2} \quad (2.39)$$

となり、この式を ρ_1 について解けば、

$$\rho_1 = \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{1 - 2\phi_1 \theta_1 + \theta_1^2} \quad (2.40)$$

が導かれる。

$\rho_h (h = 2, 3, \dots)$ については、(2.36)式の t を $t+h$ とした式の両辺に Y_t を乗じて期待値を取って得られる $\gamma_h = \phi_1 \gamma_{h-1}$ 、つまり $\rho_h = \phi_1 \rho_{h-1}$ から導くことができる。

以上をまとめると、

$$\rho_h = \phi_1^{h-1} \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{1 - 2\phi_1 \theta_1 + \theta_1^2} \quad (h = 1, 2, 3, \dots) \quad (2.41)$$

となる。

問題 2.13 平均が 0 である定常な ARMA(1,1) モデル $Y_t = 0.8Y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t-1} (V(\varepsilon_t) = 0.3)$ の ρ_1, ρ_2 および γ_0 を求めよ。

解答 2.13 自己相関 ρ_1, ρ_2 は(2.41)式を用いて計算すれば良い。計算結果は、

$$\rho_1 = 0.70, \quad \rho_2 = 0.56$$

となる。

Y_t の分散 γ_0 は(2.37)式より、

$$\frac{\sigma^2}{\gamma_0} = \frac{1 - 2\phi_1\rho_1 + \phi_1^2}{1 + \theta_1^2} \Rightarrow \frac{0.3}{\gamma_0} = \frac{1 - 2 \times 0.8 \times 0.70 + 0.8^2}{1 + 0.2^2}$$

であるので、上式を解けば $\gamma_0 = 0.60$ として導かれる。

2.3.4 自己回帰和分移動平均モデル ARIMA(p, d, q)

定義 2.10 確率過程 $\{Y_t\}$ が次の条件を満たすとき、 $\{Y_t\}$ は次数 p, d, q の自己回帰和分移動平均モデル (Autoregressive Integrated Moving Average Model) に従うといい、 $\{Y_t\} \sim ARIMA(p, d, q)$ と表す。

$\{Y_t\}$ に対して d 回の階差変換を行った後の確率過程を $\{Y_t'\}$ として、

- $\{Y_t'\}$ は定常
 - $\{Y_t'\} \sim ARMA(p, q)$
- (2.42)

ここで、階差変換とは、

$$1 \text{ 回の階差変換 } \Delta Y_t = Y_t - Y_{t-1}$$

$$2 \text{ 回の階差変換 } \Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$$

$$\vdots \quad \quad \quad \vdots$$

として定義される変換 $\Delta, \Delta^2, \Delta^3, \dots$ をいう。(つまり、(2.42)では $Y_t' = \Delta^d Y_t$ である。)

ラグ作用素 L を用いれば、

$$\Delta = 1 - L$$

$$\Delta^2 = (1 - L)^2 (= 1 - 2L + L^2)$$

$$\Delta^3 = (1 - L)^3 (= 1 - 3L + 3L^2 - L^3)$$

⋮

となるので、(2.42)は、

$$\begin{aligned} & \cdot \{(1-L)^d Y_t\} \text{は定常} \\ & \cdot \{(1-L)^d Y_t\} \sim ARMA(p, q) \end{aligned} \tag{2.43}$$

と表現することができる。

応用上現れる時系列の多くは定常性を満たすとは考えにくい、その系列の階差をとることによって定常と見なせる場合がある。以降で説明する時系列データを用いたモデルの同定、推定、診断および予測を行う際に前提として置くモデルが $ARIMA(p, d, q)$ である。

実際の応用に際しては、 $ARIMA(p, d, q)$ を前提として、

- step1 次数 p, d, q の推定 (モデルの同定)
- step2 パラメータ $c, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_p$ の推定 (モデルの推定)
- step3 推定された $ARIMA(p, d, q)$ の検証 (モデルの診断)
- step4 採用された $ARIMA(p, d, q)$ を用いた予測

という手順で分析を行う。

2.4 時系列モデルの同定・推定・診断・予測

ここでは与えられた時系列が自己回帰和分移動平均モデル $ARIMA(p, d, q)$ に従うという仮定の下に、次数の推定、パラメータの推定、推定されたモデルの検証および採用されたモデルを用いた予測について概要を説明する。

2.4.1 モデルの同定

最初に階差変換の回数である d を決定する。 d の決定に際して用いる情報はプロットされた時系列データと標本コレログラムである。

まず、プロットされた時系列データから増加 (減少) トレンドがないかを確認する。定常な時系列では平均が一定となるので、増加 (減少) トレンドを含む場合は定常性を満たしていないことになる。

次に標本コレログラム $\hat{\rho}_h$ の形状を確認する。定常な時系列ではコレログラム ρ_h が急速にゼロに収束するため、標本コレログラム $\hat{\rho}_h$ の形状から定常性を満たしているか否かおよその判断をすることができる。

定常でないとは判断される場合は、階差をとったデータに対してデータのプロットおよび標本自己相関の算出を行い、階差データの定常性を判断する。この手続を繰り返して定常性を満たすと評価できる階差変換の回数 d を決定する。(プロットされたデータの変動幅が増加トレンドにあると考えられる場合は対数変換を行うこともある。)

次に適当な階差変換によって定常な $ARMA(p, q)$ に変換されているものと仮定し、変換後の時系列データを改めて $\{y_t\}_{t=1}^n$ として残された次数である p, q をどのように推定するかについ

て考える。この場合も、標本コレログラム $\hat{\rho}_h$ が有力な情報となる。

$AR(p)$, $MA(q)$ および $ARMA(p, q)$ のコレログラムの特徴は次のようにまとめることができる。

【コレログラムの特徴】

$AR(p)$	h が大きくなると指数関数的に 0 に収束
$MA(q)$	$h = q + 1$ 以降は $\rho_h = 0$
$ARMA(p, q)$	h が大きくなると指数関数的に 0 に収束

さらに、 $\rho_h = 0$ の時、 $\hat{\rho}_h$ は漸近的に正規分布に従うことが分かっているので、これらの知識を踏まえて標本コレログラムの形状を確認し、 $\rho_h = 0$ に対する仮説検定を行うなどによって、候補としての p, q を選ぶことになる。

また、候補としての p, q を選ぶ際に、標本偏自己相関を利用することもできる。偏自己相関は次のように定義される。

定義 2.11 確率過程 $\{Y_t\}$ の自己相関を $\rho_1, \rho_2, \rho_3, \dots$ として、次の連立方程式を満たす ϕ_{hh} を Y_t と Y_{t-h} の偏自己相関という。

$$\begin{cases} \rho_1 = \phi_{h1} + \phi_{h2}\rho_1 + \dots + \phi_{hh}\rho_{h-1} \\ \rho_2 = \phi_{h1}\rho_1 + \phi_{h2} + \dots + \phi_{hh}\rho_{h-2} \\ \vdots \\ \rho_h = \phi_{h1}\rho_{h-1} + \phi_{h2}\rho_{h-2} + \dots + \phi_{hh} \end{cases} \quad (2.44)$$

この連立方程式は自己回帰モデル $AR(p)$ で説明したユール=ウォーカー方程式の p を h に置き換え、 $\phi_i (i = 1, 2, \dots, p)$ を $\phi_{hi} (i = 1, 2, \dots, h)$ としたものである。

偏自己相関は自己相関 (コレログラム) を補完する特徴を持っており、次のようにまとめることができる。

【偏自己相関の特徴】

$AR(p)$	$h = p + 1$ 以降は $\phi_{hh} = 0$
$MA(q)$	h が大きくなると指数関数的に 0 に収束
$ARMA(p, q)$	h が大きくなると指数関数的に 0 に収束

自己回帰モデル $AR(p)$ の $h = p + 1$ 以降の偏自己相関 ϕ_{hh} がゼロとなることは、ユール=ウォーカー方程式から導かれる。

さらに、標本自己相関 $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \dots$ を用いて(2.44)式から算出された推定量 $\hat{\phi}_{hh}$ (標本偏自己相関) は、 $\phi_{hh} = 0$ の時、漸近的に正規分布に従うことが分かっているので、これらの知識を踏まえて標本偏自己相関の形状を確認し、 $\phi_{hh} = 0$ に対する仮説検定を行うなどによって、標本コレログラム $\hat{\rho}_h$ を補完する情報を読み取ることができる。

問題 2.14 $AR(2)$ モデル $Y_t = 2.0 + 0.5Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.6$) (問題 2.6 と同じモデル) について、偏自己相関 $\phi_{11}, \phi_{22}, \phi_{33}$ を求めよ。

解答 2.14 問題 2.6 より $\rho_1 = 0.56, \rho_2 = 0.38, \rho_3 = 0.24$ であるので、求める偏自己相関は次の方程式の解である。

$$\rho_1 = \phi_{11}, \quad \begin{cases} \rho_1 = \phi_{21} + \phi_{22}\rho_1 \\ \rho_2 = \phi_{21}\rho_1 + \phi_{22} \end{cases} \quad \begin{cases} \rho_1 = \phi_{31} + \phi_{32}\rho_1 + \phi_{33}\rho_2 \\ \rho_2 = \phi_{31}\rho_1 + \phi_{32} + \phi_{33}\rho_1 \\ \rho_3 = \phi_{31}\rho_2 + \phi_{32}\rho_1 + \phi_{33} \end{cases}$$

$$\Leftrightarrow 0.56 = \phi_{11}, \quad \begin{cases} 0.56 = \phi_{21} + 0.56\phi_{22} \\ 0.38 = 0.56\phi_{21} + \phi_{22} \end{cases} \quad \begin{cases} 0.56 = \phi_{31} + 0.56\phi_{32} + 0.38\phi_{33} \\ 0.38 = 0.56\phi_{31} + \phi_{32} + 0.56\phi_{33} \\ 0.24 = 0.38\phi_{31} + 0.56\phi_{32} + \phi_{33} \end{cases}$$

よって、これを解けば $\phi_{11} = 0.56, \phi_{22} = 0.10, \phi_{33} = 0.00$ となる。

問題 2.15 $MA(2)$ モデル $Y_t = 2.0 + \varepsilon_t - 0.6\varepsilon_{t-1} - 0.2\varepsilon_{t-2}$ ($E(\varepsilon_t) = 0, V(\varepsilon_t) = 0.6$) (問題 2.9 と同じモデル) について、偏自己相関 $\phi_{11}, \phi_{22}, \phi_{33}$ を求めよ。

解答 2.15 問題 2.9 より $\rho_1 = -0.34, \rho_2 = -0.14, \rho_3 = 0$ であるので、求める偏自己相関は次の方程式の解である。

$$-0.34 = \phi_{11}, \quad \begin{cases} -0.34 = \phi_{21} - 0.34\phi_{22} \\ -0.14 = -0.34\phi_{21} + \phi_{22} \end{cases} \quad \begin{cases} -0.34 = \phi_{31} - 0.34\phi_{32} - 0.14\phi_{33} \\ -0.14 = -0.34\phi_{31} + \phi_{32} - 0.34\phi_{33} \\ 0 = -0.14\phi_{31} - 0.34\phi_{32} + \phi_{33} \end{cases}$$

よって、これを解けば $\phi_{11} = -0.34, \phi_{22} = -0.30, \phi_{33} = -0.20$ となる。

2.4.2 モデルの推定

次に次数 p, q が定まっている $AR(p), MA(q), ARMA(p, q)$ に対して、パラメータを推定する方法について説明する。

なお、以下で説明する方法以外に最尤法による方法があるが、こちらは数学的な議論が込み入ったものとなるため、ここでは省略する。興味のある方は参考文献 3 を参照されたい。

自己回帰モデル $AR(p)$ の推定

【手法 1】最小二乗法

AR(p)のパラメータの推定手法として最も一般的である。

時系列データ $\{y_t\}_{t=1}^n$ を用いて残差の二乗和である下式を最小にする係数 $\phi_0, \phi_1, \dots, \phi_p$ を推定値 $\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p$ とする。

$$\sum_{t=p+1}^n \{y_t - (\phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p})\}^2 \quad (2.45)$$

最小二乗法による係数の推定値は下式によって与えられる。(1.5 重回帰 を参照。)

$$\begin{pmatrix} \hat{\phi}_0 \\ \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = (X^T X)^{-1} X^T \begin{pmatrix} y_{p+1} \\ y_{p+2} \\ \vdots \\ y_n \end{pmatrix} \quad (2.46)$$

$$\text{ここで、} X = \begin{pmatrix} 1 & y_p & \dots & y_1 \\ 1 & y_{p+1} & \dots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_{n-1} & \dots & y_{n-p} \end{pmatrix}$$

なお、誤差項 ε_t の分散 σ^2 の推定値 $\hat{\sigma}^2$ としては、

$$\hat{\sigma}^2 = \frac{1}{n-p-1} \sum_{t=p+1}^n \{y_t - (\hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} + \dots + \hat{\phi}_p y_{t-p})\}^2 \quad (2.47)$$

を用いる。

【手法2】標本自己相関を用いる方法

標本自己相関 $\hat{\rho}_h$ とユール=ウォーカー方程式を用いて、パラメータを推定することもできる。

この場合は、

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \dots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix} \quad (2.48)$$

となる。($\hat{\phi}_0$ は $\hat{\mu} = \bar{y} = \hat{\phi}_0 / (1 - \hat{\phi}_1 - \hat{\phi}_2 - \dots - \hat{\phi}_p)$ から求める。)

問題 2.16 AR(1)に従っていると考えられる時系列データ $\{y_t\}_{t=1}^{10}$ が下表のとおり与えられている時、最小二乗法によってAR(1)のパラメータ ϕ_0, ϕ_1 を推定せよ。

t	1	2	3	4	5	6	7	8	9	10
y_t	3.00	3.54	3.78	4.02	4.06	4.15	4.05	4.14	4.15	3.98

解答 2.16 (2.46)式における各行列は次のようになる。

$$X = \begin{pmatrix} 1 & 3.00 \\ 1 & 3.54 \\ 1 & 3.78 \\ 1 & 4.02 \\ 1 & 4.06 \\ 1 & 4.15 \\ 1 & 4.05 \\ 1 & 4.14 \\ 1 & 4.15 \end{pmatrix}, \quad Y = \begin{pmatrix} 3.54 \\ 3.78 \\ 4.02 \\ 4.06 \\ 4.15 \\ 4.05 \\ 4.14 \\ 4.15 \\ 3.98 \end{pmatrix}$$

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3.00 & 3.54 & 3.78 & 4.02 & 4.06 & 4.15 & 4.05 & 4.14 & 4.15 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 9.00 & 34.89 \\ 34.89 & 136.45 \end{pmatrix}, \quad (X^T X)^{-1} = \begin{pmatrix} 12.70 & -3.25 \\ -3.25 & 0.84 \end{pmatrix}$$

これらを用いて(2.46)式に従って計算すれば、 $\begin{pmatrix} \hat{\phi}_0 \\ \hat{\phi}_1 \end{pmatrix} = \begin{pmatrix} 2.09 \\ 0.49 \end{pmatrix}$ が導かれる。

問題 2.17 $AR(1)$ に従っていると考えられる問題 2.16 のデータを用いて標本自己相関からパラメータ ϕ_0, ϕ_1 を推定せよ。

解答 2.17 定義 2.5 に従って標本自己相関 $\hat{\rho}_1$ を計算すると $\hat{\rho}_1 = 0.48$ となる。 $AR(1)$ では $\rho_1 = \phi_1$ であるので、 $\hat{\phi}_1 = 0.48$ となる。

また、 $\mu = \phi_0 / (1 - \phi_1)$ であるので、 $\hat{\mu} = \bar{y} = 3.89$ を用いて、 $\hat{\phi}_0 = \hat{\mu}(1 - \hat{\phi}_1) = 2.03$ が導かれる。

移動平均モデル $MA(q)$ の推定

【手法 1】最小二乗法

$MA(q)$ の場合は $AR(p)$ と異なりそのまま最小二乗法を適用することはできないため、時系列データ $\{y_t\}_{t=1}^n$ に加えて、時刻 0 以前の誤差項を

$$e_0 = e_{-1} = e_{-2} = \cdots = e_{1-q} = 0$$

と仮定し、残差を、

$$e_1 = y_1 - \theta_0 + \theta_1 e_0 + \theta_2 e_{-1} + \cdots + \theta_q e_{1-q}$$

$$e_2 = y_2 - \theta_0 + \theta_1 e_1 + \theta_2 e_0 + \cdots + \theta_q e_{2-q}$$

⋮

$$e_n = y_n - \theta_0 + \theta_1 e_{n-1} + \theta_2 e_{n-2} + \cdots + \theta_q e_{n-q}$$

として $\sum_{t=1}^n e_t^2$ を最小化する $\theta_0, \theta_1, \dots, \theta_q$ を推定量 $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_q$ と考える。

ただし、この場合は $e_t (t = 1, \dots, n)$ が $\theta_0, \theta_1, \dots, \theta_q$ の非線形関数となるため、明示的に解を求めることはできず、ニュートン法などの数値計算手法によって計算を行う必要がある。

【手法 2】標本自己相関を用いる方法

$MA(q)$ の定常性を確認するために導いた自己共分散とパラメータ $\theta_0, \theta_1, \dots, \theta_q$ の関係式 ((2.27) 以降の式) を用いて、標本自己相関 $\hat{\rho}_h$ からパラメータを推定することもできる。

問題 2.18 識別可能な $MA(1)$ に従っていると考えられる時系列データ $\{y_t\}_{t=1}^{10}$ が下表のとおり与えられている時、標本自己相関からパラメータ θ_0, θ_1 を推定せよ。

t	1	2	3	4	5	6	7	8	9	10
y_t	1.96	1.96	2.02	1.97	2.00	1.98	2.05	1.94	2.07	2.02

解答 2.18 定義 2.5 に従って標本自己相関 $\hat{\rho}_1$ を計算すると $\hat{\rho}_1 = -0.41$ となる。 $MA(1)$ では $\rho_1 = -\theta_1 / (1 + \theta_1^2)$ であるので、 $\hat{\rho}_1 = -0.41$ を代入すると $\hat{\theta}_1 = 0.52, 1.92$ となる。ここで、識別可能性の仮定より、 $MA(1)$ の特性方程式 $\theta(x) = 1 - \theta_1 x = 0$ の解の絶対値が 1 以上であるので、 $\hat{\theta}_1 = 0.52$ であることが分かる。
また、 $\mu = \theta_0$ であるので、 $\hat{\mu} = \bar{y} = 2.00$ を用いて、 $\hat{\theta}_0 = 2.00$ が導かれる。

自己回帰移動平均モデル $ARMA(p, q)$ の推定

$MA(q)$ と同様に、 $ARMA(p, q)$ の場合もそのまま最小二乗法を適用することは困難であるため、時系列データ $\{y_t\}_{t=1}^n$ に加えて、時刻 p 以前の誤差項を

$$e_p = e_{p-1} = \dots = e_{p-q+2} = e_{p-q+1} = 0$$

を仮定し、残差を

$$\begin{aligned} e_{p+1} &= y_{p+1} - (c + \phi_1 y_p + \dots + \phi_p y_1) + (\theta_1 e_p + \theta_2 e_{p-1} + \dots + \theta_q e_{p-q+1}) \\ e_{p+2} &= y_{p+2} - (c + \phi_1 y_{p+1} + \dots + \phi_p y_2) + (\theta_1 e_{p+1} + \theta_2 e_p + \dots + \theta_q e_{p-q+2}) \\ &\vdots \\ e_n &= y_n - (c + \phi_1 y_{n-1} + \dots + \phi_p y_{n-p}) + (\theta_1 e_{n-1} + \theta_2 e_{n-2} + \dots + \theta_q e_{n-q}) \end{aligned}$$

として $\sum_{t=p+1}^n e_t^2$ を最小化する $c, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_p$ を推定量 $\hat{c}, \hat{\theta}_1, \dots, \hat{\theta}_p, \hat{\phi}_1, \dots, \hat{\phi}_p$ と考える。

ただし、この方法も $e_t (t = p+1, \dots, n)$ が $c, \theta_1, \dots, \theta_p, \phi_1, \dots, \phi_p$ の非線形関数となるため、明示的に解を求めることはできず、ニュートン法などの数値計算手法によって計算を行う必要がある。

以上で述べたように $AR(p)$ 以外についてはパラメータの推定は相当に複雑なものとなる。そのため、一般的には市販されているパッケージプログラムを用いて推定を行うことになる。

2.4.3 モデルの診断

推定した時系列モデルの妥当性を検証するために、最初にすることは残差のプロットである。これによって、残差に特定のパターンがないか視覚的に確認することができる。

一部繰り返しになるが、各時系列モデルの残差を示しておく。

自己回帰モデル $AR(p)$ の残差

$$\begin{aligned} e_{p+1} &= y_{p+1} - (\hat{\phi}_0 + \hat{\phi}_1 y_p + \hat{\phi}_2 y_{p-1} + \cdots + \hat{\phi}_p y_1) \\ e_{p+2} &= y_{p+2} - (\hat{\phi}_0 + \hat{\phi}_1 y_{p+1} + \hat{\phi}_2 y_p + \cdots + \hat{\phi}_p y_2) \\ &\vdots \\ e_n &= y_n - (\hat{\phi}_0 + \hat{\phi}_1 y_{n-1} + \hat{\phi}_2 y_{n-2} + \cdots + \hat{\phi}_p y_{n-p}) \end{aligned}$$

移動平均モデル $MA(q)$ の残差

$$\begin{aligned} e_1 &= y_1 - \hat{\theta}_0 + \hat{\theta}_1 e_0 + \hat{\theta}_2 e_{-1} + \cdots + \hat{\theta}_q e_{1-q} \\ e_2 &= y_2 - \hat{\theta}_0 + \hat{\theta}_1 e_1 + \hat{\theta}_2 e_0 + \cdots + \hat{\theta}_q e_{2-q} \\ &\vdots \\ e_n &= y_n - \hat{\theta}_0 + \hat{\theta}_1 e_{n-1} + \hat{\theta}_2 e_{n-2} + \cdots + \hat{\theta}_q e_{n-q} \\ e_0 &= e_{-1} = e_{-2} = \cdots = e_{1-q} = 0 \end{aligned}$$

自己回帰移動平均モデル $ARMA(p, q)$ の残差

$$\begin{aligned} e_{p+1} &= y_{p+1} - (\hat{c} + \hat{\phi}_1 y_p + \cdots + \hat{\phi}_p y_1) + (\hat{\theta}_1 e_p + \hat{\theta}_2 e_{p-1} + \cdots + \hat{\theta}_q e_{p-q+1}) \\ e_{p+2} &= y_{p+2} - (\hat{c} + \hat{\phi}_1 y_{p+1} + \cdots + \hat{\phi}_p y_2) + (\hat{\theta}_1 e_{p+1} + \hat{\theta}_2 e_p + \cdots + \hat{\theta}_q e_{p-q+2}) \\ &\vdots \\ e_n &= y_n - (\hat{c} + \hat{\phi}_1 y_{n-1} + \cdots + \hat{\phi}_p y_{n-p}) + (\hat{\theta}_1 e_{n-1} + \hat{\theta}_2 e_{n-2} + \cdots + \hat{\theta}_q e_{n-q}) \\ e_p &= e_{p-1} = \cdots = e_{p-q+2} = e_{p-q+1} = 0 \end{aligned}$$

この残差系列 $\{e_t\}$ をプロットすることによって一定の判断ができるが、残差系列 $\{e_t\}$ を用いてモデル診断のための統計量を計算し、その値によってモデルの妥当性を検定することもできる。

統計量としては、ボックス=ピアス統計量 Q もしくはその修正版であるリュング=ボックス統計量 \tilde{Q} が用いられる。

【 $ARMA(p, q)$ の検定統計量 】

$$\text{ボックス=ピアス統計量} \quad Q = m \sum_{j=1}^K \hat{\rho}_j^2(e) \quad (2.49)$$

$$\text{リュング=ボックス統計量} \quad \tilde{Q} = m(m+2) \sum_{j=1}^K \frac{\hat{\rho}_j^2(e)}{m-j} \quad (2.50)$$

ここで、 $\hat{\rho}_j(e)$ は時系列データ $\{e_t\}$ から計算される時差 j の標本自己相関、 m は推定で使われるデータ数、 K は推定されたパラメータの数よりも大きな整数である。

この統計量は $\{Y_t\} \sim ARMA(p, q)$ の下で漸近的に自由度 $K - p - q$ のカイ二乗分布 $\chi^2(K - p - q)$ に従うことが分かっているので、これを用いて検定を行うことができる。(例えば、帰無仮説を $H_0 : \{Y_t\} \sim ARMA(p, q)$ 、対立仮説を $H_1 : H_0$ の否定 として、有意水準5%の検定であれば、 Q (もしくは \tilde{Q}) $\geq \chi^2_{5\%}(K - p - q)$ (上側5%点)の時に H_0 を棄却、となる。)

また、残差を用いない方法としては情報量基準によるものがある。

この方法はあらかじめ候補となる p, q を絞り込むのではなく、いくつかの候補すべてについて尤度関数を用いて定義される算出式に基づく値を算出し、その値の大小によって候補のモデルの中から最も妥当と思われるモデルの選択を行うものであり、最も有名なものが *AIC* 基準 (Akaike's Information Criterion 赤池情報量基準) である。

情報量基準によるモデルの診断については参考文献3を参照されたい。

2.4.4 時系列モデルに基づく予測

最後に、時点 n までの時系列データ $\{y_t\}$ が与えられたときの h 時点先の Y_{n+h} を予測する方法について説明しておく。

Y_{n+h} の予測としては、自然な方法として条件付期待値 $E(Y_{n+h} | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots)$ を用いることが考えられるが、これが実際に予測値として妥当であることは次のように確認することができる。

Y_{n+h} の任意の予測値を \hat{y}_{n+h} として、平均二乗誤差 $E((Y_{n+h} - \hat{y}_{n+h})^2 | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots)$ を計算すると、 $\mu_{n+h} = E(Y_{n+h} | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots)$ として、

$$\begin{aligned} E((Y_{n+h} - \hat{y}_{n+h})^2 | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots) &= E((Y_{n+h} - \mu_{n+h} + \mu_{n+h} - \hat{y}_{n+h})^2 | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots) \\ &= E((Y_{n+h} - \mu_{n+h})^2 | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots) \\ &\quad + 2(\mu_{n+h} - \hat{y}_{n+h})E((Y_{n+h} - \mu_{n+h}) | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots) \\ &\quad + (\mu_{n+h} - \hat{y}_{n+h})^2 \\ &= E((Y_{n+h} - \mu_{n+h})^2 | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots) + (\mu_{n+h} - \hat{y}_{n+h})^2 \end{aligned} \quad (2.51)$$

となるので、予測値 \hat{y}_{n+h} としては上式第二項の値をゼロとする $\hat{y}_{n+h} = \mu_{n+h}$ が平均二乗誤差の最小値を与えることが分かる。この結果はどのような時系列モデルに対しても成り立つ。

パラメータが未知であることを前提とすれば、 $E(Y_{n+h} | Y_n = y_n, Y_{n-1} = y_{n-1}, \dots)$ の式にパラメータの推定値を用いたものが予測値となる。

各時系列モデルで条件付期待値を計算すると次のようになるので、時系列データ $\{y_t\}$ および誤差データ $\{e_t\}$ を用いて下式に従って順次計算をすれば予測値を求めることができる。

自己回帰モデル $AR(p)$ の予測量

条件付期待値 $E(Y_{n+h} | Y_n, Y_{n-1}, \dots)$ を $h = 1$ から順番に求めると次のようになる。

まず、 $h = 1$ の場合を考えると、 $Y_{n+1} = \phi_0 + \phi_1 Y_n + \dots + \phi_p Y_{n-p+1} + \varepsilon_{n+1}$ を用いれば、

$$\begin{aligned}
E(Y_{n+1} | Y_n, Y_{n-1}, \dots) &= E(\phi_0 + \phi_1 Y_n + \dots + \phi_p Y_{n-p+1} + \varepsilon_{n+1} | Y_n, Y_{n-1}, \dots) \\
&= \phi_0 + \phi_1 Y_n + \dots + \phi_p Y_{n-p+1} + E(\varepsilon_{n+1}) \\
&= \phi_0 + \phi_1 Y_n + \dots + \phi_p Y_{n-p+1}
\end{aligned} \tag{2.52}$$

となる。つまり、 Y_{n+1} の式の誤差項をゼロとしたものが $h=1$ の予測量 \hat{Y}_{n+1} である。

ここで、 $Y_{n+1} = \hat{Y}_{n+1} + \varepsilon_{n+1}$ となっていることに注意されたい。

次に、 $h=2$ の場合を考える。 $Y_{n+1} = \hat{Y}_{n+1} + \varepsilon_{n+1}$ を用いれば、

$$\begin{aligned}
Y_{n+2} &= \phi_0 + \phi_1 Y_{n+1} + \phi_2 Y_n + \dots + \phi_p Y_{n-p+2} + \varepsilon_{n+2} \\
&= \phi_0 + \phi_1 (\hat{Y}_{n+1} + \varepsilon_{n+1}) + \phi_2 Y_n + \dots + \phi_p Y_{n-p+2} + \varepsilon_{n+2}
\end{aligned}$$

となるので、これを用いて、

$$\begin{aligned}
E(Y_{n+2} | Y_n, Y_{n-1}, \dots) &= E(\phi_0 + \phi_1 (\hat{Y}_{n+1} + \varepsilon_{n+1}) + \dots + \phi_p Y_{n-p+2} + \varepsilon_{n+2} | Y_n, Y_{n-1}, \dots) \\
&= \phi_0 + \phi_1 \hat{Y}_{n+1} + \phi_2 Y_n + \dots + \phi_p Y_{n-p+2} + E(\phi_1 \varepsilon_{n+1}) + E(\varepsilon_{n+2}) \\
&= \phi_0 + \phi_1 \hat{Y}_{n+1} + \phi_2 Y_n + \dots + \phi_p Y_{n-p+2}
\end{aligned} \tag{2.53}$$

となる。つまり、 Y_{n+2} の式の Y_{n+1} に予測量 \hat{Y}_{n+1} を用い、誤差項をゼロとしたものが $h=2$ の予測量 \hat{Y}_{n+2} である。

$h=3$ の場合は、 $h=2$ の結果から $Y_{n+2} = \hat{Y}_{n+2} + \phi_1 \varepsilon_{n+1} + \varepsilon_{n+2}$ が成り立っていることが確認できるので、 $h=2$ と同様に、

$$E(Y_{n+3} | Y_n, Y_{n-1}, \dots) = \phi_0 + \phi_1 \hat{Y}_{n+2} + \phi_2 \hat{Y}_{n+1} + \phi_3 Y_n + \dots + \phi_p Y_{n-p+3} \tag{2.54}$$

が導かれ、この場合も、 Y_{n+3} の式の Y_{n+1}, Y_{n+2} に予測量 $\hat{Y}_{n+1}, \hat{Y}_{n+2}$ を用い、誤差項をゼロとしたものになることが分かる。

$h=4$ 以降についても同様であり、以上をまとめて表現すると次のようになる。

【AR(p)の予測量】

予測量 \hat{Y}_{n+h} ($= E(Y_{n+h} | Y_n, Y_{n-1}, \dots)$)は、 Y_{n+h} に関する(2.1)式において $Y_{n+1}, \dots, Y_{n+h-1}$ に予測量 $\hat{Y}_{n+1}, \dots, \hat{Y}_{n+h-1}$ を代入し、誤差項 ε_{n+h} をゼロとしたものとなる。

移動平均モデルMA(q)の予測量

移動平均モデルの場合はAR(∞)表現を用いて自己回帰モデルと同様に考えることもできるが、MA(q)の定義式(2.18)のまま、誤差項 $\varepsilon_n, \varepsilon_{n-1}, \dots$ を既知として(誤差項 $\varepsilon_n, \varepsilon_{n-1}, \dots$ による条件付期待値として)考えた方が簡単である。この仮定の下では Y_n, Y_{n-1}, \dots も既知として考えていることになる。

【MA(q)の予測量】

予測量 \hat{Y}_{n+h} ($= E(Y_{n+h} | \varepsilon_n, \varepsilon_{n-1}, \dots)$)は、 Y_{n+h} に関する(2.18)式において、誤差項 $\varepsilon_{n+1}, \dots, \varepsilon_{n+h}$ をゼロとしたものとなる。

なお、誤差項 $\varepsilon_n, \varepsilon_{n-1}, \dots$ は直接観測されるものではないため、パラメータ推定等の結果として得られる残差 e_n, e_{n-1}, \dots を用いて計算するものとなる。

自己回帰移動平均モデル $ARMA(p, q)$ の予測量

自己回帰モデルおよび移動平均モデルと同様に考えれば、次の結果が導かれる。

【 $ARMA(p, q)$ の予測量 】

予測量 $\hat{Y}_{n+h} (= E(Y_{n+h} | Y_n, Y_{n-1}, \dots, \varepsilon_n, \varepsilon_{n-1}, \dots))$ は、 Y_{n+h} に関する(2.33)式において、 $Y_{n+1}, \dots, Y_{n+h-1}$ に予測量 $\hat{Y}_{n+1}, \dots, \hat{Y}_{n+h-1}$ を代入し、誤差項 $\varepsilon_{n+1}, \dots, \varepsilon_{n+h}$ をゼロとしたものとなる。

2.5 参考文献

- 1 . 石村貞夫、ステファニーリヒャルト著「Excel でやさしく学ぶ時系列 - 明日を支配する時系列」東京図書 (2002)
- 2 . 田中勝人著「計量経済学」岩波書店 (1998)
- 3 . 山本拓著「経済の時系列分析」創文社 (1988)
- 4 . A・C・ハーベイ著 国友直人・山本拓 訳「時系列モデル入門」東京大学出版会 (1999)

以 上

図2.2

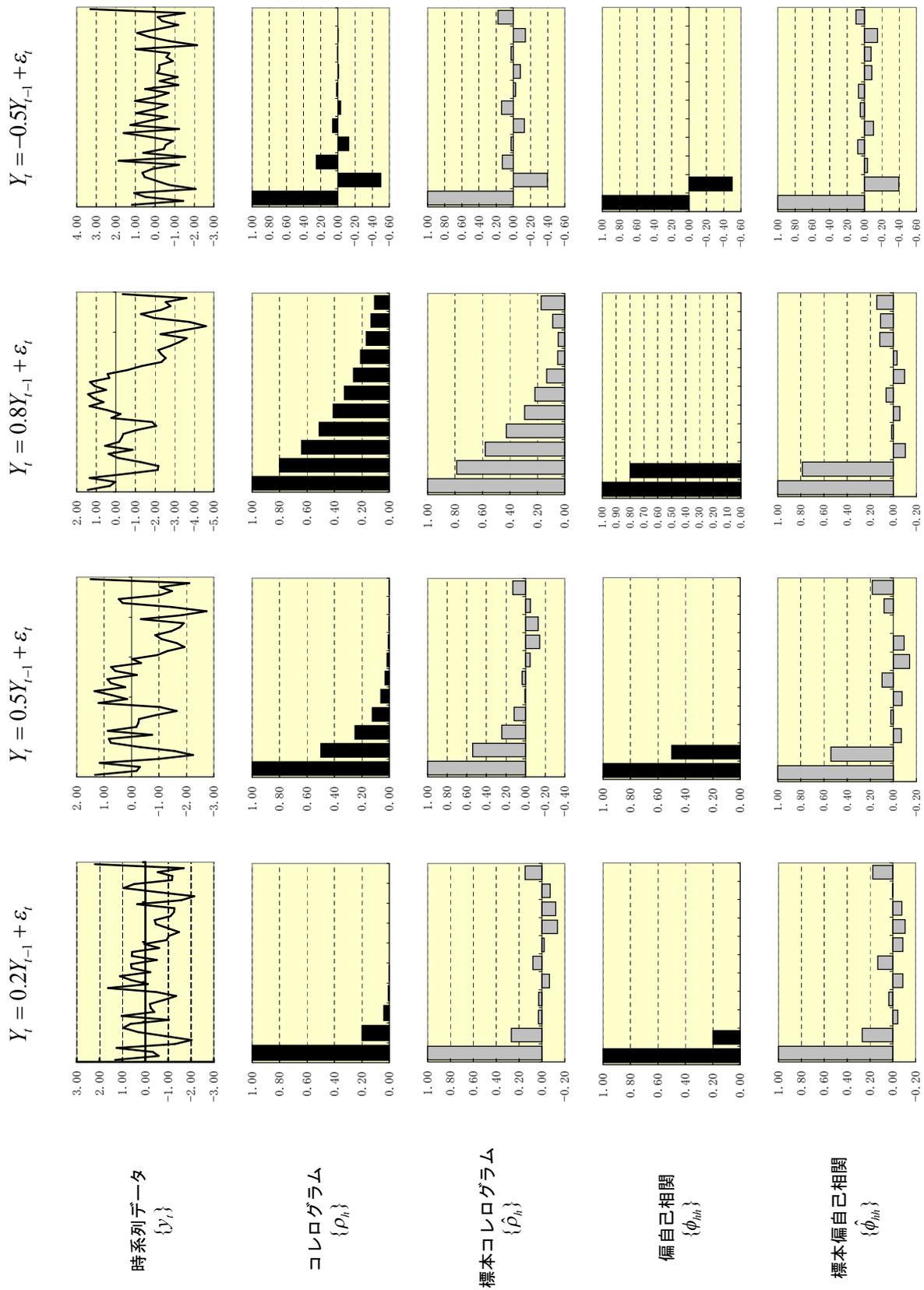


図2.3

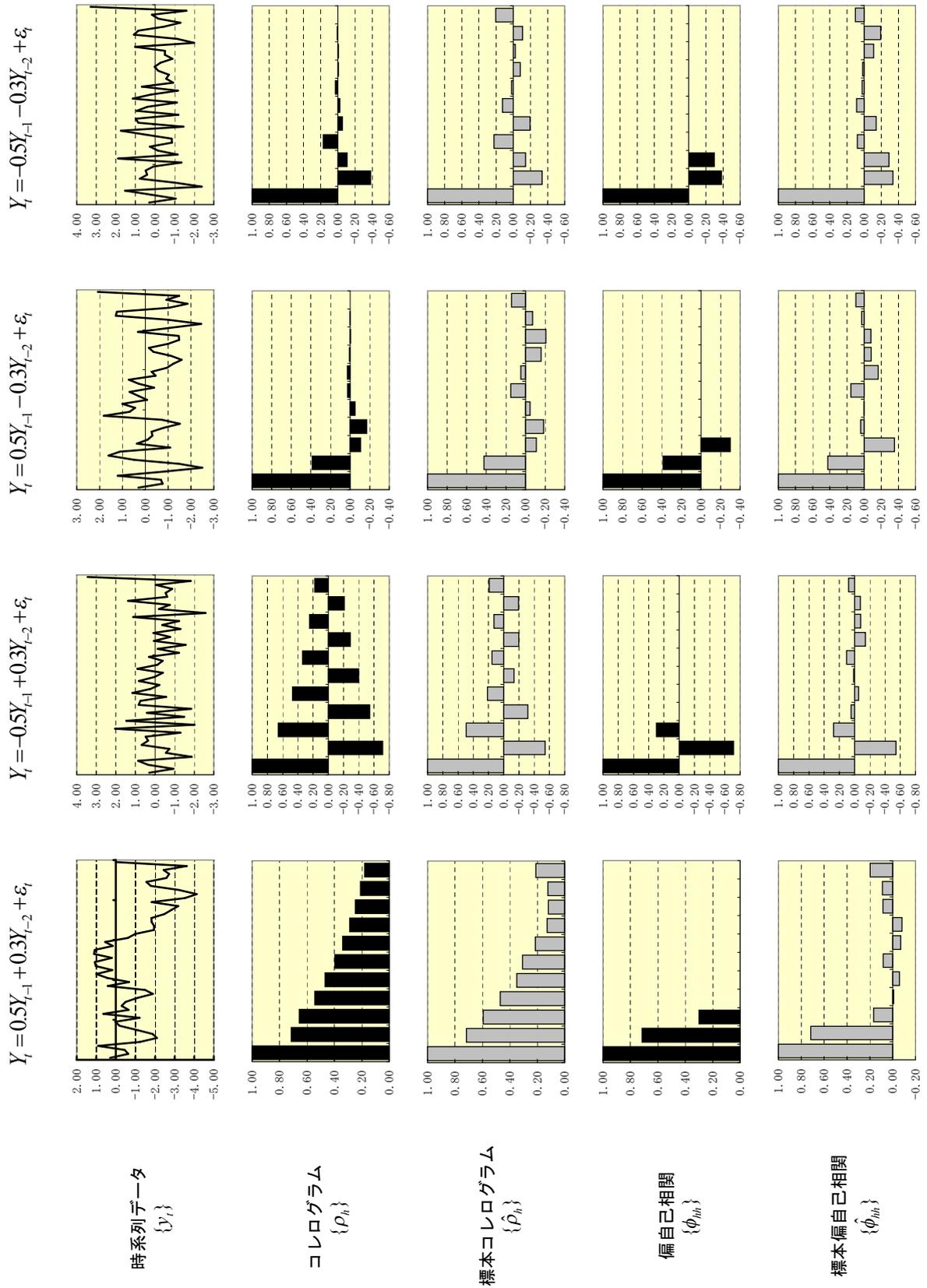


図2.4

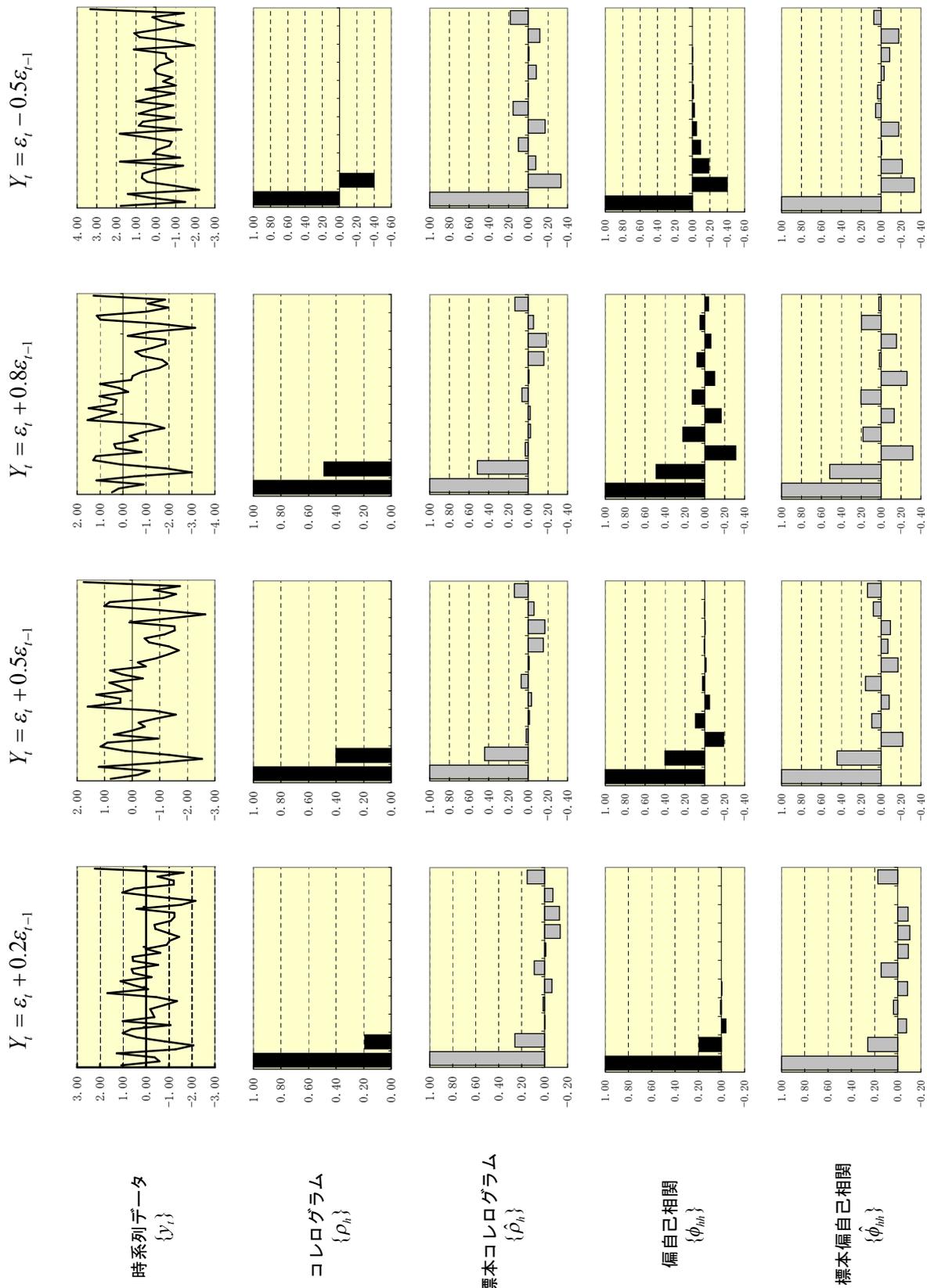
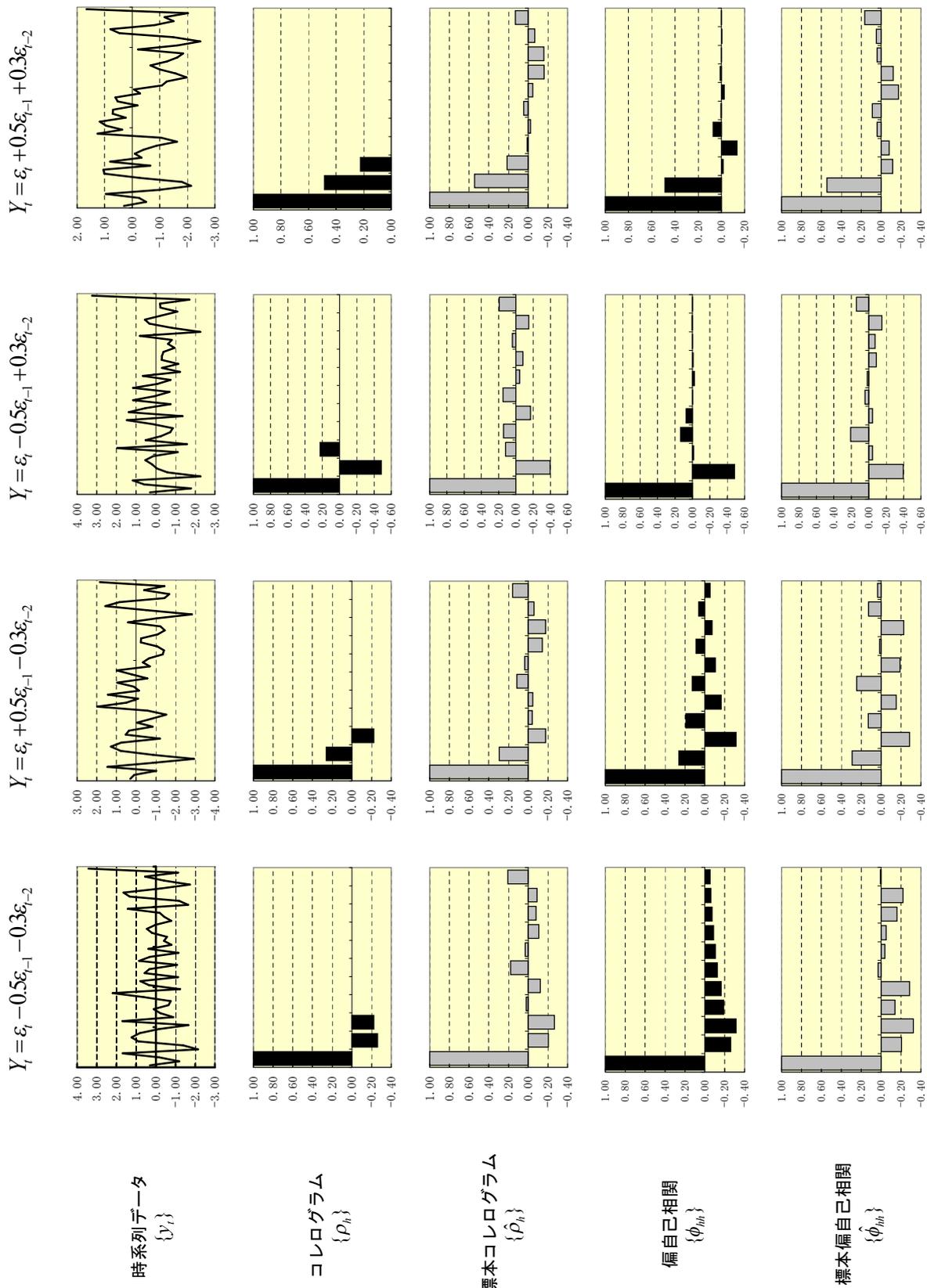


図2.5



第3章 確率過程

3.1 確率過程とは

近年、我々アクチュアリーは、リスク管理等の際に決定論的手法だけではなく、確率論的手法も用いるようになってきた。例えば、保険計理人の将来収支分析や変額年金の最低保証の実務において、確率論的モデルによる分析を行うこともできる。その際に数学的モデルとしてよく利用されるのが、**確率過程**(stochastic process)である。本章では、様々な確率過程のモデルを紹介する。

一般的に確率過程とは、様々な時刻で観察した対象を表現するモデルであり、時間の経過とともに、対象が確率的にどのように変化・変動するかを確率変数で表現しているものである。例えば、気温、債券の価格や株式配当、毎日もしくは毎年の保険事故発生件数、保険会社のサープラス等が挙げられる。(なお、気温のように時間に対して連続的に変化していくものを**連続時間型モデル**、毎日の保険事故発生件数のように一定の時間毎の数量を表すモデルを**離散時間型モデル**という。)

確率過程は、時刻を表すパラメーター t を用いて、確率変数の列 $\{X_t\}$ で表す。また、確率過程 $\{X_t\}$ の一つの実現値 x_t (t を引数とする関数)を**見本関数**(sample function)または**標本関数**という。

3.2 マルコフ連鎖

3.2.1 マルコフ連鎖とマルチンゲール

次のようなコイントスゲームを考える。最初に5点持っており、コインを1枚投げ、表が出たら1点加点(+1)され、裏が出たら1点減点(-1)されるものとする。このゲームの得点の推移を考えると、以下の条件を満たしていることが分かる。

将来のある時点の得点は、現在の得点のみに依存し、過去の得点の推移には依存していない。

例えば、最初に3回のコイントスした結果、裏・裏・表の順で出て、得点の推移が5 - 4 - 3 - 4となった場合と、表・裏・裏と出て5 - 6 - 5 - 4となった場合と、どちらの場合でも将来の得点の状況は変わらない。次のコイントスの結果、何点になるかということは、現在の得点(4点)のみによって決まり、過去の得点がどうであったかは一切関係しない。

この上記の特徴、つまり、『将来の状況は過去の推移に依存せず、ただ現在の状況のみに依存している』という特徴は、**マルコフ性** (Markov property) と呼ばれる。つまり、このゲームはマルコフ性を満たしているといえる。このマルコフ性を満たす確率過程を**マルコフ過程** (Markov process) または、**マルコフ連鎖** (Markov chains) という。なお、マルコフ性を数学的に記載すると、以下のとおりである。(以下は離散時間型モデルの場合)

任意の時刻 $t_1 < \dots < t_n < t$ および任意の x_{t_1}, \dots, x_{t_n} に対して、

$$P(N_t \leq x | N_{t_1} = x_{t_1}, \dots, N_{t_n} = x_{t_n}) = P(N_t \leq x | N_{t_n} = x_{t_n})$$

マルコフ過程という概念は非常に重要である。何故なら、マルコフ過程でない確率過程は(特に連続時間型の確率過程は) 算式によってモデル化するのが困難だからである。実際、後に登場するポアソン過程、ブラウン運動はマルコフ過程である。

ここで、先程のコイントスの問題をもう一度考えよう。コインで表が出る確率と裏が出る確率が、それぞれ $1/2$ であるとする。最初3回のコイントスの結果、現在の得点が4点であるとしよう。4回目のコイントスをした後の得点の期待値 $E(N_4 | N_3 = 4)$ は何点だろうか？

$1/2$ の確率で表が出て得点は5点になり、 $1/2$ の確率で裏が出て得点は3点になる。よって、期待値 $E(N_4 | N_3 = 4)$ は、 $5 \times 0.5 + 3 \times 0.5 = 4$ により4点である。このゲームでは、常に次の等式が成り立つことが容易に理解できるだろう。

$$E(N_{t+1} | N_1, \dots, N_t) = N_t$$

この等式が成立する確率過程を**マルチンゲール** (martingale) という。マルチンゲールは『公平な賭け』という考え方から生まれた概念である。マルチンゲールには、後述の標準ブラウン運動等がある。

3.2.2 推移確率行列とチャップマン・コルモゴロフの方程式

さて、ある会社で、会社を病欠する人の人数を把握したいとする。会社の従業員が健康に出社するか、病欠となるかを、以下のようにモデル化して分析しよう。

「健康で出社」を1で表現し、「病欠」を2で表現する。 $X(n)$ をある従業員の n 日目の状態とする。

さて、ある日の翌日に出社するか病欠となるかは、その日の健康状態のみに依存すると仮定する。例えば、健康な日の翌日に出社する確率は0.98で、逆に病欠の日の翌日に出社する確率は0.70しかないものとする。すると、 $\{X(n), n = 0, 1, 2, \dots\}$ はマ

ルコフ連鎖となる。さて、これを、行列を用いて表現する。

${}_n P_{ij}$ を状態 i の従業員が n 日後に状態 j になる確率とする。

特に、 P_{ij} を状態 i の従業員が翌日に状態 j になる確率とする。

このとき、 $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$ を推移確率行列と呼ぶ。¹

例題 3.1

上記において、推移確率行列 P を求めよ。

解答 3.1

上記の仮定より、 $P_{11} = 0.98$ 、 $P_{21} = 0.70$ である。

よって、 $P = \begin{pmatrix} 0.98 & 0.02 \\ 0.70 & 0.30 \end{pmatrix}$ である。

例題 3.2

「健康で入社」している者が、2 日後に「病欠」になっている確率 ${}_2 P_{12}$ および「健康で入社」している者が、3 日後に「病欠」になっている確率 ${}_3 P_{12}$ を求めよ。

解答 3.2

$${}_2 P_{12} = P_{11} \times P_{12} + P_{12} \times P_{22} = 0.98 \times 0.02 + 0.02 \times 0.3 = 0.0256$$

$${}_3 P_{12} = {}_2 P_{11} \times P_{12} + {}_2 P_{12} \times P_{22} = 0.9744 \times 0.02 + 0.0256 \times 0.3 = 0.027168$$

上記の例題からも分かるように、より一般的に次式が成立する。

$${}_{n+m} P_{ij} = {}_n P_{i1} \times {}_m P_{1j} + {}_n P_{i2} \times {}_m P_{2j} = \sum_{k=1,2} {}_n P_{ik} \times {}_m P_{kj}$$

この公式をチャップマン・コルモゴロフの方程式(Chapman-Kolmogorov equation)という。なお、チャップマン・コルモゴロフの方程式は、次式のように表現されることもある。

$$P(N_t = x) = \sum_{\text{all } x'} P(N_{t'} = x') \times P(N_t = x | N_{t'} = x')$$

このとき、このチャップマン・コルモゴロフの方程式を用いれば、状態 i の者が n 日

¹ 確率過程が定常性を持つ場合にのみ、推移行列による表現が可能となる。なお、定常性の説明については、3.3.1 節にて記載しているので、ご覧頂きたい。

後に状態 j になる確率 ${}_n P_{ij}$ は、 P^n の (i, j) 成分であることを証明できる。(証明は章末問題としたので、各自証明せよ。)

例 3.1

健康な従業員が一週間後に病欠である確率は、

$$P^7 = \begin{pmatrix} 0.972226 & 0.027774 \\ 0.972091 & 0.027909 \end{pmatrix} \text{の}(1,2)\text{成分であるので、} 0.027774 \text{である。}$$

さて、状態 i の者が n 日後に状態 j になる確率 ${}_n P_{ij}$ は、 n が十分大きくなれば、定常状態になり、ある確率に収束するものと考えられる。その極限確率 ${}_{\infty} P_{ij} = \lim_{n \rightarrow \infty} {}_n P_{ij}$ を求めよう。

$$P^{\infty} = \begin{pmatrix} {}_{\infty} P_{11} & {}_{\infty} P_{12} \\ {}_{\infty} P_{21} & {}_{\infty} P_{22} \end{pmatrix} \text{とすると、} P^{\infty} P = P^{\infty} \text{が成立する。これを解いて、}$$

$$P^{\infty} = \begin{pmatrix} 0.972222 & 0.027778 \\ 0.972222 & 0.027778 \end{pmatrix} \text{となる。}^2$$

3.2.3 推移行列の応用

さて、この従業員の出勤・病欠モデルを、もう少し一般化することを考える。つまり、ある日に健康に出勤するか病欠となるかは、前日の体調だけでなく、2日前の体調にも影響を受けるものとする。例えば、2日続けて健康な人は、本日は健康だが昨日は病欠の人より、次の日に出勤する可能性が高いものとする。また、2日続けて病欠している人は、1日だけ病欠の人より病気が重く、次の日も病欠する可能性が高いものとする。

具体的には、以下のとおりの確率としよう。本日は健康で、昨日も健康な人が、次の日に出勤する確率は 0.99。本日は健康だが、昨日は病欠の人が、次の日に出勤する確率は 0.95。本日は病欠だが、昨日は健康だった人が、次の日に出勤する確率は 0.80。

² 別解として以下のような解答もある。

$\pi = (\pi_1 \quad \pi_2)$ を、定常状態になったときの健康な従業員と病欠の従業員の比を表す行列とする。つまり、次のとおりとする。

π_1 : 定常状態時における健康な従業員の占める割合

π_2 : 定常状態時における病欠の従業員の占める割合 (よって、 $\pi_1 + \pi_2 = 1$)

$$\pi = \pi \cdot P \text{を解いて、} \pi = (0.972222 \quad 0.027778)$$

本日も昨日も病欠の人が、次の日に出社する確率は、0.60。

さて、先程と同様に「健康で出社」と「病欠」という2つの状態でのモデルを表現すると、このモデルはマルコフ過程にはならない。しかし、以下のとおり4つの状態モデル化すれば、マルコフ過程のモデルで表現することができる。

状態1：本日は健康で、昨日も健康

状態2：本日は健康だが、昨日は病欠

状態3：本日は病欠だが、昨日は健康

状態4：本日も昨日も病欠

例題 3.3

このモデルの推移確率行列を求めよ。

解答 3.3

$$P = \begin{pmatrix} 0.99 & 0 & 0.01 & 0 \\ 0.95 & 0 & 0.05 & 0 \\ 0 & 0.80 & 0 & 0.20 \\ 0 & 0.60 & 0 & 0.40 \end{pmatrix}$$

例題 3.4

本日も昨日も健康な従業員が、1週間後に病欠となっている確率を求めよ。

解答 3.4

本日も昨日も健康な従業員が一週間後に病欠である確率は、

$$P^7 = \begin{pmatrix} 0.976086 & 0.010239 & 0.010271 & 0.003404 \\ 0.975770 & 0.010426 & 0.010286 & 0.003518 \\ 0.972705 & 0.012250 & 0.010426 & 0.004619 \\ 0.969998 & 0.013856 & 0.010552 & 0.005594 \end{pmatrix}$$

の(1,3)成分と(1,4)成分の合計であるので、0.013675である。

3.3 ポアソン過程

3.3.1 ポアソン過程

交通調査として、ある道路の脇で1時間、通り過ぎる車の台数を数えるものとする。 t 時間経過後までに、目の前を通り過ぎた車の台数を N_t で表すことにしよう。このように、何らかの事象（この場合、目の前を車が通り過ぎるという事象）の発生回数を数える確率過程 $\{N_t\}$ を計数過程 (counting process) という。なお、数学的に計数過

程を定義すれば、以下のとおりとなる。

連続時間型の確率過程 $\{N_t\}$ が計数過程であるとは、

以下の3つの条件を満たすことをいう。

- $N_0 = 0$
- $\{N_t\}$ の見本関数は非負の整数値で、 t に関し単調増加。
- $\{N_t\}$ の見本関数は、右連続。³

さて、この計数過程に対し、以下の3条件を仮定するのは、不自然ではないだろう。

たまたまある時間帯に通った車が多かったとしても、

その後から来る車が多くなったりまたは少なくなったりしない。

最初の1分間でも、最後の1分間でも通り過ぎる車の台数の可能性は同じ。

同一時刻に2台以上の車が通り過ぎることはない。

条件 1 は、サイコロで説明するのが分かりやすいだろう。サイコロを1秒間に1回振り、1が出る回数を数えていくこととする。サイコロで1が出る確率は(その他の目と同じく)6分の1である。だから1分間では、1の目は平均して10回くらい出るだろう。さて、ある1分間に1の目が20回出たとしよう。または、2回しか出なかったとしよう。さて、次の1分間に1が出る回数は、10回より多くなるだろうか？それとも少なくなるだろうか？

イカサマサイコロであるという可能性を否定すれば、前の1分間に1が多く出ようと、または少なく出ようと、次に1が出る確率はやはり6分の1であり、次の1分間に1が出る回数は、やはり平均して10回くらいなのである。つまり、サイコロは条件 1 を満たしていることになる。この条件 1 を満たす確率過程を一般的に加法過程 (additive process) または独立増分過程という。なお、加法過程は、マルコフ過程となっている。(証明は章末問題とする。)

逆に、条件 1 を満たさない例としては、商店街等で行われるくじ引き等が挙げられる。1秒間に1回くじを引き、当たりの個数を数えていくこととする。ただし、当たりを引いても、はずれを引いても、くじは元に戻さないものとする。最初、全部で3600

³ 関数 $f(x)$ が右連続とは、任意の点において、その点の右側から近づくと、その点に収束するということである。数学的に記載すれば以下のとおりである。

$$\lim_{t \rightarrow 0} f(x+t) = f(x) \quad (t > 0)$$

つまり、 $f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$ は右連続であり、 $f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$ は右連続でない。

本のくじが入っており、そのうち 60 本が当たりくじであるとしよう。つまり、当たりの出る確率は $1/60$ である。さて、最初の 1 分間に当たりくじが (奇跡的に) 60 本出たとしよう。すると、次の 1 分間に出る当たりくじは、(全ての当たりくじが既に出てしまったことになるから、) 0 本である。逆に最初の 10 分間に当たりくじが 1 本も出なかったとしよう。すると、当たりの出る確率は、(はずれくじが減ったことにより) $60/3000 = 1/50$ に上がっていることになる。よって、一般的なくじ引きでは条件は満たされていないことが分かる。

条件は、通り過ぎる車の台数は、時刻ではなく時間の幅のみに依存することを意味している。ラッシュ・アワー等の例外的な時間を無視すれば、自然な仮定といえるであろう。逆に、この 1 時間の間にラッシュ・アワーが含まれているならば、この仮定は成立しないであろう。この条件を満たす確率過程を、一般に**定常過程**(stationary process) という。

条件に関しては、ほぼ同一時刻に 2 台以上の車が通り過ぎることはあるが、写真判定のようなものを用いることを考えれば、まったく同時に 2 台以上の車が通り過ぎることはなく、ほんの僅かの差はあると考えれば自然だろう。

上記の から の条件を満たす計数過程を**ポアソン過程**(Poisson process) という。ポアソン過程は、ポアソン分布に対応する確率過程である。

さて、ポアソン過程を、数学的な表現で記載すると、以下のとおりとなる。

独立性: $0 \leq s < t \leq u < v$ $N_t - N_s$ と $N_v - N_u$ は独立
 定常性: $N_{s+t} - N_t$ と N_s は同じ分布に従う。
 $P(N_{t+h} - N_t \geq 2) = o(h)^4$

さて、ポアソン過程に関しては、損害保険においてクレーム件数を分析する際等に使用されるため、損保数理のテキストで説明されている。よって、詳細な説明は、そちらをご覧ください。また、ポアソン過程の一般化された確率過程として**一般ポアソン過程**がある。こちらに関しても、損保数理のテキストをご覧ください。

3.3.2 再生過程

ポアソン過程において、クレームが発生してから、次のクレームが発生するまでの時間 t は、指数分布 $\Gamma(1, \lambda) = \lambda e^{-\lambda t}$ に従う (証明は演習問題)。これを一般化し、あるクレームから次のクレームまでの間の時間が同じ確率分布に従う計数過程を、**再生過程** (renewal process) という。当然のことであるが、ポアソン過程は再生過程である。

⁴ $o(h)$ とは、 h より十分小さいことを意味する記号である。数学的に表現すれば、

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0 \text{ を満たす関数である。}$$

3.4 ブラウン運動

3.4.1 ブラウン運動

さて、次のようなマルコフ連鎖を考える。

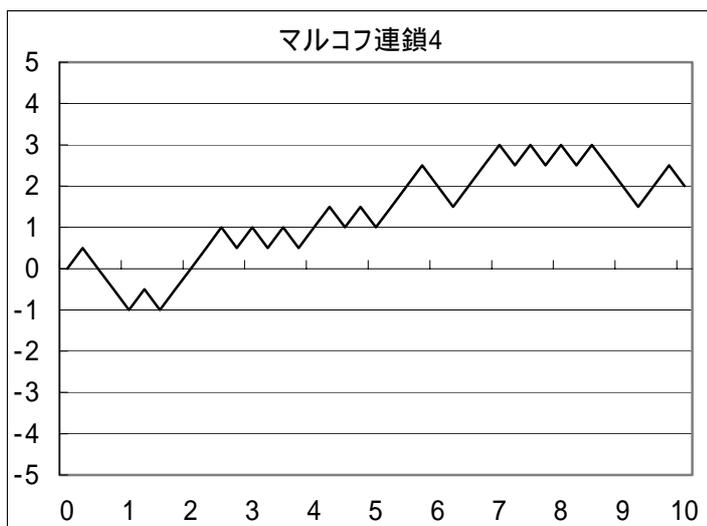
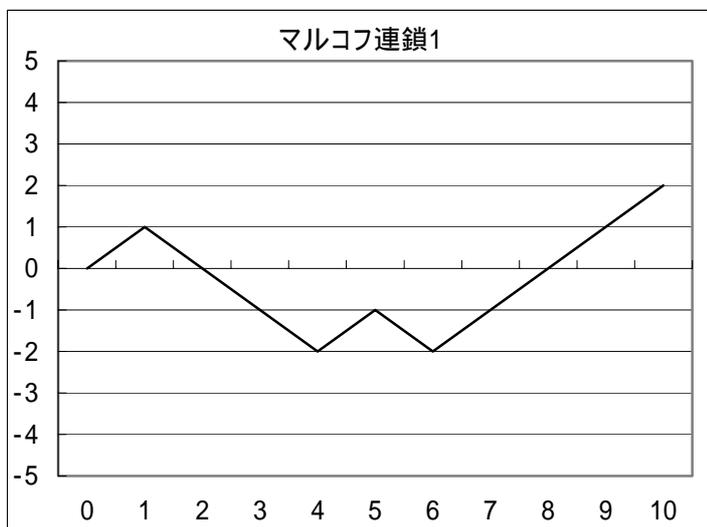
マルコフ連鎖 1 : 1 秒に 1 回コインを投げ、表は + 1 点、裏は - 1 点

マルコフ連鎖 2 : 1/2 秒に 1 回コインを投げ、表は $+\frac{1}{\sqrt{2}}$ 点、裏は $-\frac{1}{\sqrt{2}}$ 点

.....

マルコフ連鎖 n : 1/n 秒に 1 回コインを投げ、表は $+\frac{1}{\sqrt{n}}$ 点、裏は $-\frac{1}{\sqrt{n}}$ 点

図 3.1





例題 3.5

マルコフ連鎖 n において、 t 秒間に得られる得点の期待値および分散を求めよ。ただし、 $t = t' + t''$ ($t' \times n$ は整数かつ $0 \leq t'' < \frac{1}{n}$) とする。

解答 3.5

1 回コインを投げて得られる得点を X で表し、 t 秒間に得られる得点を S で表すと、

$$E(X) = 0$$

$$V(X) = \frac{1}{n}$$

t 秒間にコイントスは $t' \times n$ 回行われるので、

$$E(S) = t'n \times E(X) = 0$$

$$V(S) = t'n \times V(X) = t'$$

よって、期待値は 0、分散は t' 。

さて、ここで n を無限大まで大きくすることを考える。すると、それはある確率過程となる。その確率過程はブラウン運動とよばれる確率過程である。

ブラウン運動 (Brownian motion) とは、以下を満たす確率過程 $\{X_t\}$ をいう。

連続性: X_t の見本関数は t に関する連続関数

独立増分性: $0 \leq s < t \leq u < v$ $X_t - X_s$ と $X_v - X_u$ は独立

定常増分性: $X_{s+t} - X_t$ と X_s は同じ分布に従う。

なお、ブラウン運動は、以下の定理 3.1 のとおり正規分布に対応する確率過程であり、正規分布がガウス分布とも呼ばれることから、**ガウス過程** (Gaussian process) とも呼ばれる。

ブラウン運動がさらに次の条件を満たすときに、**標準ブラウン運動** (standard Brownian motion) または**ウィーナー過程** (Wiener process) という。

$$X_0 = 0$$

$$E(X_t) = 0$$

$$V(X_t) = t$$

なお、ウィーナー過程は、標準正規分布に対応する確率過程である。

定理 3.1

X_t をブラウン運動とすると、 $X_t - X_0$ は、平均 $E(X_t - X_0)$ ・分散 $V(X_t - X_0)$ の正規分布に従う。

なお、証明は難解であるため省略する。⁵

⁵ 証明は、Gihman and Skorohod [1980]を参考にして頂きたい。

3.5 モデリングの金融工学への適用

3.5.1 株価モデル

ウィーナー過程は、大変扱いやすい（計算しやすい）という性質をもっており、金融工学等の分野でよく用いられる。例えば、**ブラック・ショールズ式**（The Black-Scholes formula）と呼ばれるオプションのプライシング公式も、このブラウン運動を元に計算を行っている。

ある株の期待リターン⁶が μ 、ボラティリティ（分散）が σ^2 であるとする。期待リターンが μ であるとは、概ね μ という利回りで株価が上昇するということである。しかし、株価というものは、上昇する場合もあれば、下落する場合もあるものである。それを表す概念がボラティリティ σ^2 である。ボラティリティは変動のしやすさを表す概念であり、ボラティリティが高いほど変動しやすいことを表している。逆に、ボラティリティがゼロならば、確実に μ という利回りで株価は上昇し、そこから上下にはまったく変動しないということを表す。

さて、一般的に、株価 S_t は以下のとおりモデル化される。

$$dS_t = S_t \mu dt + S_t \sigma dB_t$$

図 3-2



これは、以下のことを意味している。

dt という大変短い時間における株価の変動を考える。左辺の dS_t とは、その短い時間における株価の上昇額を意味している。数学的に表現すれば、 $dS_t = S_{t+dt} - S_t$ である。

本節は、金融工学の専門的な内容に踏み込むため、アクチュアリー試験の対象外とする。
⁶ 後ほど明らかになるが、ここでの期待リターンとは、通常の年間利回りの意味ではなく、利力または連続複利の意味で用いており、1の資産が t 年後に、 $(1 + \mu)^t$ ではなく、 $e^{\mu t}$ になる点に注意。

その短い時間に株価は、期待リターン μ で運用されるので、 $S_t \times \mu \times dt$ だけリターンが得られることになる。つまり、 $S_t \mu dt$ だけ株価が上昇することを、右辺の第1項は表している。なお、この第1項をドリフト項という。しかし、先にも述べたように株価は上昇するか下落するか分からない。それを表現しているのが右辺第2項である。この第2項は拡散項という。 dB_t がウィーナー過程を表しており、株価が $S_t \sigma dB_t$ だけランダムに変動するということを意味している。なお、 $dB_t = B_{t+dt} - B_t = B_{dt}$ であるから、 $E(S_t \sigma dB_t) = 0$ および $V(S_t \sigma dB_t) = S_t^2 \sigma^2 dt$ である。

以下、この株価の式を理解するために、以下のレンマを用いて式変形を行う。

伊藤のレンマ (Ito's Lemma)

X_t が

$$dX_t = a(X_t, t) \cdot dt + b(X_t, t) \cdot dB_t$$

と表されるとき、 X_t と t の関数 $f(X_t, t)$ は、次の確率微分方程式を満たす。

$$df = \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} \cdot a(X_t, t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \cdot \{b(X_t, t)\}^2 \right) \cdot dt + \frac{\partial f}{\partial x} \cdot b(X_t, t) \cdot dB_t$$

ただし、 $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial x}(X_t, t)$ は、 $f(x, t)$ の x での偏微分 $\frac{\partial f(x, t)}{\partial x}$ に、 X_t を代入したものである。

なお、伊藤のレンマの証明は専門書をご覧ください。

さて、上記の確率微分方程式に、 $f(S_t, t) = \log S_t$ を代入すると、

$$d \log S_t = \left(0 + \frac{1}{S_t} \cdot S_t \mu - \frac{1}{2} \frac{1}{S_t^2} \cdot \{S_t \sigma\}^2 \right) \cdot dt + \frac{1}{S_t} \cdot S_t \sigma \cdot dB_t = \left(\mu - \frac{1}{2} \sigma^2 \right) \cdot dt + \sigma \cdot dB_t$$

これを積分して、⁷

⁷ 本来、確率積分は、通常の積分のように積分できない。例えば、 $dS_t = S_t \mu dt + S_t \sigma dB_t$ を解く際に、 $\frac{dS_t}{S_t} = \mu dt + \sigma dB_t$ を積分して、 $\log S_t = \mu \cdot t + \sigma \cdot B_t$ とは出来ない。確率積分

の際に注意すべき点があり、本来であれば厳密な議論が必要であるが、ここでは省略する。興味のある方は、伊藤積分について勉強していただきたい。伊藤積分に関しては、蓑谷千鳳彦[2000]が詳しく分かりやすい。

$$\log S_t = \left(\mu - \frac{1}{2} \sigma^2 \right) \cdot t + \sigma \cdot B_t$$

よって、次式が成立する。

$$S_t = S_0 \cdot e^{\left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma \cdot B_t}$$

3.5.2 株価モデル⁸

近年、金融工学の発達に伴い、様々な金融商品が開発されている。そのひとつがコール・オプションである。コール・オプションとは、ある特定の商品（例えば、株式）を、将来のある一定時点またはそれ以前に、あらかじめ定められた価格（行使価格）で、購入する権利のことである。例えば、現在 100 円の株式を、1 年後に 110 円で購入する権利である。このコール・オプションを購入した場合、1 年後に株価が 110 円以下ならばその価格で株式を購入すればよいし、110 円以上ならばコール・オプションを行使することにより 110 円で株式を購入することが出来る。つまり、将来購入したい株式の価格上昇リスクを回避することができるのである。このコール・オプションの価格を評価するのに、ブラック・ショールズ式が利用される。具体的には、 $C(S_t, t)$ をコール・オプションの価格、 S_t を現時点 t の株価、 K を時点 T における行使価格、 r を安全資産の収益率⁹として、ブラック・ショールズ式は以下のとおり表される。

$$C(S_t, t) = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

$$\text{ただし、 } d_1 = \frac{\log \frac{S_t}{K} + \left(r + \frac{1}{2} \sigma^2 \right) (T-t)}{\sigma \sqrt{T-t}}, \quad d_2 = \frac{\log \frac{S_t}{K} + \left(r - \frac{1}{2} \sigma^2 \right) (T-t)}{\sigma \sqrt{T-t}}$$

なお、ここで、 $N(d) = \int_{-\infty}^d \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ は、標準正規分布の累積密度関数である。

また、プット・オプションの価格 $P(S_t, t)$ は、以下のとおり表される。

$$P(S_t, t) = K e^{-r(T-t)} N(-d_2) - S_t N(-d_1)$$

ブラック・ショールズ式は、市場が無裁定 (No Arbitrage)¹⁰であるという条件の下

⁸ ブラック・ショールズ・モデルに関しては、養谷千凰彦[2000]が詳しい。

⁹ 注 6 に同じ。ここでの収益率は、利子または連続複利の意味で用いており、1 の資産が t 年後に e^{rt} になることを想定している。

¹⁰ 無裁定とは、裁定機会がないことを指す。裁定機会とは、元手なしで確実にお金を儲ける機会を指す。例えば、2%の利回りで百万円借りることができ、3%の利回りで百万円貸すことができるならば、確実に 1 万円を稼ぐことができる。または、東京証券取引所で 6 万

で導かれるが、その証明は専門書をご覧いただきたい。

例題 3.6

プット・オプションの価格式 $P(S_t, t) = Ke^{-r(T-t)}N(-d_2) - S_tN(-d_1)$ とコール・オプションの価格式 $C(S_t, t) = S_tN(d_1) - Ke^{-r(T-t)}N(d_2)$ の間に、プット・コール・パリティの関係式 $C(S_t, t) + Ke^{-r(T-t)} = P(S_t, t) + S_t$ が成り立つことを証明せよ。

解答 3.6

$N(-d_1) = 1 - N(d_1)$ および $N(-d_2) = 1 - N(d_2)$ を代入すれば、明らかである。

さて、ブラック・ショールズ式には、2種類の弱点がある。

1つ目は、ヨーロピアン型のコール・オプションやプット・オプションの価格は計算できるが、複雑なオプションであるエキゾチック・オプションについては、価格計算が出来ないということである。例えば、満期までの期間に観察される原資産価格の最大値または最小値を用いるルック・バック・オプションや、行使価格が満期までの特定の期間に観察される原資産価格の平均値に等しいアジアン・オプション等である。こういったオプションの評価は、ブラック・ショールズ式が使用できないため、モンテカルロ法等のシミュレーションに頼って価格を算出するしかない。

2つ目は、ブラック・ショールズ式が株価の動きを適切に表現しているかという問題である。ブラック・ショールズ式では、株価の動きを $dS_t = S_t\mu dt + S_t\sigma dB_t$ と仮定している。しかし、倒産等の理由で株が下落する際は、徐々に株価が下がるのではなく、一気に下がる。例えば、1987年10月のブラック・マンデーがそうであった。このような「ジャンプ」は、ウィーナー過程では表現できない。現在では、「ジャンプ」を表現するために、レヴィ過程という確率過程を用いた研究が進んでいる。

3.5.3 株価モデルとレヴィ過程¹¹

ブラック・ショールズ式の長所は、計算が容易である点にある。しかし、ブラック・ショールズ式が株価の動きを適切に表現しているかということ、そうではない面もある。例えば、倒産等の理由で株が下落する際は、徐々に株価が下がるのではなく、一気に下がる。それを、表現するために、加法過程の一種である**レヴィ過程** (Levy process)

円で取引されている株式が、大阪証券取引所で5万円で取引されていれば、その株式を大阪で買い東京で売れば、確実に1万円を稼ぐことが出来る。このように確実にお金を儲ける手段が存在しないことを、無裁定という。

¹¹ 株価モデルに用いるレヴィ過程の分類に関しては、宮原孝夫[2004]が参考になる。

という確率過程を用いた研究が進んでいる。レヴィ過程がウィーナー過程と異なるのは、ウィーナー過程が連続な確率変数であったのに対し、レヴィ過程では不連続な点（ジャンプ）が存在し、価格が上または下へジャンプするのである。

図 3.3



イメージとしては、レヴィ過程は、ウィーナー過程とポアソン過程を組み合わせた確率過程を考えていただいて差し支えない。なお、ウィーナー過程やポアソン過程はレヴィ過程の一種である。（実際には、レヴィ過程とは、不連続な点が存在しても、数学的に確率連続な加法過程であり、第 1 種不連続過程（D 過程）であるものである。D 過程とは、見本関数が各点で右連続かつ各点で有限な左極限を持つ確率過程である。関数 $f(x)$ の各点で有限な左極限を持つとは、ある点において、その点の左側から近づくととき、（その点に収束しなくても良いが、）何らかの点には収束するということである。数学的に記載すれば以下のとおりである。

$$\exists \lim_{t \rightarrow 0} f(x-t) \quad (t > 0)$$

3.5.4 代表的な金利モデル¹²

金利モデルとして有名なものには、**バシチェック・モデル**（Vasicek model）、**ハル・ホワイト・モデル**（Hull-White model）、**コックス・インガソル・ロス・モデル**（Cox Ingersoll Ross model）（C I R モデル）がある（金利モデルとしては、その他に Ho-Lee

¹² 金利モデルに関しては、田畑吉雄[2001]に詳しくまとめられている。また、ハル・ホワイト・モデルに関しては、日本債券信用銀行 商品開発グループ[1996]が詳しい

モデル、Heath-Jarrow-Morton モデル等があるが、ここでは省略する。

金利に関する確率過程は少々難解であるため、ここでは、モデルの導き方およびモデルから導かれる結果について触れずに、単にモデルの形のみを記載する。なお、これらのモデルは、単にモデルが最初から与えられているのではなく、何らかの前提を置いて、その他の一般的な確率過程を変形してやることで求められている。その際、バシチェック・モデル、ハル・ホワイト・モデルは無裁定 (No arbitrage) アプローチ¹³を、CIRモデルは均衡 (Equilibrium) アプローチ¹⁴を採用している。

バシチェック・モデルは、以下のように表される。

$$dr(t) = a(b - r(t))dt + \sigma dB_t$$

a 、 b 、 σ はどれも非負の値であるが、 b は長期的に収束していく金利を表現している。 a は b に収束していく強さを表しており、平均回帰レート (Reversion rate) と呼ばれる。また σ は金利の変動の大きさを表現している。

金利が b よりも大きいとき、第 1 項はマイナスの数である。よって、金利は b に向かって下がっていく。逆に金利が b よりも小さいときは、第 1 項はプラスであり、金利は b に向かって上昇する。

ハル・ホワイト・モデルは、以下のように表される。

$$dr(t) = a\left(\frac{\theta(t)}{a} - r(t)\right)dt + \sigma dB_t$$

ハル・ホワイト・モデルでは、金利の期間構造を $\frac{\theta(t)}{a}$ と捉えている。

CIRモデルは、以下のようなモデルである。

$$dr(t) = a(b - r(t))dt + \sigma\sqrt{r(t)}dB_t$$

バシチェック・モデル、ハル・ホワイト・モデルでは金利がマイナスになってしまうことがあるという弱点があったが、CIRモデルでは常に金利は非負の値をとる。

¹³ 市場は無裁定であるという仮定をもとに導く手法

¹⁴ 効用関数をもとにし、利潤を最大化させる手法

3.6 練習問題

- 1 .マルチンゲールは、任意の k に対して次式が成り立つことを数学的帰納法を用いて証明せよ。

$$E(N_{t+k} | N_1, \dots, N_t) = N_t$$

- 2 . 状態 i の者が n 日後に状態 j になる確率 ${}_n P_{ij}$ は、 P^n の (i, j) 成分であることを、チャップマン・コルモゴロフの方程式を用い、数学的帰納法により証明せよ。
- 3 . 3.2.2 節のモデルにおいて、1,000 人の従業員がいて、本日は 970 人が健康で 30 人が病欠とする。このとき、翌日出社する人数の期待値を求めよ。
- 4 . 3.2.3 節のモデルにおいて、従業員が病欠となっている極限確率を求めよ。
- 5 . 3.2.3 節のモデルにおいて、1,000 人の従業員がいて、本日は、状態 1 : 950 人、状態 2 : 30 人、状態 3 : 15 人、状態 4 : 5 人である。翌日、病欠になる従業員の数の期待値を求めよ。
- 6 . 次のようなコイントスゲームを考える。最初に 3 点持っており、コインを 1 枚投げ、表が出たら 1 点加点 (+ 1) され、裏が出たら 1 点減点 (- 1) されるものとする。点数が 5 点になったら勝ち、点数が 1 点になったら負けで、その時点で終了とする。表が出る確率と裏が出る確率が共に $1/2$ であるとき、5 回のコイントス後に、 i 点になっている確率をそれぞれ求めよ。 ($i = 1, 2, 3, 4, 5$)
- 7 . 問題 6 のゲームで、最初の持ち点が 2 点の場合、5 回のコイントス後に、 i 点になっている確率をそれぞれ求めよ。 ($i = 1, 2, 3, 4, 5$)
- 8 . 問題 6 のゲームで、表が出る確率が a 、裏が出る確率が b とするとき、確率推移行列 P を求めよ。
- 9 . 問題 8 のゲームで、最初の持ち点が i 点であるとき、このゲームに勝つ確率 p_i とし、 $p = {}^T(p_1 \ p_2 \ p_3 \ p_4 \ p_5)$ とする。このとき、 $p = P \cdot p$ が成立することを示せ。

10. 問題8のゲームにおいて、 $p = {}^T(p_1 \ p_2 \ p_3 \ p_4 \ p_5)$ を求めよ。

11. 加法過程は、マルコフ過程であることを証明せよ。

12. ポアソン過程において、あるクレームが発生してから、次のクレームが発生するまでの時間 t が、指数分布 $\Gamma(1, \lambda) = \lambda e^{-\lambda t}$ に従うことを示せ。

3.7 練習問題解答

1. 帰納法より k のとき与式が成立するとして、 $k+1$ のときを証明する。

$$\begin{aligned} E(N_{t+k+1} | N_1, \dots, N_t) &= E(E(N_{t+k+1} | N_1, \dots, N_t, \dots, N_{t+k}) | N_1, \dots, N_t) \quad (\text{一般的に成立}) \\ &= E(N_{t+k} | N_1, \dots, N_t) \quad (\text{マルチンゲール性より}) \\ &= N_t \quad (\text{帰納法の仮定より}) \end{aligned}$$

よって、題意は証明された。

2.

$n=1$ のとき、 $P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$ であるので、題意は自明である。

$n=m$ のとき、題意が成立する ($P^m = \begin{pmatrix} {}^m P_{11} & {}^m P_{12} \\ {}^m P_{21} & {}^m P_{22} \end{pmatrix}$) と仮定して、

$n=m+1$ のとき、題意が成立する ($P^{m+1} = \begin{pmatrix} {}^{m+1} P_{11} & {}^{m+1} P_{12} \\ {}^{m+1} P_{21} & {}^{m+1} P_{22} \end{pmatrix}$) ことを示す。

チャップマン・コルモゴロフの定理より、

$${}^{m+1} P_{ij} = {}^m P_{i1} \times P_{1j} + {}^m P_{i2} \times P_{2j} = \sum_{k=1,2} {}^m P_{ik} \times P_{kj}$$

一方、

$$P^{m+1} = P^m \times P = \begin{pmatrix} {}^m P_{11} & {}^m P_{12} \\ {}^m P_{21} & {}^m P_{22} \end{pmatrix} \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} {}^m P_{11} \times P_{11} + {}^m P_{12} \times P_{21} & {}^m P_{11} \times P_{12} + {}^m P_{12} \times P_{22} \\ {}^m P_{21} \times P_{11} + {}^m P_{22} \times P_{21} & {}^m P_{21} \times P_{12} + {}^m P_{22} \times P_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1,2}^m P_{1k} \times P_{k1} & \sum_{k=1,2}^m P_{1k} \times P_{k2} \\ \sum_{k=1,2}^m P_{2k} \times P_{k1} & \sum_{k=1,2}^m P_{2k} \times P_{k2} \end{pmatrix} = \begin{pmatrix} {}_{m+1}P_{11} & {}_{m+1}P_{12} \\ {}_{m+1}P_{21} & {}_{m+1}P_{22} \end{pmatrix}$$

よって、題意は証明された。

3 . $970 \times 0.98 + 30 \times 0.7 = 971.6$ で 971.6 人。

4 . $\pi = \pi \cdot P$ を解いて、 $\pi = (0.976027 \quad 0.010274 \quad 0.010274 \quad 0.003425)$ であるから、 $0.010274 + 0.003425 = 0.013699$

5 . $950 \times 0.01 + 30 \times 0.05 + 15 \times 0.2 + 5 \times 0.4 = 16.0$ で 16 人

6 . ${}_t p_i$ を t 回のコイントス後に i 点になっている確率とし、

${}_t p = {}^T({}_t p_1 \quad {}_t p_2 \quad {}_t p_3 \quad {}_t p_4 \quad {}_t p_5)$ とすると、

$${}_1 p = {}^T(0 \quad \frac{1}{2} \quad 0 \quad \frac{1}{2} \quad 0)$$

$${}_2 p = {}^T(\frac{1}{4} \quad 0 \quad \frac{1}{2} \quad 0 \quad \frac{1}{4})$$

$${}_3 p = {}^T(\frac{1}{4} \quad \frac{1}{4} \quad 0 \quad \frac{1}{4} \quad \frac{1}{4})$$

$${}_4 p = {}^T(\frac{3}{8} \quad 0 \quad \frac{1}{4} \quad 0 \quad \frac{3}{8})$$

$${}_5 p = {}^T(\frac{3}{8} \quad \frac{1}{8} \quad 0 \quad \frac{1}{8} \quad \frac{3}{8})$$

7 . 問題 6 と同様にして、

$${}_1 p = {}^T(\frac{1}{2} \quad 0 \quad \frac{1}{2} \quad 0 \quad 0)$$

$${}_2 p = {}^T(\frac{1}{2} \quad \frac{1}{4} \quad 0 \quad \frac{1}{4} \quad 0)$$

$${}_3 p = {}^T(\frac{5}{8} \quad 0 \quad \frac{1}{4} \quad 0 \quad \frac{1}{8})$$

$${}_4 p = {}^T(\frac{5}{8} \quad \frac{1}{8} \quad 0 \quad \frac{1}{8} \quad \frac{1}{8})$$

$${}_5 p = {}^T (11/16 \quad 0 \quad 1/8 \quad 0 \quad 3/16)$$

8 . 以下のとおり

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ b & 0 & a & 0 & 0 \\ 0 & b & 0 & a & 0 \\ 0 & 0 & b & 0 & a \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

9 . P^∞ の定義より、 $P^\infty = P \cdot P^\infty$ および

$$p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = P^\infty \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ が成立する。よって、}$$

$$p = P^\infty \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = P \cdot P^\infty \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = P \cdot p$$

10 . $p = P \cdot p$ を解いて、 $p = (0 \quad \frac{a^3}{1-2ab} \quad \frac{a^2}{1-2ab} \quad \frac{a(1-ab)}{1-2ab} \quad 1)$

11 . (略) 損保数理のテキストを参照せよ (7 . 7 練習問題の問 4)

12 . 損保数理のテキストを参照のこと (7 . 7 練習問題の問 2)

クレームが 0 件の場合には、 t は、損保数理テキスト 7 . 7 練習問題の問 2 における T_1 に等しいので、題意は証明される。また、ポアソン過程の定常性を考えれば、題意が成立することは直感的に理解できるであろう。

3.8 参考文献

- ・ 伊藤清[1991], 『確率論』, 岩波書店
- ・ 田畑吉雄[2001], 『金融工学入門』, エコノミスト社
- ・ テキスト委員会[2000], 『損保数理』, 日本アクチュアリー会
- ・ 日本債券信用銀行 商品開発グループ[1996],
『EXCELでわかるハル・ホワイト・モデル』, 金融財政事情研究会
- ・ 野口悠紀雄・藤井真理子[2000], 『金融工学』, ダイアモンド社
- ・ 福島正俊[1998], 『数学シリーズ 確率論』, 裳華房
- ・ 蓑谷千鳳彦[2000], 『よくわかるブラック・ショールズモデル』, 東洋経済新報社
- ・ 宮原孝夫[2003], 『株価モデルとレヴィ過程』, 朝倉書店
- ・ S . N . ネフツィ[1999], 『ファイナンスへの数学』, 朝倉書店
- ・ D . ラムベルトン・B . ラペール[2000], 『ファイナンスへの確率解析』, 朝倉書店
- ・ Gihman,I.I. and Skorohod,A.V.[1980],
Introduction a la Theorie des Processus Aleatoires , Mir

第4章 シミュレーション

4.1 乱数の生成法

4.1.1 はじめに

シミュレーションとは、現象を「真似ること」とか「模擬すること」と翻訳することができるが、研究対象となる現象をそのまま真似て実験すること全般を指すようである。シミュレーションで使用する手法に、モンテカルロ法として知られている手法がある。モンテカルロ法は、確率的な現象を前提とし、数学的な問題を解析的に解く代わりに、または、解析的には解けないので、大量に生成した乱数を利用して、答えを直接求めようとするものである。シミュレーションはおもにオペレーション・リサーチの分野で応用されている。モンテカルロ法は、多くの乱数を必要とするので、電子計算機の発展と共に可能となった技術である。

金融工学の分野では、近年、モンテカルロ法を応用して、金融商品の価格を計算する研究が盛んである。金融商品の価格を解析的に解くことができない場合であっても、モンテカルロ法によれば、価格を求めることができることがある。

金融商品の価格の計算にあつては、次のような積分計算を求められることが多い。すなわち密度関数 f に対して g なる関数があり

$$E[g(X)] = \int g(x)f(x)dx$$

を計算することである。この積分を解析的に求めることができれば問題ないのだが、解析的には求まらない場合、モンテカルロ法を使用することができる。まず、 f に従う独立な確率変数 X_1, \dots, X_n を生成し、 $E[g(X)]$ を

$$\frac{g(X_1) + \dots + g(X_n)}{n}$$

で近似する。大数の法則によれば、

$$E[g(X)] = \lim_{n \rightarrow \infty} \frac{g(X_1) + \dots + g(X_n)}{n}$$

となる。 n をどのぐらい大きくすれば、十分有効な近似値がもとまるか、どのようにしてある密度関数に従う確率変数を生成できるか、ということが課題となる。また、多くの場合、 n は数千から数万のオーダーであり、いかに計算プロセスを効率化することも課題である。特定の密度関数に従う確率変数を、一様分布に従う確率変数から、直接的または間接的に生成することができる。以下では、一様分布に従う確率変数を生成することができたとして、これを使用して特定の密度関数に

従う確率変数を生成する方法を述べる。計算時間の効率をはかるため、個々の分布に対して様々な工夫を施した算法(ロジック)が提案されている。本稿ではそのうちの代表的なものを紹介する。

生成方法の効率は実装されるシステムに依存する。また、基本となる一様乱数の生成方法に問題があると、これらを利用して求められる一般の確率変数にも問題が生じる。最終的には生成する乱数の検証を行った上で使用することが必要となる¹。

¹ 一様乱数の生成方法とその検証、および、これから求められる一般の確率変数の検証に関しては、後述する。参考文献としては[1]、[2]、[3]、[6]、[7]および[8]があげられる。

4.1.2 一般の確率変数を生成する一般的な技法

以下では(0,1)上の一様分布に従う確率変数をもとにして、任意の統計分布に従う確率変数を生成する方法を述べる²。

4.1.2.1 逆関数法

分布関数が F である確率分布に従う確率変数を求めるには、 U を(0,1)上の一様分布からの確率変数とし、その逆関数値 $F^{-1}(U)$ を求めれば、これが分布関数 F からの確率変数となっている。一般的な連続型の分布関数は単調増加関数であるので、その逆関数は一意に求まる。分布関数 F の逆関数が計算可能であるときに有効な手法である。

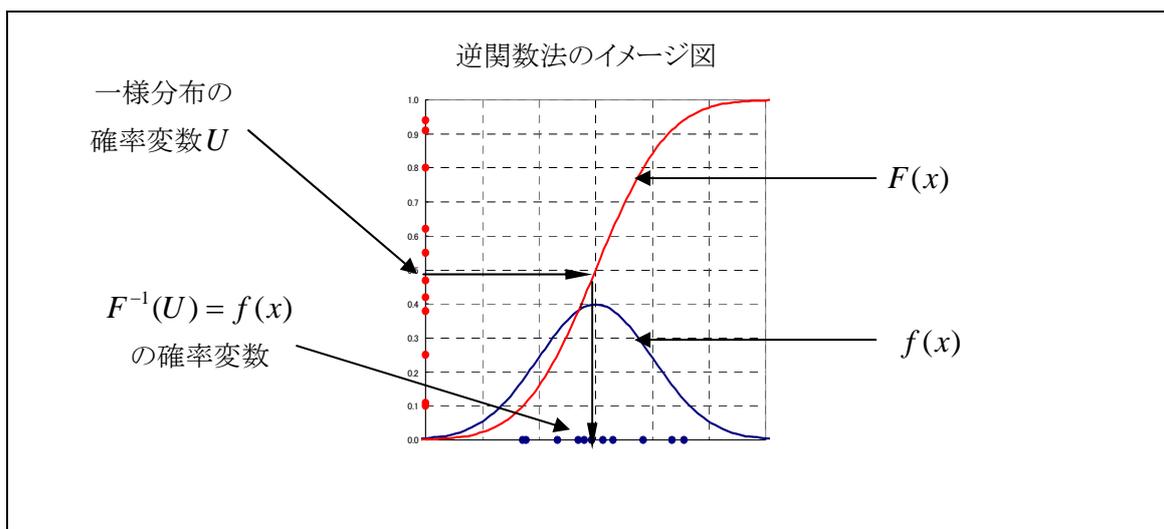
連続型の分布関数が単調増加関数でない場合、および、離散型の分布関数の場合を含め、一般的に、

$$F^{-1}(u) = \inf \{x \mid F(x) \geq u\}$$

と定義することができる。

4.1.2.1.1 逆関数法のイメージ

一様分布の確率変数は(0,1)上に一様に生成されている。これを縦線上の●で表現した。これら



² ここでは代表的な「逆関数法」、「棄却法」および「合成法」を説明するが、他に「奇偶法」、「一様乱数の比率を用いる方法」、「二者択一法(別名法またはエリマス法ともいう)」等がある。参考文献[1]、[2]、[4]、[6]および[7]を参照されたい。

の F^{-1} による写像を横軸上の●で表現した。密度関数 f の密度の高い部分(高い関数値となる横軸部分)では、分布関数 F の傾きが急であるため、多くのものが写像されてきている。 f に従う確率変数としてこれらを採用する。

分布関数が連続型の場合、次の命題が成立する。

4.1.2.1.2 命題:逆関数法(連続型の場合)

U を $(0,1)$ 上の一様分布からの確率変数とする。連続型の分布関数 F が単調増加関数であるとき、

$$X = F^{-1}(U) = \{x \mid F(x) = U\}$$

によって確率変数 X を定義すれば、確率変数 X は分布関数 F を持つ。

証明

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{F^{-1}(U) \leq a\} \\ &= P\{U \leq F(a)\} \\ &= F(a) \end{aligned}$$

3 番目の等号は、 $F(x)$ が単調増加関数であり、 $U \leq F(a)$ である場合に限り $F^{-1}(U) \leq a$ となることから成立する。◆

4.1.2.1.3 例題:指数分布の確率変数

パラメータ1の指数分布の分布関数は $F(x) = 1 - e^{-x}$ であり、 $u = 1 - e^{-x}$ とおくと $x = -\log(1-u)$ であるから、その逆関数は、

$$F^{-1}(u) = -\log(1-u)$$

である。ここに \log は自然対数とする(以下同じ)。例えば下表のように $(0,1)$ 上の一様分布に従う確率変数 U として 0.37、0.16、0.45 等が与えられたとすると、パラメータ1の指数分布として、0.46、0.17、0.60 等が得られる。

試行	1	2	3	4	5	6	7
U	0.37	0.16	0.45	0.38	0.47	0.83	0.11
$1-U$	0.63	0.84	0.55	0.62	0.53	0.17	0.89
$-\log(1-U)$	0.46	0.17	0.60	0.48	0.63	1.77	0.12

指数分布に従う確率変数の生成方法の詳細は、後述する。◆

4.1.2.1.4 逆関数法(離散型の場合)

分布関数が離散型の場合でも、同様に、まず累積確率を求め、連続型の分布における逆関数法を応用する。例えば、確率 P_1, \dots, P_n で値 x_1, \dots, x_n をとる離散型確率変数 X 、すなわち

$$P\{X = x_j\} = P_j, \quad j = 1, \dots, n, \quad \sum_{j=1}^n P_j = 1$$

である確率変数 X を生成するには、 U を $(0,1)$ 上の一様分布に従う確率変数とし、

$$X = \begin{cases} x_1, U \leq P_1 \text{ のとき} \\ x_2, P_1 < U \leq P_1 + P_2 \text{ のとき} \\ \vdots \\ x_n, \sum_{i=1}^{n-1} P_i < U \leq \sum_{i=1}^n P_i \text{ のとき} \end{cases}$$

とおくと、

$$P\{X = x_j\} = P\left\{\sum_{i=1}^{j-1} P_i < U \leq \sum_{i=1}^j P_i\right\} = P_j$$

なので、 X は求める分布を持っていることになる。

4.1.2.1.5 例題: 離散型

X は 1 から 10 までの整数をとる確率変数とし、 X の値をシミュレーションしたい。その確率および累積確率は以下のとおり。

j	1	2	3	4	5	6	7	8	9	10
x_j	1	2	3	4	5	6	7	8	9	10
P_j	0.11	0.12	0.09	0.08	0.12	0.10	0.09	0.09	0.10	0.10
累積	0.11	0.23	0.32	0.40	0.52	0.62	0.71	0.80	0.90	1.00

例えば下表のように $(0,1)$ 上の一様分布に従う確率変数 U として 0.37、0.16、0.45 等が与えられたとすると、 X として 4、2、5 等が得られる。

試行	1	2	3	4	5	6	7
U	0.37	0.16	0.45	0.38	0.47	0.83	0.11
X	4	2	5	4	5	9	1



4.1.2.2 棄却法

密度関数が g である分布の確率変数を生成する手段を持っているとする。一方、密度関数が f である分布の確率変数を生成したいとする。この際、 f/g が有界であるならば、すなわち

$$\text{適当な定数 } c \text{ に対して、} f(y)/g(y) \leq c \text{ for all } y$$

であるならば、次の手順により g を用いて f に従う確率変数 X を求めることができる。

- 手順 1: 密度関数が g である確率変数 Y を生成する。
- 手順 2: $(0,1)$ 上の一様分布に従う確率変数 U を生成する。
- 手順 3: $U \leq f(Y)/\{cg(Y)\}$ ならば、 $X = Y$ とする (X として Y を採用する)。そうでないならば手順 1 にもどる。

手順 3 で X の値を、確率 $P\{U \leq f(Y)/(cg(Y))\} = 1/c$ で得ることになるので(次の命題参照)、上記の手順の反復の回数は定数 c の取り方に依存する。 $1/c$ を大きくすれば手順数は減少する(手順 3 で手順 1 に戻る場合が少なくなる)。つまり最小の c を選択すれば計算手順数は効率化できる。またひとつの標本値を得るために必要な手順数は、平均 c の first success distribution となる。

ここに、パラメータ p (平均 $1/p$) の first success distribution (以下 FSD) に従う確率変数 X は以下の性質をもつ。

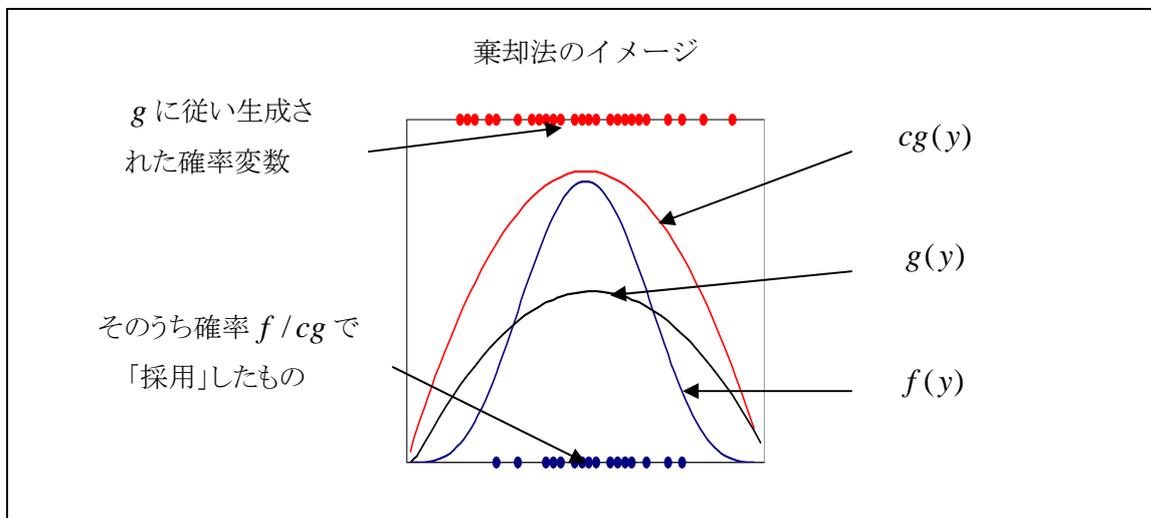
$$P\{X = j\} = p(1-p)^{j-1}, \quad j = 1, 2, 3, \dots$$

$$P\{X < j\} = \sum_{i=1}^{j-1} P\{X = i\} = 1 - P\{X > j-1\} = 1 - (1-p)^{j-1}$$

$$\text{平均} = \frac{1}{p}, \quad \text{分散} = \frac{1-p}{p^2}$$

FSD は成功確率 p の試行を繰り返したとき、初めての成功事象が発生するまでの試行回数(成功したときの試行を含む)の分布である。これに対しパラメータ p (平均 $(1-p)/p$) の幾何分布は、成功確率 p の試行を繰り返したとき、初めての成功事象が発生するまでの失敗の試行回数(成功したときの試行を含まない)の分布である。なお「4.1.4.1 幾何分布に従う確率変数」も参照されたい。

4.1.2.2.1 棄却法のイメージ



棄却法の手順をイメージ図を使って説明する。まず手順1で g に従う確率変数を生成する。この確率変数は、密度関数 g の値の大ききなところ、つまり密度の高いところでは多く生成される。これを図では上の線上の●で表現した。手順2と3では、これらから確率 $f/(cg)$ で選択し、選ばれたものを密度関数 f の確率変数として採用する。採用されたものが下の線上の●である。

4.1.2.2.2 命題:棄却法

棄却法によって生成された確率変数 X は、密度関数 f を持っている。

証明

得られた値を X とするとき、

$$\begin{aligned}
 P\{X \leq x\} &= P\{Y \leq x | U \leq f(Y)/(cg(Y))\} \\
 &= \frac{P\{Y \leq x, U \leq f(Y)/(cg(Y))\}}{K} \\
 &= \frac{\int P\{Y \leq x, U \leq f(Y)/(cg(Y)) | Y = y\} g(y) dy}{K} \\
 &= \frac{\int_{-\infty}^x (f(y)/(cg(y))) g(y) dy}{K} \\
 &= \frac{\int_{-\infty}^x f(y) dy}{Kc}
 \end{aligned}$$

である。ここに、 $K = P\{U \leq f(Y)/(cg(Y))\}$ である。 $x \rightarrow \infty$ とすることにより、 $P\{X \leq x\} \rightarrow 1$ であるから

$$1 = \int_{-\infty}^{+\infty} f(y)dy / Kc = 1/(Kc)$$

より、 $K = 1/c$ であることがわかる。よって $P\{X \leq x\} = \int_{-\infty}^x f(y)dy$ となる。これは、 X が密度関数 f に従うことを示している。◆

4.1.2.2.3 例題:連続型の場合の棄却法

棄却法を使用し次の密度関数をもつ確率変数 X を生成する。

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1$$

この確率変数(パラメータ(2,4)のベータ分布)は、区間(0,1)上で定義されているので、

$$g(x) = 1, \quad 0 < x < 1$$

を用いた棄却法を考える。 $g(x)$ は(0,1)上の一様分布の確率密度である。 $f(x)/g(x) \leq c$ となる定数 c を求めるため、まず

$$\frac{f(x)}{g(x)} = 20x(1-x)^3$$

の最大値を求める。この微分、

$$\frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] = 20 \left[(1-x)^3 - 3x(1-x)^2 \right]$$

を0とおくことで、 x が1/4のとき f/g が最大値をとることがわかる。すなわち

$$\frac{f(x)}{g(x)} \leq 20 \left(\frac{1}{4} \right) \left(\frac{3}{4} \right)^3 = \frac{135}{64} \equiv c$$

よって、

$$\frac{f(x)}{cg(x)} = \frac{256}{27} x(1-x)^3$$

となる。まとめると、この場合の棄却法による手順は以下のとおりとなる。

手順 1: 独立な $(0,1)$ 上の一様分布に従う確率変数 U_1 および U_2 を生成する(密度関数 g に従う確率変数を生成する)。

手順 2: $U_2 \leq 256/27 \times U_1(1-U_1)^3$ のとき $X = U_1$ とする。そうでなければ手順 1 に戻る。

手順 2 を通過するまでの回数の平均値は $c = 135/64 \approx 2.11$ である。すなわち、平均 2.11 回の試行で 1 個の X が求まる。例えば U_1, U_2 が以下の表のとおり得られたとすると、生成される X は次の通りである。この例の場合、7 回の試行から 4 個の X が求まっている。

試行	1	2	3	4	5	6	7
U_1	0.37	0.16	0.45	0.38	0.47	0.83	0.11
U_2	0.19	0.83	0.80	0.05	0.70	0.55	0.58
右辺	0.877202	0.899154	0.709867	0.858687	0.66344	0.038663	0.735257
$U_2 < \text{右辺}$	はい	はい	いいえ	はい	いいえ	いいえ	はい
X	0.37	0.16	-	0.38	-	-	0.11

ここに右辺 = $256/27 \times U_1(1-U_1)^3$ である。◆

4.1.2.2.4 例題: 離散型の場合の棄却法

先の例題 4.1.2.1.5 と同様、 X は 1 から 10 までの整数をとる確率変数とし、 X の値をシミュレーションしたい。 X のとる値とその確率 P_j は以下のとおり。

j	1	2	3	4	5	6	7	8	9	10
x_j	1	2	3	4	5	6	7	8	9	10
P_j	0.11	0.12	0.09	0.08	0.12	0.10	0.09	0.09	0.10	0.10

一つの方法は累積確率をもとめその逆関数を利用することである(例題 4.1.2.1.5 の方法)。別の方法は、次の分布関数を設定し、棄却法によるものである。すなわち、

Y	1	2	3	4	5	6	7	8	9	10
Q_y	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10

とおくと、 $c = \max(P_j / Q_y) = 1.2$ である。このとき、

手順 1 [1,10] 上の一様分布から Y を生成する(Y は 1 から 10 までの整数値。生成方法

は後述する)。

手順2 (0,1) 上の一様分布から U を生成する。

手順3 $U < P_y / (1.2 \times Q_y)$ ならば、 $X = Y$ とする (X の値として Y を使う)、そうでないならば手順1にもどる。

という手順で X の値をシミュレーションできる。 $c = 1.2$ であるので、平均 1.2 回の試行で 1 個の X が求まる。例えば Y 、 U が以下の表の様に与えられたとき、生成される X は次の通りである。この例の場合、7 回の試行から 7 個の X が求まっている。

試行	1	2	3	4	5	6	7
Y	4	2	5	4	5	9	2
U	0.19	0.83	0.8	0.05	0.7	0.55	0.58
P_y	0.08	0.12	0.12	0.08	0.12	0.10	0.12
Q_y	0.10	0.10	0.10	0.10	0.10	0.10	0.10
右辺	0.67	1.00	1.00	0.67	1.00	0.83	1.00
$U <$ 右辺	はい						
X	4	2	5	4	5	9	2

ここに右辺 = $P_y / (1.2 \times Q_y)$ である。◆

4.1.2.3 合成法

密度関数が $g_i, i = 1, \dots, n$ である分布があるとし、これら各々の確率変数を生成する手段をもっているとする。一方、密度関数 f に従う確率変数を求めたいが、 f が h_i の重み付けの平均で表現されているとする。すなわち、

$$f(x) = h_1 g_1(x) + \dots + h_n g_n(x)$$

$$\sum_{i=1}^n h_i = 1, h_i \geq 0, i = 1, \dots, n$$

であるとき、 h_i の確率で番号 i を選択し、選択された i に対して g_i に従う確率変数 X を生成すれば、 X は分布 f に従った確率変数である。

連続分布の場合も同様に、

1. $f(x) = \int g(x; \theta) h(\theta) d\theta$
2. $h(\theta)$ はパラメータ θ の重みを表現する密度関数であり、これに従う確率変数を生成する手段をもっている。
3. $g(x; \theta)$ は θ をパラメータとする密度関数であり、これに従う確率変数を生成する手段をもっている。

とき、始めに $h(\theta)$ に従う乱数 θ を求め、その θ をパラメータとする $g(x; \theta)$ に従う確率変数 X を求

めれば、これが分布 f に従う確率変数である。

離散型と連続型が混合している場合も同様である。一般に、累積密度関数 F が、重み h_i と確率変数の生成方法が既知の分布 F_i によって、

$$F(x) = \sum h_i F_i(x), \quad \sum h_i = 1, \quad h_i \geq 0$$

とあらわされている場合に応用できる。

4.1.2.3.1 例題: 合成法

分布関数、

$$F(X) = (1/3)(X + X^2 + X^3), 0 \leq X \leq 1$$

に従う確率変数を生成する場合、分布関数の逆関数 $F^{-1}(U)$ は計算できないが、 $F_i(X) = X^i$ 、 $h_i = 1/3, i = 1, 2, 3$ とすれば $F_i^{-1}(U)$ は計算できるので、 $[0, 1]$ 上の一様分布に従う確率変数 U_1 、 U_2 を用意し、

$$\begin{aligned} 0 \leq U_1 \leq 1/3 \text{ ならば } X &= F_1^{-1}(U_2) \\ 1/3 < U_1 \leq 2/3 \text{ ならば } X &= F_2^{-1}(U_2) \\ 2/3 < U_1 \leq 1 \text{ ならば } X &= F_3^{-1}(U_2) \end{aligned}$$

として X が求まる。◆

4.1.3 連続型分布に従う確率変数の生成

以下では特定の連続型分布に従う確率変数を生成する代表的な方法を述べる。前述の一般的方法に加えて、数値処理の効率を考慮して、様々な方法が考案されている。

4.1.3.1 一様分布に従う確率変数

領域 $[a, b]$ 上の一様分布は以下の性質を持つ。

$$\text{密度関数: } f(x) = \frac{1}{b-a}, a \leq x \leq b$$

$$\text{分布関数: } F(x) = \frac{x-a}{b-a}, a \leq x \leq b$$

$$\text{平均} = \frac{a+b}{2}, \quad \text{分散} = \frac{(b-a)^2}{12}$$

4.1.3.1.1 逆関数法による方法

$[a, b]$ 上の一様分布に従う確率変数 X は、逆関数法を用いて、 $[0, 1]$ 上の一様分布に従う確率変数 U に対して、

$$X = (b-a)U + a$$

から求めることができる。

$(0, 1]$ 、 $[0, 1)$ および $(0, 1)$ 上の一様分布に従う確率変数は、 $[0, 1]$ 上の一様分布を生成した際、0 または 1、またはその両方を除くことによって求めることができる。

4.1.3.2 整数値をとる一様分布

$1, \dots, k$ (k は正の整数) を均等な確率でとる確率変数は、 U を $[0, 1)$ 上の一様分布に従う確率変数として

$$\text{int}(kU) + 1$$

によって求めることができる。

$0, \dots, k$ (k は正の整数) を均等な確率でとる確率変数は、 U を $[0, 1)$ 上の一様分布に従う確率変数として

$$\text{int}((k+1)U)$$

によって求めることができる。ここに、 $\text{int}(x)$ は x を超えない最大の整数とする。

4.1.3.3 指数分布に従う確率変数³

パラメータ λ の指数分布は以下の性質を持つ。

$$\text{密度関数: } f(x) = \lambda e^{-\lambda x}, 0 \leq x$$

$$\text{分布関数: } F(x) = 1 - e^{-\lambda x}, 0 \leq x$$

$$\text{平均} = \frac{1}{\lambda}, \quad \text{分散} = \frac{1}{\lambda^2}$$

4.1.3.3.1 逆関数法による方法

逆関数の例題でパラメータ 1 の指数分布に従う確率変数を求める方法を示した。すなわち、 U が $(0,1)$ 上の一様分布に従う確率変数のとき、 $F^{-1}(U) = -\log(1-U)$ はパラメータ 1 の指数分布に従う確率変数となる。 X がパラメータ 1 の指数分布に従う確率変数の場合、 $(1/\lambda)X$ がパラメータ λ (平均 $1/\lambda$) の指数分布に従う確率変数なので、パラメータ λ (平均 $1/\lambda$) の指数分布に従う確率変数 X は、

$$X = -(1/\lambda) \log(1-U)$$

で求められる。

さらに、 $1-U$ も同じく $(0,1)$ 上の一様分布に従う確率変数なので、 $-\log U$ もパラメータ 1 の指数分布に従う確率変数である。よって、

$$X = -(1/\lambda) \log U$$

もパラメータ λ (平均 $1/\lambda$) の指数分布に従う確率変数を与える。

4.1.3.4 正規分布に従う確率変数⁴

パラメータ (μ, σ^2) の正規分布は以下の性質を持つ。

$$\text{密度関数: } f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

分布関数: 一般式は求まらない。

$$\text{平均} = \mu, \quad \text{分散} = \sigma^2$$

³ その他、奇偶法、不定個数の一様乱数の最小値を利用する方法がある(参考文献[1][2])。

⁴ その他、奇偶法、一様乱数の比を用いる方法(参考文献[1][2])、合成法(参考文献[6])等が考案されている。

その他の性質

1. 正規分布の再生性

独立な確率変数 X_1, X_2, \dots, X_n が各々パラメータ (μ_i, σ_i^2) の正規分布に従うならば、 $X_1 + X_2 + \dots + X_n$ はパラメータ $(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$ の正規分布に従う。

4.1.3.4.1 中心極限定理を用いた方法

一般に中心極限定理により、「同一の分布に従う n 個の独立な確率変数の算術平均は、 n が大なるとき、正規分布で近似できる」。よって $[0,1]$ 上の一様分布に従う n 個の独立な確率変数 $U_i, i=1, \dots, n$ の算術平均、

$$\bar{U} = \frac{1}{n}(U_1 + \dots + U_n)$$

は、 n が大なるとき、正規分布に近似した分布を持つ。 $[0,1]$ 上の一様分布の平均が $1/2$ 、分散が $1/12$ であるので、12個の独立な $[0,1]$ 上の一様乱数を生成し、

$$X = U_1 + \dots + U_{12} - 6$$

とおくと、中心極限定理から、 X は平均0、分散1の正規分布(標準正規分布)に従う確率変数となる。一般に平均 μ 、分散 σ の正規分布に従う確率変数 Z は上記の X に対して

$$Z = \mu + \sigma X$$

として求められる。

この方法の特徴は以下のとおり。

実装が簡単である。

一様分布に従う確率変数が高速に生成できる場合は、ボックス・マーラーの変換式による方法(後述する)のように対数計算、三角関数計算を行わない分、効率的である(計算が速い)。

標準正規分布は理論的には $(-\infty, +\infty)$ の範囲の値をとるのに対して、この方法は $[-6, +6]$ の範囲の値しかとらないので、分布の裾が薄くなる傾向にあると考えられる。ただし、標準正規分布の値が、例えば、+6以上となるのは 10^{-9} のオーダーであるので、実務の必要に応じて利用すればよい。

4.1.3.4.2 棄却法による方法

標準正規分布(平均0、分散1)に従う確率変数 Z を、棄却法によって生成する。まず Z の絶対値の密度関数は

$$f(x) = \frac{2}{\sqrt{2\pi}} \exp(-x^2/2), \quad 0 < x < \infty$$

である。gとしてパラメータ1の指数分布

$$g(x) = e^{-x}, \quad 0 < x < \infty$$

をとると、

$$\frac{f(x)}{g(x)} = \sqrt{2e/\pi} \exp\left\{-\frac{(x-1)^2}{2}\right\} \leq \sqrt{2e/\pi} = c$$

$$\frac{f(x)}{cg(x)} = \exp\left\{-\frac{(x-1)^2}{2}\right\}$$

であるので、密度関数 f に従う確率変数 X は、棄却法により、次の手順で得られる。

- 手順1 独立な確率変数 Y および U を生成する。ここに Y はパラメータ1の指数分布に従う確率変数、 U は $(0,1)$ 上の一様分布に従う確率変数とする。
- 手順2 $U \leq \exp\{-(Y-1)^2/2\}$ のとき、または、同じく

$$-\log U \geq (Y-1)^2/2$$

のとき、 $X = Y$ とする。そうでなければ手順1に戻る。

$c = \sqrt{2e/\pi} \approx 1.32$ であるので、上記の手順2を通過するまでの回数は、平均1.32の幾何分布に従う。

一旦、密度関数 f を持っている確率変数 X を生成できたなら、 Z を X か $-X$ のいずれかを等しい確率でとる確率変数とおくことで、標準正規分布に従う確率変数 Z を生成することができる。平均 μ 、分散 σ^2 の正規分布に従う確率変数は、 $\mu + \sigma Z$ で求まる。

4.1.3.4.3 ボックス・マーラーの変換式による方法(1)

$(0,1)$ 上の独立な一様分布に従う確率変数 U_1, U_2 から、

$$X = (-2 \log U_1)^{1/2} \cos(2\pi U_2),$$

$$Y = (-2 \log U_1)^{1/2} \sin(2\pi U_2)$$

と変換すると、 X, Y は独立な2変数標準正規分布に従う。すなわち2つの独立な一様分布に従う確率変数から独立な標準正規分布に従う確率変数が2つ同時に求まる。これをボックス・マーラー(Box-Muller)の変換式という。以下これを求める。

X および Y を独立な標準正規分布に従う確率変数とすると、結合密度関数は

$$f(x, y) = \frac{1}{2\pi} \exp\left\{-\frac{(x^2 + y^2)}{2}\right\}, \quad -\infty < x < \infty, -\infty < y < \infty$$

となる。今、点 (X, Y) の極座標を考え、

$$\begin{aligned} R^2 &= X^2 + Y^2 \\ \Theta &= \tan^{-1} Y/X \end{aligned}$$

とおく。 R^2 と Θ の結合密度を求めるため、変換

$$d = x^2 + y^2, \quad \theta = \tan^{-1} y/x$$

を考えると、この変換のヤコビアン行列式が

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial d}{\partial x} & \frac{\partial d}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} \\ &= \begin{vmatrix} 2x & 2y \\ \frac{1}{1+y^2/x^2} \left(\frac{-y}{x^2}\right) & \frac{1}{1+y^2/x^2} \left(\frac{1}{x}\right) \end{vmatrix} = 2 \begin{vmatrix} x & y \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix} = 2 \end{aligned}$$

であるところから、 R^2 と Θ の結合密度は、

$$\begin{aligned} f_{R^2, \Theta}(d, \theta) &= \frac{1}{2\pi} e^{-d/2} \frac{1}{2} \\ &= \frac{1}{2} e^{-d/2} \frac{1}{2\pi}, \quad 0 < d < \infty, 0 < \theta < 2\pi \end{aligned}$$

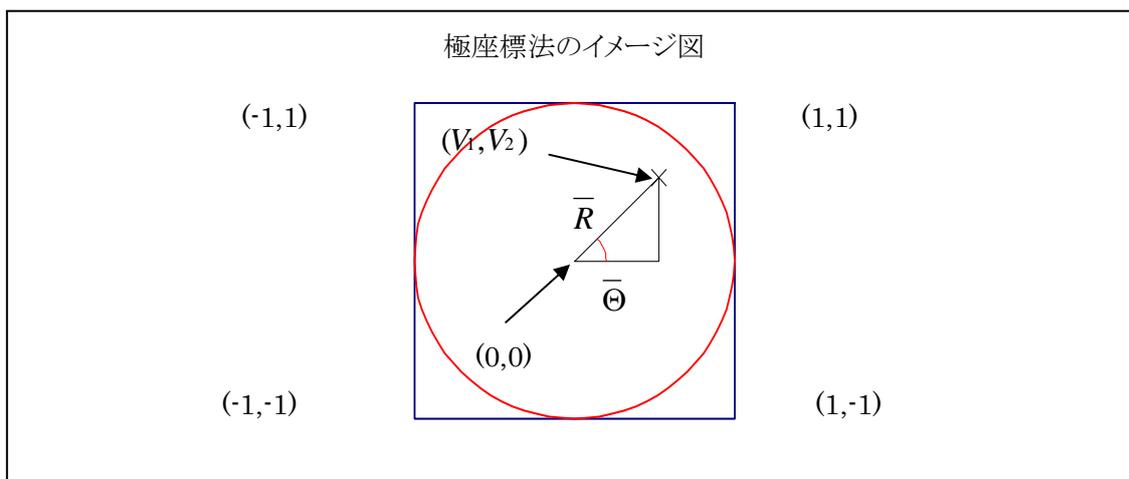
となる。式の形から R^2 と Θ は独立であると結論できる。 R^2 はパラメータ $1/2$ の指数分布、 Θ は $(0, 2\pi)$ 上の一様分布に従う確率変数である。

今度は逆に、極座標を直交座標に変換する。上記から、パラメータ $1/2$ の指数分布に従う確率変数 W (W は R^2 の役割を果たす) および W と独立な $(0, 2\pi)$ 上の一様分布に従う確率変数 V (V は、 Θ の役割を果たす) から始めれば、 $X = \sqrt{W} \cos V$ 、 $Y = \sqrt{W} \sin V$ は独立な標準正規分布に従う確率変数となる。従って U_1 と U_2 が $(0, 1)$ 上の独立した一様分布に従う確率変数のとき、

$$\begin{aligned} X &= (-2 \log U_1)^{1/2} \cos(2\pi U_2), \\ Y &= (-2 \log U_1)^{1/2} \sin(2\pi U_2) \end{aligned}$$

は独立な標準正規分布に従う確率変数となる。

$X^2 + Y^2$ はパラメータ $1/2$ の指数分布に従う確率変数だが、一方、カイ 2 乗分布の定義によって、 $X^2 + Y^2$ は自由度 2 のカイ 2 乗分布に従う確率変数でもある。これらの 2 つの分布は同一である。



4.1.3.4.4 ボックス・マーラーの変換式による方法(極座標法)

上記の方法では対数と三角関数の計算が伴うため、計算時間が非効率であり、かつ、十分な計算結果の精度を得られないことがあるが、少なくとも三角関数の計算を回避する方法がある。

まず U が $(0,1)$ 上の一様分布に従う確率変数であるとき、 $2U$ が $(0,2)$ 上の一様分布に従う確率変数、 $2U - 1$ は $(-1,1)$ 上の一様分布に従う確率変数であることに注意する。よって、 $(0,1)$ 上の一様分布に従う確率変数 U_1 および U_2 を生成し、

$$V_1 = 2U_1 - 1$$

$$V_2 = 2U_2 - 1$$

とおけば、 (V_1, V_2) は $(0,0)$ を中心とする面積 4 の正方形の区域上で一様に分布する。

今 $(0,0)$ を中心とする半径 1 の円に含まれるものを得るまで、引き続いて対 (V_1, V_2) を生成する、すなわち、生成された (V_1, V_2) のうち $V_1^2 + V_2^2 \leq 1$ を満たす (V_1, V_2) のみを採用すると、その対 (V_1, V_2) は一様に円内に分布し、この対の極座標を $\bar{R}, \bar{\Theta}$ とすると、 \bar{R} と $\bar{\Theta}$ は独立である。 \bar{R}^2 は $(0,1)$ 上の一様分布、 $\bar{\Theta}$ は $(0, 2\pi)$ 上の一様分布に従う確率変数である。

$$\sin \bar{\Theta} = V_2 / \bar{R} = \frac{V_2}{\sqrt{V_1^2 + V_2^2}},$$

$$\cos \bar{\Theta} = V_1 / \bar{R} = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$

であるので、別の乱数 U を生成し

$$X = (-2 \log U)^{1/2} V_1 / \bar{R},$$

$$Y = (-2 \log U)^{1/2} V_2 / \bar{R}$$

とすることにより、独立な標準正規分布に従う確率変数 X および Y を生成することができる。実際、 $V_1^2 + V_2^2 \leq 1$ という条件から \bar{R}^2 が $(0,1)$ 上の一様分布に従う確率変数であり、 $\bar{\Theta}$ と独立なので、新しい乱数 U を生成する代わりに、これを使用することができる。したがって

$$X = \left(-2 \log \bar{R}^2\right)^{1/2} V_1 / \bar{R} = \sqrt{\frac{-2 \log S}{S}} V_1,$$

$$Y = \left(-2 \log \bar{R}^2\right)^{1/2} V_2 / \bar{R} = \sqrt{\frac{-2 \log S}{S}} V_2$$

は独立な標準正規分布に従う確率変数である。ここに、

$$S = \bar{R}^2 = V_1^2 + V_2^2$$

である。まとめると、1 対の独立な標準正規分布 X および Y を生成する手順は、次のとおり。

手順 1: $(0,1)$ 上の一様分布に従う独立な確率変数 U_1 および U_2 を生成する。

手順 2: $V_1 = 2U_1 - 1, V_2 = 2U_2 - 1, S = V_1^2 + V_2^2$ とする。

手順 3: $S > 1$ のとき、手順 1 にもどる。

手順 4:

$$X = \sqrt{\frac{-2 \log S}{S}} V_1, \quad Y = \sqrt{\frac{-2 \log S}{S}} V_2$$

とおく。

上記は、極座標法と呼ばれる。正方形内の任意のポイントが円内に落ちる確率が、 $\pi/4$ (円の領域割る正方形の領域) であるので、極座標法では、手順 1 を平均 $4/\pi = 1.273$ 回反復することにより、独立な標準正規分布に従う確率変数を 2 つ得ることになる。

4.1.3.5 ガンマ分布に従う確率変数

パラメータ (α, β) のガンマ分布は次の性質を持つ。

$$\text{密度関数: } f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, 0 < x$$

$$\text{ここに、 } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

一般には分布関数は求まらないが、 α が正の整数のとき、

$$\text{分布関数: } F(x) = 1 - e^{-\beta x} \sum_{j=0}^{\alpha-1} \frac{(\beta x)^j}{j!}, 0 < x$$

$$\text{平均} = \alpha / \beta, \quad \text{分散} = \alpha / \beta^2$$

(α を形状パラメータ、 β をスケール・パラメータと称する。)

その他の性質

1. パラメータ λ の指数分布=パラメータ $(1, \lambda)$ のガンマ分布
2. 自由度 n のカイ2乗分布=パラメータ $(n/2, 1/2)$ のガンマ分布
3. ガンマ分布の再生性
独立な確率変数 X_1, X_2 がそれぞれパラメータ $(\alpha_1, \beta), (\alpha_2, \beta)$ のガンマ分布に従うならば、 $X_1 + X_2$ はパラメータ $(\alpha_1 + \alpha_2, \beta)$ のガンマ分布に従う。
4. X がパラメータ $(\alpha, 1)$ のガンマ分布に従うならば、 X / β はパラメータ (α, β) のガンマ分布に従う。
5. 独立な確率変数 X_1, X_2, \dots, X_n がそれぞれパラメータ λ の指数分布に従うならば $X_1 + X_2 + \dots + X_n$ はパラメータ (n, λ) のガンマ分布に従う。

4.1.3.5.1 指数分布の和による方法⁵

パラメータ α が整数 n である場合のガンマ分布に従う確率変数は、 n 個の独立なパラメータ λ の指数分布に従う確率変数の和が、パラメータ (n, λ) のガンマ分布に従う(上記5.)ことを利用して求めることができる。すなわち、 U_1, \dots, U_n が $(0, 1)$ 上の独立した一様分布に従う確率変数である場合、

$$X = -\frac{1}{\lambda} \sum_{i=1}^n \log U_i = -\frac{1}{\lambda} \log \left(\prod_{i=1}^n U_i \right)$$

はパラメータ (n, λ) のガンマ分布に従っている。右端の式は対数計算を一度のみ行うので、一般

⁵ n が大きな整数であるときは生成の効率が良くない。 α が整数でない場合の生成方法を含め、参考文献[1][2][6][7]を参照されたい。

的には計算が効率的である。

4.1.3.6 カイ2乗分布に従う確率変数

自由度 n のカイ2乗分布は次の性質を持つ。

$$\text{密度関数 : } f(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}$$

$$\text{平均} = n, \text{ 分散} = 2n$$

その他の性質

1. パラメータ $(n/2, 1/2)$ のガンマ分布 = 自由度 n のカイ2乗分布。

2. n 個の独立な標準正規分布の2乗の和 = 自由度 n のカイ2乗分布

3. カイ2乗分布の再生性

自由度 n のカイ2乗分布と自由度 m のカイ2乗分布の和は、自由度 $n+m$ のカイ2乗分布。

4.1.3.6.1 標準正規分布から求める方法

自由度 n のカイ2乗分布に従う確率変数 χ_n^2 は、 $Z_i, i=1, \dots, n$ を n 個の独立な標準正規分布に従う確率変数として、

$$\chi_n^2 = Z_1^2 + \dots + Z_n^2$$

で求められる。

4.1.3.6.2 指数分布から求める方法

上記の方法では、標準正規分布を生成するために多くの一様分布を生成しなければならない。生成する一様分布の数を節約する方法は以下のとおり。

まず、 $Z_1^2 + Z_2^2$ がパラメータ $1/2$ の指数分布であることから、 n が偶数の場合(例えば $n = 2k$)、 χ_{2k}^2 はパラメータ $(k, 1/2)$ のガンマ分布と等しくなる。

よって、 $-2\log(\prod_{i=1}^k U_i)$ は自由度 $2k$ のカイ2乗分布となる。

n が奇数の場合(例えば $n = 2k+1$)、標準正規分布に従う確率変数 Z を生成し、 χ_{2k}^2 に Z^2 を加えれば自由度 $2k+1$ のカイ2乗分布に従う確率変数を得ることができる。つまり、

$$\chi_n^2 = \begin{cases} -2\log\left(\prod_{i=1}^k U_i\right), n = 2k \\ Z^2 - 2\log\left(\prod_{i=1}^k U_i\right), n = 2k+1 \end{cases}$$

である。ここに Z と U_1, \dots, U_n は独立であり、 Z は標準正規分布、他は $(0,1)$ 上の一様分布に従う確率変数とする。結局、 k 個の指数分布の和と、1 個の標準正規分布から求めることができる。

4.1.3.7 ベータ分布に従う確率変数⁶

パラメータ (n, m) のベータ分布は以下の性質を持つ。

$$\text{密度関数} : f(x) = \frac{(n+m-1)!}{(n-1)!(m-1)!} x^{n-1} (1-x)^{m-1}, \quad 0 < x < 1$$

$$\text{平均} = \frac{n}{n+m}, \quad \text{分散} = \frac{n \times m}{(n+m)^2 (n+m+1)}$$

その他の性質

1. 独立な確率変数 X_1, X_2 がそれぞれパラメータ $(\alpha_1, \beta), (\alpha_2, \beta)$ のガンマ分布に従うならば、 $X_1 / (X_1 + X_2)$ はパラメータ (α_1, α_2) のベータ分布に従う。
2. X がパラメータ (n, m) のベータ分布に従うならば、 $1 - X$ はパラメータ (m, n) のベータ分布に従う。

4.1.3.7.1 ガンマ分布から求める方法

X, Y をそれぞれパラメータ $(n, 1), (m, 1)$ のガンマ分布に従う確率変数とすると、 $X / (X + Y)$ はパラメータ (n, m) のベータ分布に従う確率変数となる。よって $(0,1)$ 上の一様分布に従う確率変数 U_1, \dots, U_{n+m} を生成し、

$$X = \frac{-\log \prod_{i=1}^n U_i}{-\log \prod_{i=1}^n U_i - \log \prod_{i=n+1}^{n+m} U_i} = \frac{-\log \prod_{i=1}^n U_i}{-\log \prod_{i=1}^{m+n} U_i}$$

とおけば、 X はパラメータ (n, m) のベータ分布に従う確率変数となる。結局、指数分布の和の比を計算することで求めることができる。

⁶ ここではパラメータ n, m が正の整数の場合を紹介するにとどめる。一般のベータ分布に対して多くの方法が提案されている。参考文献[2][6][7]を参照されたい。

4.1.4 離散型分布に従う確率変数の生成

連続型分布の場合と同様、以下では代表的な方法を説明する。離散型にあっても効率性を重視して様々な方法が提案されている。

4.1.4.1 幾何分布に従う確率変数⁷

パラメータ p の幾何分布に従う確率変数 X は以下の性質をもつ。

$$\begin{aligned}P\{X = j\} &= p(1-p)^j, \quad j = 0, 1, 2, \dots \\P\{X < j\} &= \sum_{i=0}^{j-1} P\{X = i\} = 1 - P\{X > j-1\} = 1 - (1-p)^j \\ \text{平均} &= \frac{1-p}{p}, \quad \text{分散} = \frac{1-p}{p^2}\end{aligned}$$

4.1.4.1.1 逆関数法による方法

逆関数法を使用する。すなわち、 $(0,1)$ 上の一様分布に従う確率変数 U を生成して、 X を次のような j とおくことにより、幾何分布に従う確率変数を得ることができる。

$$\begin{aligned}X &= \{j : 1 - (1-p)^j \leq U < 1 - (1-p)^{j+1}\} \\ &= \{j : (1-p)^{j+1} < 1-U \leq (1-p)^j\}\end{aligned}$$

$1-U$ が U と同じ分布を持っているので、結局、

$$\begin{aligned}X &= \min \left\{ j : (1-p)^{j+1} < U \right\} = \min \left\{ j : j+1 > \frac{\log U}{\log(1-p)} \right\} \\ &= \text{int} \left(\frac{\log U}{\log(1-p)} \right)\end{aligned}$$

によって X をもとめることができる。

⁷ その他、ベルヌーイ試行による方法、二者択一法を利用する方法などがある。参考文献[1][2]を参照されたい。

4.1.4.2 二項分布に従う確率変数⁸

パラメータ (n, p) の二項分布は以下の性質をもつ。

$$P\{X = j\} = \binom{n}{j} p^j (1-p)^{n-j}, \quad j = 0, 1, \dots, n$$

$$P\{X < j\} = \sum_{i=0}^{j-1} \binom{n}{i} p^i (1-p)^{n-i}, \quad j = 0, 1, \dots, n$$

$$\text{平均} = np, \quad \text{分散} = np(1-p)$$

その他の性質

1. X_1, X_2, \dots, X_n が独立なパラメータ p のベルヌーイ試行ならば

$X_1 + X_2 + \dots + X_n$ はパラメータ (n, p) の二項分布に従う確率変数である。

2. 二項分布の再生性

X_1, X_2, \dots, X_m が独立なパラメータ (n_i, p) の二項分布に従う確率変数ならば

$X_1 + X_2 + \dots + X_m$ はパラメータ $(n_1 + n_2 + \dots + n_m, p)$ の二項分布に従う確率変数である。

4.1.4.2.1 ベルヌーイ試行の合計による方法

パラメータ (n, p) の二項分布に従う確率変数は、 n 個の独立したベルヌーイ試行の合計として表現することができる。すなわち、 U_1, \dots, U_n が独立した $(0, 1)$ 上の一様分布に従う確率変数で、

$$X_i = \begin{cases} 1, & \text{if } U_i < p \\ 0, & \text{上記以外} \end{cases}$$

とすると、 $X \equiv \sum_{i=1}^n X_i$ がパラメータ (n, p) の二項分布に従う確率変数となる。

4.1.4.3 負の二項分布に従う確率変数

パラメータ (r, p) の負の二項分布は以下の性質を持つ。

⁸ その他、幾何分布と指数分布を応用する方法(参考文献[2])、ベータ分布を利用する方法(参考文献[1][2])が考案されている。

$$P\{X = j\} = \binom{r+j-1}{j} p^r (1-p)^j, j = 0, 1, 2, \dots$$

$$P\{X < j\} = \sum_{i=0}^{j-1} \binom{r+i-1}{i} p^r (1-p)^i, j = 0, 1, 2, \dots$$

$$\text{平均} = r(1-p)/p, \quad \text{分散} = r(1-p)/p^2$$

その他の性質

1. X_1, X_2, \dots, X_r が独立なパラメータ p の幾何分布に従う確率変数ならば
 $X_1 + X_2 + \dots + X_r$ はパラメータ (r, p) の負の二項分布に従う確率変数である。
2. G がパラメータ $(r, 1)$ のガンマ分布に従う確率変数ならば
 パラメータ $(G(1-p)/p)$ のポワソン分布は、パラメータ (r, p) の負の二項分布である。
3. 負の二項分布の再生性
 X_1, X_2, \dots, X_n が独立なパラメータ (r_i, p) の負の二項分布に従う確率変数ならば
 $X_1 + X_2 + \dots + X_n$ はパラメータ $(\sum_{i=1}^n r_i, p)$ の負の二項分布に従う確率変数である。

4.1.4.3.1 幾何分布から求める方法

X_1, \dots, X_r を r 個の独立したパラメータ p の幾何分布に従う確率変数とすると、 $\sum_{i=1}^r X_i$ はパラメータ (r, p) の負の二項分布に従う。これを用いて負の二項分布に従う確率変数を得ることができる。

4.1.4.3.2 ガンマ分布とポワソン分布から求める方法

G をパラメータ $(r, 1)$ のガンマ分布に従う確率変数とすると、パラメータ $G(1-p)/p$ のポワソン分布は、パラメータ (r, p) の負の二項分布に従う。これを用いて、負の二項分布に従う確率変数を得ることができる。

4.1.4.4 ポワソン分布に従う確率変数⁹

パラメータ λ のポワソン分布は次の性質を持つ。

⁹ その他、ガンマ分布と二項分布を利用する方法などがある(参考文献[1])。

$$P\{X = x\} = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, \dots$$

$$P\{X < x\} = \sum_{y=0}^{x-1} \frac{e^{-\lambda} \lambda^y}{y!}$$

平均 = λ , 分散 = λ

その他の性質

1. ポワソン分布の再生性

X_1, X_2, \dots, X_m が独立でそれぞれパラメータ λ_i のポワソン分布に従う確率変数のとき、 $X_1 + X_2 + \dots + X_m$ はパラメータ $\lambda_1 + \dots + \lambda_m$ のポワソン分布に従う確率変数である。

4.1.4.4.1 指数分布から求める方法

平均 λ のポアソン分布に従う確率変数をシミュレートするために、 $(0,1)$ 上の独立した一様分布に従う確率変数 U_1, U_2, \dots を生成し、

$$N+1 = \min \left\{ n : \prod_{i=1}^n U_i < e^{-\lambda} \right\} = \min \left\{ n : \log \left(\prod_{i=1}^n U_i \right) < -\lambda \right\}$$

とすると、確率変数 N は求める分布を持っている。それは

$$N = \max \left\{ n : \sum_{i=1}^n -\log U_i < \lambda \right\}$$

であることからわかる。つまり、 $-\log U_i$ はパラメータ1の指数分布に従う確率変数であり、 $-\log U_i, i \geq 1$ をパラメータ1のポアソン過程の到着間隔時間と解釈すれば、 N が時刻 λ までの事象の数と等しいことがわかる。従って、 N は平均 λ のポアソン分布に従う確率変数である¹⁰。

¹⁰ 指数分布とポワソン分布の関係は、参考文献[4][6]を参照されたい。

4.1.5 練習問題

4.1.5.1 SOA(2000年5月コース3問題14:一部改題)

密度関数 $g(x)$ および定数 c を用いて、密度関数 $f(x)$ を有する確率変数を、棄却法で生成したい。

1. $f(x) = 12(x - 2x^2 + x^3), 0 < x < 1$
2. $g(x) = 1, 0 < x < 1$

定数 c を適当に定めることにより、 $f(x)$ に従う確率変数を求めるための繰り返し回数を最小にした。 c をもとめよ。

選択肢:(A)1.52 (B)1.78 (C)1.86 (D)2.05 (E)2.11

4.1.5.1.1 解答

f/g を微分することにより、 f/g は $x = 1/3$ で最高値 1.78 を取ることがわかる。

答え:(B)

4.1.5.2 SOA(2000年5月コース3問題32)

1. ある保険会社はある町の除雪コストをカバーしている。冬は4ヶ月間である。
2. 月々の免責額は10,000(毎月10,000を差し引いた除雪コストを支払う)
3. 保険会社はその町の月々の除雪コストは独立で、平均15,000、標準偏差2,000の正規分布に従っていると想定している。
4. 累積密度関数の逆関数を用いる方法で4ヶ月間の純支払額をシミュレートした。(小なる確率変数は、小額の純支払額に対応するとする)
5. $[0,1]$ 区間の一様分布として、次の確率変数を得た。

0.5398 0.1151 0.0013 0.7881

4ヶ月間の純支払額のシミュレーション結果を求めよ。

選択肢:(A)13,400 (B)14,400 (C)17,800 (D)20,000 (E)26,600

4.1.5.2.1 解答

$X = 15,000 + 2,000 \times Z$ 、 Z を標準正規分布とおき、 Z を逆関数法で求める(試験に添付されている標準正規分布のテーブルから参照する)。

U	$Z = F^{-1}(U)$	X	$X-10,000$	$Max(0, X - 10,000)$
0.5398	0.0999	15,200	5,200	5,200
0.1151	-1.1998	12,600	2,600	2,600
0.0013	-3.0115	8,977	-1,023	0
0.7881	0.7998	16,600	6,600	6,600
計				14,400

答え:(B)

4.1.5.3 SOA(2001年5月コース3問題8)

株価の変動をシミュレートした。

1. 価格は幾何ブラウン運動するとし、 $\mu = 0.01$ 、 $\sigma^2 = 0.0004$ とする。
2. 時間の単位は各ステップごとに1とする
3. 価格変動は、分布の逆関数法によるものとする。
4. $t = 0$ における価格は100とする。
5. $[0,1]$ 区間上の一様分布の確率変数が、0.1587、0.9332の順であったとする。

シミュレーションの結果、 $t = 1$ における価格が F 、 $t = 2$ における価格が G であったとする。

$G - F$ を求めよ。

選択肢:(A)1 (B)2 (C)3 (D)4 (E)5

4.1.5.3.1 解答

$\sigma = 0.0004^{(1/2)} = 0.02$ 、また、標準正規分布のテーブルから

t	U	$Z = F^{-1}(U)$
1	0.1587	-1
2	0.9332	1.5

$$F = 100 \exp(0.01 + Z1 \times 0.02) = 100 \exp(-0.01) = 99$$

$$G = F \times \exp(0.01 + Z2 \times 0.02) = 99 \times \exp(0.04) = 103.05$$

$$G - F = 4.05$$

答え:(D)

4.1.5.4 SOA(2001年5月コース3問題11:一部改題)

ある特別な2年満期の定期保険の一時払純保険料(Z)を、累積関数の逆関数による方法でシミュレーションしたい。

1. 契約年齢は70歳である。

2. 2種類の死亡要因に対する死亡率は以下のとおり。

k	${}_kq_{70}^{(1)}$	${}_kq_{70}^{(2)}$
0	0.10	0.10
1	0.10	0.50

3. 各保険年度末に支払われる死亡給付額は以下のとおり。

保険年度	(1)による死亡	(2)による死亡
1	1,000	1,100
2	1,150	1,200

4. 予定利率=0.0

5. [0,1] 上の一様分布から得られた確率変数は、0.35 であった。

6. 大なる確率変数は、高額 of 給付金支払いに対応するとする。

このシミュレーションにおける Z を求めよ。

選択肢:(A)1,000 (B)1,050 (C)1,100 (D)1,150 (E)1,200

4.1.5.4.1 解答

$$F(0) = 1 - 0.1 - 0.1 - 0.1 - 0.5 = 0.2$$

$$F(1,000) = F(0) + {}_0q_{70}^{(1)} = 0.2 + 0.1 = 0.3$$

$$F(1,100) = F(1,000) + {}_0q_{70}^{(2)} = 0.3 + 0.1 = 0.4$$

$$F(1,150) = F(1,100) + {}_1q_{70}^{(1)} = 0.4 + 0.1 = 0.5$$

$$F(1,200) = F(1,150) + {}_1q_{70}^{(2)} = 0.5 + 0.5 = 1.0$$

$$F(1,000) = 0.3 < 0.35 < F(1,100) = 0.4 \text{ なので } Z = 1,100$$

答え:(c)

4.1.5.5 SOA(2001年5月コース3問題12)

保有契約3件に対して死亡および解約返戻金の支払いを、累積密度関数の逆関数による方法でシミュレーションした。シミュレーションの手順は以下のとおりであり、確率変数はシミュレーションに従って利用するものとする。

手順1. 契約番号1から始め、死亡給付が発生したかどうかをシミュレートする。発生していなければ手順2へ、していれば次の契約にすすむ。

手順2. その契約が解約するかどうかシミュレートする。その後、次の契約に対して手順1

を繰り返す。

- [0,1] 上の一様分布に従う確率変数は以下の順で生成された。

0.3 0.5 0.1 0.4 0.8 0.2 0.3 0.4 0.6 0.7...

- 契約内容、死亡率、解約率は以下のとおり。

契約番号	年齢	死亡率	解約率	死亡給付額	解約返戻金額
1	100	0.4	0.15	10	10
2	91	0.2	0.15	25	20
3	96	0.3	0.15	20	15

このシミュレーションの結果得られた支払額合計を求めよ。

選択肢:(A)30 (B)35 (C)40 (D)45 (E)50

4.1.5.5.1 解答

契約番号	U	死亡率	$U < \text{死亡率か}$	U	解約率	$U < \text{解約率か}$
1	0.3	0.4	はい(給付 10)	-	-	-
2	0.5	0.2	いいえ	0.1	0.15	はい(解約返戻金 20)
3	0.4	0.3	いいえ	0.8	0.15	いいえ

上表から給付合計は $10+20=30$

答え:(A)

4.1.5.6 SOA(2002年11月コース3問題10)

混合分布 Y をシミュレートしたい。

- Y は確率 0.3 で平均 0.5 の指数分布に従い、
- 確率 0.7 で $[-3, 3]$ 上の一様分布に従うものとする。
- $[0, 1]$ 上の一様分布に従う確率変数として 0.25 および 0.69 をこの順で得た。

シミュレーションの結果(Y)を求めよ。ただし、

- 始めの確率変数の値は、小なる値が指数分布に対応し、次の確率変数の値は、小なる値が、 Y の小なる値に対応することとする。
- また指数分布を求める際、逆関数法を使用すること。

選択肢:(A)0.19 (B)0.38 (C)0.59 (D)0.77 (E)0.95

4.1.5.6.1 解答

まず、 $0.25 < 0.3$ なので指数分布を選択することになる。

$$Y = -0.5 \times \log(1 - 0.69) = 0.5855$$

答え:(c)

4.2 分散減少法

4.2.1 シミュレーションの統計的分析

金融商品価格の計算では、ある密度関数 $f(x)$ を前提にして、ある関数 $g(x)$ の平均値、

$$\theta = E[g(X)] = \int g(x)f(x)dx$$

を求めようとすることが多い。前節まででは、この際、密度 f に従う互いに独立な確率変数 X_1, \dots, X_n を生成することによって、 $g(X_i), i=1, \dots, n$ をシミュレートし、

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n g(X_i)$$

を θ の近似値とすると述べた。これをもう少し統計学的に記述すると、 θ を母集団の平均とし、 $\hat{\theta}_n$ をその標本平均¹¹とすると、大数の強法則から

$$\Pr\left\{\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta\right\} = 1$$

が成立するので、 $\hat{\theta}_n$ を θ の推定量とする、となる¹²。さらに標本平均 $\hat{\theta}_n$ は不偏推定量 (Unbiased Estimator)¹³である。すなわち、その期待値が母集団の平均値に等しくなる。

$$E[\hat{\theta}_n] = \frac{1}{n} \sum_{i=1}^n E[g(X_i)] = \frac{1}{n} \cdot n \cdot \theta = \theta$$

また、標本分散

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (g(X_i) - \hat{\theta}_n)^2$$

も母集団の分散 σ^2 の不偏推定量である。すなわち、

$$E[\hat{\sigma}_n^2] = \sigma^2$$

である。また、それぞれの推定量の分散は以下のとおりである。

$$\text{Var}[\hat{\theta}_n] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[g(X_i)] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

¹¹ 標本の平均値と区別し、確率変数(変量)の平均という意味で「標本変量平均」と呼ばれることもある。

¹² これを概収束という。「参考」を参照されたい。

¹³ 「参考」を参照されたい

$$\text{Var}[\hat{\sigma}_n^2] = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right),$$

$$\text{ただし } \mu_4 = \int (g(x) - \theta)^4 f(x) dx$$

さらに、 $\hat{\theta}_n$ と θ の平均 2 乗誤差(mean square error)は $\hat{\theta}_n$ の分散に等しく、

$$E\left[(\hat{\theta}_n - \theta)^2\right] = E\left[(\hat{\theta}_n - E[\hat{\theta}_n])^2\right] = \text{Var}[\hat{\theta}_n] = \sigma^2/n$$

となる。

σ は母集団の標準偏差であるので事前に知ることはできないが、

$$\int (g(x) - \theta)^2 f(x) dx < \infty$$

であれば、 $\hat{\theta}_n$ と θ の標準誤差(平均 2 乗誤差の平方根)は c/\sqrt{n} (c は定数)という形をしていることがわかる。これは、 $\hat{\theta}_n$ という推定量を用いて、モンテカルロ法により θ を推定する場合、その標準偏差は n をシナリオ数とすると \sqrt{n} のオーダーである(これを $o(\sqrt{n})$ とかく)という意味である。 n を増やしてもシミュレーションの精度は \sqrt{n} に比例して増加するに過ぎないことを意味する。例えば 1000 本のシナリオを用いたモンテカルロ法により得られた値の精度を、もう一桁上げるためには、その 100 倍である 10 万本のシナリオが必要となる、という意味である。

より小さな標準誤差を与える推計量を利用できれば、精度を下げずシナリオ数を減らしたり、また逆にシナリオ数を増加させずに精度を上げることができるので、シミュレーション全体の効率を上げることができる。以下に紹介する分散減少法は、上記の標準誤差 c/\sqrt{n} の定数部分(c)を如何に減少させるかを工夫したものである¹⁴。

分散減少法に「負の相関法(Antithetic Variables)」、「制御変量法(control Variates)」、「層化別法(Stratified Sampling)」、「加重サンプリング法(Importance Sampling)」、「条件付分散減少法(Variance Reduction by Conditioning)」、またこれらの組み合わせなど、多数が考案されている¹⁵。本稿では特に金融工学の分野でよく利用される「負の相関法」と「制御変量法」を解説する。

¹⁴ モンテカルロ法の標準誤差が $o(\sqrt{n})$ であるのに対して、数値積分の標準誤差は $o(n^{-2/d})$ (n は積分の区分数、 d は次元)、準乱数を用いた準モンテカルロ法では、解の絶対誤差は $O(n^{-1})$ (n はシナリオ数)といわれている。この分にはモンテカルロ法の「効率」はよくない。ただし、 $d = 4$ 以上の数値積分(多重積分)にあつてはモンテカルロ法は数値積分より効率がよい。また準モンテカルロ法の応用範囲は限られている。準モンテカルロ法に関しては参考文献[8][10]を参照されたい。

¹⁵ 詳細は参考文献[4][5][6][7][8][10]を参照されたい。

その前に、以下では、シミュレーションにより母集団の平均 θ を推定したときの信頼区間を求めよう。

4.2.2 母集団の平均の信頼区間

前述のように標本平均 $\hat{\theta}_n$ は母集団の平均 θ の不偏推定量である。ここで、ある信頼係数(信頼度ともいう)の下、 θ の信頼区間をもとめよう。まず、 $E[\hat{\theta}_n] = \theta$ 、 $Var[\hat{\theta}_n] = \frac{\sigma^2}{n}$ であり、中心極限定理¹⁶から、

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \leq Z_\alpha \right\} = \int_{-\infty}^{Z_\alpha} (1/\sqrt{2\pi}) \exp(-x^2/2) dx$$

である。右辺は標準正規分布である。上記の等号を

$$\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \approx N(0,1)$$

と表現する。 $N(0,1)$ は標準正規分布(平均0、分散1の正規分布)を意味する。 n が十分大きいとき、左辺の分布は右辺の分布で近似できると読む。 σ は母集団の標準偏差であり未知であるが、これを標本標準偏差 $\hat{\sigma}_n$ で置き換えたものも標準正規分布で近似できることが知られている(Slutsky's Theorem)。すなわち、

$$\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n/\sqrt{n}} \approx N(0,1)$$

である¹⁷。このとき、

$$P \left\{ \left| \hat{\theta}_n - \theta \right| < Z_{\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}} \right\} \approx \int_{-Z_{\alpha/2}}^{Z_{\alpha/2}} (1/\sqrt{2\pi}) \exp(-x^2/2) dx = 1 - \alpha$$

であるので、 θ に対する信頼係数 $(1-\alpha)100$ パーセント ($0 < \alpha < 1$ 。例えば $\alpha = 0.05$ のとき信頼係数は 95 パーセント) の信頼区間は、

¹⁶ 「参考」を参照されたい。

¹⁷ Y_1, \dots, Y_n がもともと正規分布に従う確率変数である場合、この統計量は自由度 $n-1$ の t 分布に従うことが知られている。 n が十分大きいとき t 分布は正規分布で近似できるので、本文のように正規分布を前提に区間推定をすることができる。 n が小なるときは、 t 分布を前提に区間推定をすることもできる。

$$\left(\hat{\theta}_n - Z_{\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}, \quad \hat{\theta}_n + Z_{\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}} \right)$$

となる。ここに $Z_{\alpha/2}$ は標準正規分布の $\alpha/2$ 分位数であり、 Z が標準正規分布に従う確率変数のとき、

$$P\{Z > Z_{\alpha/2}\} = \alpha/2$$

を満たす。例えば $\alpha = 0.05$ のとき $Z_{0.05/2} = 1.96$ 、 $P\{Z > 1.96\} = 0.025$ である。

また、この信頼区間から、ある信頼係数の下、必要な精度を得るためのシミュレーション実行回数（シナリオ数）の目安を計算することができる。例えば信頼係数 $100(1-\alpha)$ パーセントのもと（例えば 95パーセント）、母集団のパラメータ θ との誤差を d 以下とするためには、

$$Z_{0.05/2} \frac{\hat{\sigma}_n}{\sqrt{n}} = 1.96 \frac{\hat{\sigma}_n}{\sqrt{n}} \leq d$$

となるまで、シミュレーションを繰り返せばよい。ただしシミュレーションを繰り返しながらその停止条件を判断する算法（計算ロジック）を採用すると、場合によってはシミュレーションが停止しないという問題がある。実務では、これをさけるため、本格的なシミュレーションを実行する前に事前調査を行い、この調査にもとづき n を定め、本格的なシミュレーションを再度行う、という手順を推奨する。まとめると、次の手順が実務的である。

1. 適当なシミュレーション回数 n' （例えば30回）のシミュレーションを行い、 $\hat{\sigma}_{n'}$ を求める。
2. 信頼係数 $100(1-\alpha)$ と必要な精度 d を定める。
3. $(Z_{\alpha/2} \hat{\sigma}_{n'} / d)^2 \leq n$ なる n を定め、再度 n 回のシミュレーションを行う。

4.2.3 負の相関法(Antithetic Variables)

前節までは、母集団の平均を求めるための推定量として標本平均を前提としていた。標本平均は不偏推定量であり、その分散は平均2乗誤差に等しい。ここでもし標本平均の分散より小さな分散を持つ不偏推定量があれば、シミュレーションの結果がより正確なものになると期待できる。また、より正確な値となる不偏推定量ならば、シナリオ数(実行回数)を減らすことができ、シミュレーションがより効率化する。負の相関法と呼ばれる方法は、独立な確立変数を生成するのではなく、負の相関関係にあるものを生成し、これを利用するものである。例えば $n = 2$ とし、 X と Y を共通の分布に従う(独立とは限らない)確率変数、 σ^2 をその分散とし、

$$\hat{\theta}_2^{(1)} = \frac{1}{2}(X + Y)$$

とおくと、これも不偏推定量であるが、

$$\begin{aligned} \text{Var}[\hat{\theta}_2^{(1)}] &= \frac{1}{4}\text{Var}[X + Y] \\ &= \frac{1}{4}(\text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]) \\ &= \frac{1}{2}\sigma^2 + \frac{1}{2}\text{Cov}[X, Y] \end{aligned}$$

となる¹⁸。ここで X と Y に負の相関関係があれば、共分散(Cov)が負であるので、

$$\text{Var}[\hat{\theta}_2^{(1)}] \leq \frac{1}{2}\sigma^2 = \text{Var}[\hat{\theta}_2]$$

となり、シミュレーションの精度を上げることができる。一般に $X_1, \dots, X_M, Y_1, \dots, Y_M$ がすべて同じ分布に従い、 X_i と Y_i ($i = 1, \dots, M$) 以外はすべて互いに独立とすると $X_i + Y_i$ と $X_j + Y_j$ ($i \neq j$) は独立であるから、標本平均

$$\hat{\theta}_{2M}^{(1)} = \frac{1}{2M} \sum_{i=1}^M (X_i + Y_i)$$

の分散は、

¹⁸ $n \geq 2$ の場合の一般式は、 $\text{Var}[\hat{\theta}_n^{(1)}] = \frac{1}{n}\sigma^2 + \frac{1}{n^2} \sum_{i \neq j} \text{Cov}[Y_i, Y_j]$ である。

$$\begin{aligned}
\text{Var}\left[\hat{\theta}_{2M}^{(1)}\right] &= \frac{1}{(2M)^2} \sum_{i=1}^M \text{Var}[X_i + Y_i] \\
&= \frac{M}{(2M)^2} \text{Var}[X + Y] \\
&= \frac{M}{(2M)^2} \{\text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]\} \\
&= \frac{1}{2M} \{\text{Var}[X] + \text{Cov}[X, Y]\}
\end{aligned}$$

となり、 X と Y に負の相関関係があれば、

$$\text{Var}\left[\hat{\theta}_{2M}^{(1)}\right] \leq \frac{1}{2M} \sigma^2 = \text{Var}\left[\hat{\theta}_{2M}\right]$$

となる。次の定理および命題が知られている。

4.2.3.1 定理

1. X を分布 F に従う確率変数とする
2. f と g は増加関数とする。

ならば、

$$\text{Cov}[f(X), g(X)] = E[f(X)g(X)] - E[f(X)]E[g(X)] \geq 0$$

である。

証明

f および g は増加関数であるので、任意の $x \geq y$ (または $x \leq y$) である x 、 y に対して

$$[f(x) - f(y)][g(x) - g(y)] \geq 0$$

である。なぜなら2項とも正(または負)であるからである。よって分布 F に従う独立な数 X 、 Y に対しても

$$[f(X) - f(Y)][g(X) - g(Y)] \geq 0$$

である。すると

$$E\{[f(X) - f(Y)][g(X) - g(Y)]\} \geq 0$$

であり、よって

$$E[f(X)g(X)] + E[f(Y)g(Y)] \geq E[f(X)g(Y)] + E[f(Y)g(X)]$$

である。ところが X と Y は同じ分布にしたがっているので、

$$E[f(Y)g(Y)] = E[f(X)g(X)]$$

更に X と Y の独立性から

$$E[f(X)g(Y)] = E[f(Y)g(X)] = E[f(X)]E[g(X)]$$

であるので、

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)]$$

である。よって、題意は証明された¹⁹。◆

4.2.3.2 命題

1. U を $[0,1]$ 上の一様分布に従う確率変数とする。
2. h を U に関する単調(増加または減少)関数とする。

このとき

$$\text{Cov}[h(U), h(1-U)] \leq 0$$

である。

証明

まず、 h は増加関数とする。 $-h(1-U)$ が U の増加関数になっているので、定理より

$$\text{Cov}[h(U), -h(1-U)] \geq 0$$

$$\text{よって、} \text{Cov}[h(U), h(1-U)] = -\text{Cov}[h(U), -h(1-U)] \leq 0$$

h が減少関数の場合も同様である。◆

¹⁹ 一般の次元の確率変数ベクトル $X = (X_1, \dots, X_n)$ の場合も同様に証明される。次の命題も同じ。参考文献[4][5]を参照されたい。

ここで、分布 F の逆関数 F^{-1} が存在するとする。 $U_i, i=1, \dots$ を $[0,1]$ 上の一様分布に従う確率変数とすると、 $1-U_i, i=1, \dots$ も $[0,1]$ 上の一様分布に従う確率変数となるので

$$X_i = F^{-1}(U_i) \quad Y_i = F^{-1}(1-U_i)$$

とおくと、分布 F に従う確率変数 X_i および Y_i を生成できる。母集団の平均値 θ の推定量として

$$\begin{aligned} \hat{\theta}_{2M}^{(1)} &= \frac{1}{2M} \sum_{i=1}^M \{g(X_i) + g(Y_i)\} \\ &= \frac{1}{2M} \sum_{i=1}^M \{g(F^{-1}(U_i)) + g(F^{-1}(1-U_i))\} \end{aligned}$$

をとると、これは不偏推定量、すなわち

$$E[\hat{\theta}_{2M}^{(1)}] = \theta$$

である。 F は増加関数であるので F^{-1} も増加関数であり、ここで g が単調関数であるなら、 $g \circ F^{-1}$ も単調関数であるので、上記の命題より

$$\begin{aligned} \text{Var}[\hat{\theta}_{2M}^{(1)}] &= \frac{\sigma^2}{2M} + \frac{1}{2M} \text{Cov}[g(F^{-1}(U)), g(F^{-1}(1-U))] \\ &\leq \frac{\sigma^2}{2M} = \text{Var}[\hat{\theta}_{2M}] \end{aligned}$$

となる。すなわち $\hat{\theta}_{2M}^{(1)}$ は推定量 $\hat{\theta}_{2M}$ (標本平均) より小なる平均2乗誤差を与える推定量である。よって $\hat{\theta}_{2M}^{(1)}$ のほうがより母集団の平均に近い結果を推定する推定量であると考えられる。また、一様分布に従う確率変数を $2M$ 個生成する必要はなく、 M 個のみ生成すればよいので、一様分布に従う確率変数を生成する時間も節約できる。金融工学に登場する平均を求めたい関数 ($g(x)$) の多くは単調性を持つので、負の相関法はよく利用される。さらに、分布 F が標準正規分布のとき、

$$F^{-1}(1-U) = -F^{-1}(U)$$

という関係があるので²⁰、更に簡便になり、標準正規分布に従う確率変数 ξ_i に対して、

$$\hat{\theta}_{2M}^{(1)} = \frac{1}{2M} \sum_{i=1}^M \{g(\xi_i) + g(-\xi_i)\}$$

と置くことができる。

4.2.3.3 例

²⁰ 一般に「左右対称な分布」に対して成立する。

$f(x)$ を $[0,1]$ 上の一様分布の密度関数、 $g(x) = \exp(x)$ として

$$\theta = E[g(x)] = \int_0^1 g(x)f(x)dx = \int_0^1 e^x dx$$

を推定しよう。もちろん $\theta = e - 1$ であることは解析的にもとまるのであるが、分散減少法の効果を計算してみる。 U を $[0,1]$ 上の一様分布に従う確率変数とする。まず、

$$E[g(X)] = E[e^U] = e - 1 \doteq 1.7183$$

$$\begin{aligned} \text{Var}[g(X)] &= \text{Var}[e^U] = E[e^{2U}] - E[e^U]^2 \\ &= \int_0^1 e^{2x} dx - (e - 1)^2 = (e^2 - 1)/2 - (e - 1)^2 \doteq 0.2420 \end{aligned}$$

$$E[e^{1-U}] = E[e^U], \quad \text{Var}[e^{1-U}] = \text{Var}[e^U]$$

$$\text{Cov}[e^U, e^{1-U}] = E[e^U e^{1-U}] - E[e^U]E[e^{1-U}] = e - (e - 1)^2 \doteq -0.2342$$

である。 $2M$ 個の独立な一様分布に従う確率変数を使った場合、

$$\text{Var}\left(\frac{1}{2M} \sum_{i=1}^{2M} e^{U_i}\right) = \frac{\text{Var}(e^U)}{2M} = \frac{1}{2M} 0.2420$$

一方、 M 個の一様分布 $U_i, i = 1, \dots, M$ と、 M 個の $1 - U_i, i = 1, \dots, M$ を使った場合、

$$\begin{aligned} \text{Var}\left[\frac{1}{2M} \sum_{i=1}^M (e^{U_i} + e^{1-U_i})\right] &= \frac{1}{2M} (\text{Var}[e^U] + \text{Cov}[e^U, e^{1-U}]) \\ &= \frac{1}{2M} 0.0078 \end{aligned}$$

負の相関法によって分散は $(1 - 0.0078/0.2420) \times 100 = 96.8$ パーセント減少すると期待できる。

(参考文献[4]より) ◆

4.2.3.4 例

「4.2.3.3例」を、実際に一様分布にしたがう乱数を生成してシミュレーションした。まず 10 個の乱数 $(U_i, i = 1, \dots, 10; n = 10)$ を生成し、 $g(U_i) = \exp(U_i)$ を計算した。これらは以下のとおりである。

i	1	2	3	4	5	6	7	8	9	10
U_i	0.5457	0.9319	0.0097	0.5340	0.1758	0.9601	0.0299	0.9894	0.8624	0.1588
$\exp[U_i]$	1.7258	2.5393	1.0097	1.7057	1.1922	2.6120	1.0304	2.6896	2.3688	1.1721

標本平均、標本分散および標本平均の分散は以下のとおりである。

$$\text{標本平均 } \hat{\theta}_{10} = \frac{1}{10} \sum_{i=1}^{10} e^{U_i} \doteq 1.8046$$

$$\text{標本分散 } \hat{\sigma}_{10}^2 = \frac{1}{9} \sum_{i=1}^{10} (e^{U_i} - \hat{\theta}_{10})^2 \doteq 0.4795$$

$$\text{標本平均 } \hat{\theta}_{10} \text{ の分散} = \text{Var}[\hat{\theta}_{10}] = \frac{1}{10} \hat{\sigma}_{10}^2 \doteq 0.04795$$

次に負の相関法の効果を見るために、上記の乱数のうち U_1, \dots, U_5 までと $1-U_1, \dots, 1-U_5$ までを使用して次の表を得る。

i	1	2	3	4	5	平均
U_i	0.5457	0.9319	0.0097	0.5340	0.1758	—
$1-U_i$	0.4543	0.0681	0.9903	0.4660	0.8242	—
$Exp[U_i]$	1.7258	2.5393	1.0097	1.7057	1.1922	1.6346
$Exp[1-U_i]$	1.5751	1.0705	2.6920	1.5936	2.2801	1.8422

負の相関法による平均と分散、および、平均の分散は以下のとおりである。

$$\text{負の相関法による平均 } (\hat{\theta}_{10}^{(2)}) = \frac{1}{10} \sum_{i=1}^5 (e^{U_i} + e^{1-U_i}) \doteq 1.7384$$

$$\text{負の相関法による分散} = \frac{1}{9} \sum_{i=1}^5 \left\{ (e^{U_i} - \hat{\theta}_{10}^{(2)})^2 + (e^{1-U_i} - \hat{\theta}_{10}^{(2)})^2 \right\} \doteq 0.3521$$

$$\text{共分散} = \frac{1}{4} \sum_{i=1}^5 \left\{ (e^{U_i} - 1.634) \cdot (e^{1-U_i} - 1.8422) \right\} \doteq -0.3662$$

$$\begin{aligned} & \text{負の相関法の平均 } \hat{\theta}_{10}^{(2)} \text{ の分散} \\ & = \text{Var}[\hat{\theta}_{10}^{(2)}] = \text{Var}[\hat{\theta}_{10}] + \frac{1}{10} \text{共分散} \doteq 0.04795 - \frac{1}{10} 0.3662 \doteq 0.01133 \end{aligned}$$

この例では、負の相関法によると、分散は $(1 - 0.01133 / 0.04795) \times 100 = 76.4$ パーセント減少している。◆

4.2.3.5 演習

「4.2.3.4例」を $n = 1000$ としてシミュレーションしてみよ。分散の減少効果は「4.2.3.3例」の理論値に近づくであろうか。◆

4.2.4 制御変量法(Control Variates)

負の相関法と同じく、標本平均の分散より小さな分散を与える、母集団の平均値の不偏推計量を考える。まず、その平均が既知である関数 $h(x)$ があると仮定し、その平均を $E[h(x)] = \theta_h$ とする。また、 $X_i, i = 1, \dots$ を密度関数 f に従う確率変数とする。このとき、

$$\hat{\theta}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n \{g(X_i) - (h(X_i) - \theta_h)\}$$

とおくと、 $E[\hat{\theta}_n^{(2)}] = \theta$ であることは明らかであり、 $\hat{\theta}_n^{(2)}$ も θ の不偏推計量である。また、

$$\begin{aligned} \text{Var}[\hat{\theta}_n^{(2)}] &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[g(X_i) - h(X_i)] \\ &= \frac{1}{n} (\text{Var}[g(X)] + \text{Var}[h(X)] - 2\text{Cov}[g(X), h(X)]) \\ &= \frac{1}{n} (\sigma^2 + \sigma_h^2 - 2\text{Cov}[g(X), h(X)]) \end{aligned}$$

$$\text{ここに、} \sigma_h^2 = \text{Var}[h(x)] \text{ とする。}$$

であり、

$$-1 \leq \text{Corr}[g(X), h(X)] = \frac{\text{Cov}[g(X), h(X)]}{\sqrt{\text{Var}[g(X)]} \sqrt{\text{Var}[h(X)]}} = \frac{\text{Cov}[g(X), h(X)]}{\sigma \cdot \sigma_h} \leq 1$$

であるので、

$$-2\sigma \cdot \sigma_h \leq -2\text{Cov}[g(X), h(X)] \leq 2\sigma \cdot \sigma_h$$

よって

$$\frac{1}{n} (\sigma - \sigma_h)^2 \leq \text{Var}[\hat{\theta}_n^{(2)}] \leq \frac{1}{n} (\sigma + \sigma_h)^2$$

である。左の等号は $\text{Corr} = 1$ のとき、右の等号は $\text{Corr} = -1$ のとき成立する。よって $h(x)$ として、

1. $g(x)$ との相関係数が 1 に近く、
2. 分散 (標準偏差) の差が小さく、
3. 平均値 $E[h(x)] = \theta_h$ が既知のもの

が選択できれば $\hat{\theta}_n^{(2)}$ の分散を小さくできる。このような $h(x)$ を制御変量 (Control Variate) という。

以上のアイデアを念頭に、 c を定数として、

$$\hat{\theta}_n^{(3)} = \frac{1}{n} \sum_{i=1}^n \{g(X_i) + c(h(X_i) - \theta_h)\}$$

とおくと、 $\hat{\theta}_n^{(3)}$ は同じく θ の不偏推定量であり、

$$\begin{aligned} \text{Var}(\hat{\theta}_n^{(3)}) &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n \{g(X_i) + c(h(X_i) - \theta_h)\}\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[g(X_i) + c(h(X_i) - \theta_h)] \\ &= \frac{1}{n} \text{Var}[g(X) + c(h(X) - \theta_h)] \\ &= \frac{1}{n} \{ \text{Var}[g(X)] + 2 \cdot c \cdot \text{Cov}[g(X), h(X)] + c^2 \text{Var}[h(X)] \} \end{aligned}$$

である。これは c の 2 次関数と見ることができ、 $\text{Var}[h(X)] > 0$ であるから、

$$c = c^* \equiv -\frac{\text{Cov}[g(X), h(X)]}{\text{Var}[h(X)]}$$

のとき最小値をとる。このとき

$$\begin{aligned} \text{Var}(\hat{\theta}_n^{(3)}) &= \frac{1}{n} \left\{ \text{Var}[g(X)] - \frac{\text{Cov}[g(X), h(X)]^2}{\text{Var}[h(X)]} \right\} \\ &= \text{Var}(\hat{\theta}_n) - \frac{1}{n} \left\{ \frac{\text{Cov}[g(X), h(X)]^2}{\text{Var}[h(X)]} \right\} \leq \text{Var}(\hat{\theta}_n) \end{aligned}$$

となり、標本平均の分散より小さな分散が得られている。両者を比較すると、

$$\begin{aligned} \frac{\text{Var}(\hat{\theta}_n^{(3)})}{\text{Var}(\hat{\theta}_n)} &= \left\{ \text{Var}[g(X)] - \frac{\text{Cov}[g(X), h(X)]^2}{\text{Var}[h(X)]} \right\} / \text{Var}[g(X)] \\ &= 1 - \frac{\text{Cov}[g(X), h(X)]^2}{\text{Var}[h(X)] \cdot \text{Var}[g(X)]} \\ &= 1 - \text{Corr}[g(X), h(X)]^2 \end{aligned}$$

となる。つまり、制御変量法によって分散は $\text{Corr}[g(X), h(X)]^2 \times 100$ パーセント減少する。

$\text{Cov}[g(X), h(X)]$ および $\text{Var}[h(X)]$ は事前には不明であるので、これらをシミュレーションの結果を用いてその推定量

$$\widehat{\text{Cov}}[g(X), h(X)] = \frac{1}{n-1} \sum_{i=1}^n (g(X_i) - \hat{g}(X)) (h(X_i) - \hat{h}(X))$$

$$\widehat{\text{Var}}[h(X)] = \frac{1}{n-1} \sum_{i=1}^n (h(X_i) - \hat{h}(X))^2$$

$$\text{ただし、 } \hat{g}(X) = \frac{1}{n} \sum_{i=1}^n g(X_i), \quad \hat{h}(X) = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

で代用して、 c^* を次の \widehat{c}^* で推定する。

$$\widehat{c}^* = - \frac{\widehat{\text{Cov}}[g(X), h(X)]}{\widehat{\text{Var}}[h(X)]}$$

以上より

$$\hat{\theta}_n^{(3)} = \hat{g}(X) - \frac{\widehat{\text{Cov}}[g(X), h(X)]}{\widehat{\text{Var}}[h(X)]} (\hat{h}(X) - \theta_h)$$

$$\text{Var}(\hat{\theta}_n^{(3)}) = \frac{1}{n} \left\{ \widehat{\text{Var}}[g(X)] - \frac{\widehat{\text{Cov}}[g(X), h(X)]^2}{\widehat{\text{Var}}[h(X)]} \right\}$$

$$\text{ただし、 } \widehat{\text{Var}}[g(X)] = \frac{1}{n-1} \sum_{i=1}^n (g(X_i) - \hat{g}(X))^2$$

を得る²¹。

²¹ $\hat{g}(X)$ 、 $\widehat{\text{Var}}[g(X)]$ はそれぞれ標本平均 $\hat{\theta}_n$ 、標本分散 $\hat{\sigma}_n^2$ であるが、ここでの表記を統一するため導入した。

制御変量法が、金融工学で有効に活用されている例として、算術平均エイジアン・オプション (Arithmetic Average Asian Option、上記の記述では $g(X)$ に相当) の価格付けに、幾何平均エイジアン・オプション (Geometric Average Asian Option、上記に記述では $h(X)$ に相当) を制御変量として利用する例があげられる。後者の価格 (上記の記述では θ_h に相当) は解析的にもとまるのでこれを利用するわけである。

4.2.4.1 回帰分析との関係

制御変量法と回帰分析には、次のような関係がある。まず、

$$\hat{\theta}_n^{(3)} = \hat{g}(X) + \hat{c}^* (\hat{h}(X) - \theta_h)$$

である (再掲)。ここで線形回帰モデル

$$g(X) = \alpha + \beta \cdot h(X)$$

を考える。 α と β の最小2乗推定量 $\hat{\alpha}$ 、 $\hat{\beta}$ は

$$\begin{aligned} \hat{\alpha} &= \hat{g}(X) - \hat{\beta} \cdot \hat{h}(X) \\ \hat{\beta} &= \frac{\sum_{i=1}^n (g(X_i) - \hat{g}(X)) (h(X_i) - \hat{h}(X))}{\sum_{i=1}^n (h(X_i) - \hat{h}(X))^2} \end{aligned}$$

であり、 $\hat{\beta} = -\hat{c}^*$ の関係にあることがわかる。すると、

$$\begin{aligned} \hat{\theta}_n^{(3)} &= \hat{g}(X) - \hat{\beta} (\hat{h}(X) - \theta_h) \\ &= \hat{\alpha} + \hat{\beta} \cdot \theta_h \end{aligned}$$

つまり、 $\hat{\theta}_n^{(3)}$ は $h(X) = \theta_h$ に対応する回帰直線上の値として求まる。観測値の平均値の点 $(\hat{h}(X), \hat{g}(X))$ もこの回帰直線上にあるが、 $\hat{\theta}_n^{(3)}$ は、制御変量 $h(X)$ の、標本平均 $\hat{h}(X)$ と理論値 θ_h の差の分だけ、 $g(X)$ の標本平均 $\hat{g}(X)$ をずらしたものに相当する、ととらえることができる。

4.2.4.2 例

「4.2.3.3例」と同様 $g(X) = \exp(X)$ を考える。制御変量として $h(X) = X$ をとることとする。 U を $[0, 1]$ 上の一様分布に従う確率変数とすると、

$$E[h(X)] = E[U] = 1/2 \quad \text{Var}[h(X)] = \text{Var}[U] = 1/12$$

$$\begin{aligned} \text{Cov}[g(X), h(X)] &= \text{Cov}[e^U, U] = E[e^U \cdot U] - E[e^U]E[U] \\ &= \int_0^1 x e^x dx - (e-1)/2 \\ &= 1 - (e-1)/2 (\doteq 0.1408) \end{aligned}$$

$$\text{Var}[g(X)] = \text{Var}[e^U] = (e^2 - 1)/2 - (e-1)^2 (\doteq 0.2420)$$

これより、制御変量法による分散は、

$$\begin{aligned} \text{Var}[\hat{\theta}_n^{(3)}] &= \frac{1}{n} \left\{ \text{Var}[e^U] - \frac{\text{Cov}[e^U, U]^2}{\text{Var}[U]} \right\} \\ &= \frac{1}{n} \left\{ \text{Var}[e^U] - 12 \text{Cov}[e^U, U]^2 \right\} \doteq \frac{1}{n} 0.0039 \end{aligned}$$

一方、標本平均の分散は、

$$\text{Var}[\hat{\theta}_n] = \frac{1}{n^2} n \text{Var}[e^U] \doteq \frac{1}{n} 0.2420$$

よって制御変量法によって分散は $(1 - 0.0039/0.2420) \times 100 = 98.4$ パーセント減少することが期待できる。◆

4.2.4.3 例

「4.2.4.2例」を実際に10個の一樣乱数を生成してシミュレーションした。 $n = 10$ である。生成した一樣分布に従う乱数等は「4.2.3.4例」と同じである。結果は以下のとおり(再掲)。

i	1	2	3	4	5	6	7	8	9	10
U_i	0.5457	0.9319	0.0097	0.5340	0.1758	0.9601	0.0299	0.9894	0.8624	0.1588
$\exp(U_i)$	1.7258	2.5393	1.0097	1.7057	1.1922	2.6120	1.0304	2.6896	2.3688	1.1721

標本平均、標本分散は以下のとおり。例えば、平均の理論値は、 $E[U] = 0.5$ (これは既知としている)、 $E[\exp(U)] = e - 1 = 1.7183$ (これを求めようとしている)、分散の理論値およびその他の理論値は「4.2.4.2例」で求めたものである。

	平均	分散	平均の理論値	分散の理論値
$h(U) = U$	0.5198	0.1613	0.5000	1/12
$g(U) = \exp(U)$	1.8046	0.4795	1.7183	0.2420

標本平均、標本平均の分散は、以下のとおり。

標本平均	$\hat{\theta}_{10}$	$Var[\hat{\theta}_{10}]$
観測値	1.8046	0.0480
理論値	1.7183	0.0242
理論値との差	5%	98%

ここに、標本平均の分散 ($Var[\hat{\theta}_{10}]$) の観測値 = 標本分散 / 10 = 0.0480 である。また、

	共分散 $\widehat{Cov}[e^U, U]$	\hat{c}^*
観測値	0.2764	-1.7139
理論値	0.1408	-1.69031

制御推定量およびその分散は、以下のとおり。

制御推定量	$\hat{\theta}_{10}^{(3)}$	$Var[\hat{\theta}_{10}^{(3)}]$
観測値	1.7707	0.00058
理論値	1.7183	0.00039
理論値との差	3%	49%

この例では、分散は $(1 - 0.00058 / 0.0480) \times 100 = 98.8$ パーセント減少している。また、推定量の理論値との差も少なくなっている。ただし、この例では「標本平均の分散」および「制御推定量の分散」としてそれぞれの「推定量」を「観測値」としている。実際にはどれほど分散が減少するかを「観測」するには、こういったシミュレーションを何度か行い、それぞれのシミュレーションで得られた平均値の分散を求める必要がある。以下の「4.2.4.6演習」を参照されたい。◆

4.2.4.4 演習

- 「4.2.4.2例」および「4.2.4.3例」における制御変量を、 $h(X) = X$ と同じ分散を持ち、 $g(X)$ を同じ相関係数を持つ関数、例えば $h(X) = X + 1$ に変更しても、制御推定量の値、制御推定量の分散 (の推定量) とともに変わらないことを確認せよ。

2. $h(X) = -X$ 、 $h(X) = 1 - X$ ではどうか。◆

4.2.4.5 演習

1. 「4.2.4.3例」を、 $n = 1000$ としてシミュレーションし、分散の減少効果を確認せよ。
2. この際求められる \hat{c}^* と「4.2.4.3例」の \hat{c}^* を比較し、評価せよ。
3. 95 パーセント信頼区間を求めよ。◆

4.2.4.6 演習

「4.2.4.3例」を 100 回繰り返すこととする。その際、合計 1000 個の一樣乱数を使用するがこれらが、すべて独立であるよう注意する。100 個の標本平均と 100 個の制御推定量が「標本」として得られる。

1. これらの「標本」の分散の理論値は何か。
2. これらの「標本」の分散を実際に求め、「4.2.4.3例」の推定量と比較せよ。
3. これらの「標本」の分散を比較し、分散減少効果を確認せよ。◆

4.2.4.7 演習

1. 「4.2.4.3例」の観測値をもとに、母集団の平均の 95 パーセント信頼区間を、標本平均と制御推定量各々で求めよ。
2. 「4.2.4.6演習」で計算された 100 組の標本平均と制御推定量のうち何割が、各々上記の信頼区間に入っているか、確認せよ。
3. 「4.2.4.2例」から標本平均と制御推定量の理論的な 95 パーセント信頼区間を求めよ。
4. 2.と同じく、「4.2.4.6演習」で計算された 100 組の標本平均と制御推定量の何割が、各々 3.の「理論的」信頼区間内にあるか、確認せよ。◆

4.2.4.8 演習

次のようなシミュレーションを行った。

1. まず「4.3 例」を行い、 \hat{c}^* を得る。
2. 次に、この \hat{c}^* を使い、

$$Y_i = g(X_i) + \hat{c}^*(h(X_i) - \theta_h)$$

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

とおく。

3. あらたに n 個の乱数を生成させて \hat{Y} をシミュレーションする。
このとき、

1. 実際に n 個 (例えば $n = 1000$) の一様乱数を生成し \hat{Y} を求めてみよ。
2. 標本値から、このときの \hat{Y} の分散の推定量を求めよ。
3. この値が、 $\text{Var}[\hat{\theta}_n^{(3)}] = (1/n)0.0039$ に近いことを確認せよ。◆

4.2.4.9 演習

「4.2.4.3例」を、手持ちの回帰分析ツールを利用して再計算し、制御推定量の値が同じとなるか、確認せよ。「4.2.4.5演習」の場合はどうか、確認せよ。◆

4.2.5 演習解答

4.2.4.4.2 相関係数の 2 乗が同じであるので、制御推定量、制御推定量の分散ともに同じである。

4.2.4.6.1 それぞれ $Var[\hat{\theta}_{10}] = 0.2420/10$ 、 $Var[\hat{\theta}_{10}^{(3)}] = 0.0039/10$ である。「4.2.4.3例」では、この値を推定した。これらは 1000 個の乱数を使用しているものの、 $Var[\hat{\theta}_{1000}] = 0.2420/1000$ 、 $Var[\hat{\theta}_{1000}^{(3)}] = 0.0039/1000$ ではないことに注意。

4.2.4.7.1 標本平均:

$$(1.8046 - 1.96 \times \sqrt{0.048}, 1.8046 + 1.96 \times \sqrt{0.048}) = (1.3754, 2.2338)$$

制御推定量:

$$(1.7707 - 1.96 \times \sqrt{0.00058}, 1.7707 + 1.96 \times \sqrt{0.00058}) = (1.7235, 1.8179)$$

個の例では、この信頼区間内に真の値(1.7183)が含まれていないことに注意されたい。 n を十分大きく取る必要がある。

4.2.4.7.3 標本平均:

$$(1.7183 - 1.96 \times \sqrt{0.0242}, 1.7183 + 1.96 \times \sqrt{0.0242}) = (1.4134, 2.0232)$$

制御推定量:

$$(1.7183 - 1.96 \times \sqrt{0.00039}, 1.7183 + 1.96 \times \sqrt{0.00039}) = (1.6796, 1.7570)$$

4.2.4.8.2 ヒント

$$Var[\hat{Y}] = (1/n)[Var(g(X)) + 2\hat{c} * Cov(g(X), h(X)) + \hat{c}^2 * Var(h(X))]$$

であり、これにシミュレーションの結果を代入数する。

4.2.6 参考

本稿で参照した統計学上の基本的な法則・定理などをまとめた。

4.2.6.1 大数の弱法則(Weak law of large number)

1. $\{Z_n\}, n = 1, 2, \dots$ を確率変数の列 (独立性、分布の同一性は仮定しない) とする。
2. $\theta_n = E[\sum_{i=1}^n Z_i]$ および $\sigma_n^2 = \text{Var}[\sum_{i=1}^n Z_i]$ が存在する。
3. $\lim_{n \rightarrow \infty} (\sigma_n^2 / n^2) = 0$ が成立する。

ならば、

$$\lim_{n \rightarrow \infty} P \left\{ w : \left| \frac{\sum_{i=1}^n Z_i(w) - \theta_n}{n} \right| > \varepsilon \right\} = 0, \text{ for } \forall \varepsilon > 0$$

である。

4.2.6.2 大数の強法則(Strong law of large number)(1)

1. $\{Z_n\}, n = 1, 2, \dots$ を独立な確率変数の列とし、同一の分布に従うとする。
2. $\{Z_n\}, n = 1, 2, \dots$ 各々に平均値 θ (共通) が存在する。

ならば、

$$P \left\{ w : \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n Z_i(w)}{n} = \theta \right\} = 1$$

である。

4.2.6.3 大数の強法則(Strong law of large number) (2)

1. $\{Z_n\}, n = 1, 2, \dots$ を独立な確率変数の列 (分布の同一性は仮定しない) とする。
2. $\{Z_n\}, n = 1, 2, \dots$ 各々に平均と分散が存在する。
3. $\lim_{n \rightarrow \infty} (\sum_{i=1}^n \text{Var}(Z_i) / n^2) < \infty$ である。

ならば、

$$P \left\{ \omega : \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n Z_i(\omega) - E \left[\sum_{i=1}^n Z_i(\omega) \right]}{n} = 0 \right\} = 1$$

である。

4.2.6.4 中心極限定理(Central limit theory)

1. $\{Z_n\}, n = 1, 2, \dots$ を独立な確率変数の列とし、同一の分布に従うとする。
2. $\{Z_n\}, n = 1, 2, \dots$ 各々に平均値 $E[Z_n] = \theta$ (共通)、かつ、
3. 分散 $\text{Var}[Z_n] = \sigma^2 (\sigma > 0)$ (共通) が存在する。

ならば、

$$\lim_{n \rightarrow \infty} P \left\{ \omega : \frac{(1/n) \sum_{i=1}^n Z_i(\omega) - \theta}{\sigma/\sqrt{n}} \leq x \right\} = \Phi(x)$$

$$\Phi(x) = \int_{-\infty}^x \left(1/\sqrt{2\pi}\right) \exp(-z^2/2) dz \quad (\text{標準正規分布の分布関数})$$

である。

4.2.6.5 収束のまとめ

$\{Z_n\}, n = 1, 2, \dots$ を確率変数の列 (独立性、分布の同一性は仮定しない) とし、 Z を別の確率変数とする。確率変数列 $\{Z_n\}$ が確率変数 Z に「収束する」という場合、次の 4 つが考えられる。

概収束 (Almost everywhere convergence):

$$P\{\lim_{n \rightarrow \infty} Z_n = Z\} = 1$$

平均 2 乗収束:

$$\lim_{n \rightarrow \infty} E[|Z_n - Z|^2] = 0$$

確率収束 (Convergence in probability):

$$\lim_{n \rightarrow \infty} P\{|Z_n - Z| > \varepsilon\} = 0, \text{ for } \forall \varepsilon > 0$$

法則収束 (Convergence in law):

$$\lim_{n \rightarrow \infty} P\{Z_n \leq x\} = P\{Z \leq x\}$$

収束の強さは以下のとおりである。

概収束 \Rightarrow 確率収束 \Rightarrow 法則収束、および、平均 2 乗収束 \Rightarrow 確率収束 \Rightarrow 法則収束

大数の強法則は概収束、大数の弱法則は確率収束、中心極限定理は法則収束である。

4.2.6.6 不偏推定量(Unbiased Estimator)

ある推定量 $\hat{\theta} = \theta(X_1, \dots, X_n)$ が、母集団のある特性値 θ (例えば平均値) の不偏推定量であるとは、 $\hat{\theta} = \theta(X_1, \dots, X_n)$ の母集団の分布による期待値が θ になることである。すなわち、

$$E[\hat{\theta}] = \theta$$

となることである。

標本平均が母集団の平均の不偏推定量であることは本文中に示した。また、標本平均の分散の期待値も本文中に示した。

標本分散が、母集団の分散の不偏推定量であることの証明は、参考文献[17]を参照されたい。また、標本分散の分散は、参考文献[18]を参照されたい。

4.3 検定

4.3.1 適合度の検定

この節では、一般の「適合度の検定」と呼ばれる「カイ2乗検定」と「KS(Kolmogorov-Smirnov)検定」の概略を紹介する。目的は、シミュレーションで使用する乱数が、いかに目的とする分布からの乱数に適合しているかを検定することである。シミュレーションで用いる一様分布に従う乱数そのものを検定する場合は、この生成された一様分布に従うと考えられる乱数列に対して「適合度の検定」を行うことになる。「適合度の検定」の詳細に関しては、一般の数理統計学の書物を参照していただきたい。

4.3.1.1 カイ2乗検定²²

4.3.1.1.1 概要

母集団からの標本が k 種類の事象 $E_i (i = 1, 2, \dots, k)$ に区分できるとする。それぞれの事象の発生する確率を $p_i (i = 1, 2, \dots, k; \sum_{i=1}^k p_i = 1)$ とし、このような母集団から大きさ n の標本をとり、事象 E_i であったものの個数が $N_i (i = 1, 2, \dots, k)$ であったとする。このとき、母集団の確率 p_i に対して、帰無仮説

$$H_0 : p_1 = p_1^0, p_2 = p_2^0, \dots, p_k = p_k^0$$

を検定するには、統計量

$$T = \sum_{i=1}^k \frac{(N_i - n \cdot p_i^0)^2}{n \cdot p_i^0}$$

に対して、ある臨界値 x_0 を定めて、

$T > x_0$ ならば H_0 を棄却し、

$T < x_0$ ならば H_0 を採択する。

とする。 n が無限大に大きいとき、統計量 T の極限分布は、自由度 $k-1$ のカイ2乗分布であることが知られている。つまり、 H_0 が正しく、 n が十分大きければ、

$$P\{T < x_0\} \approx P\{\chi^2 < x_0\}$$

²² 紹介方法は、国沢清典[17]を参考にした。

である。ここに χ^2 はカイ 2 乗分布に従う確率変数である。

自由度は区分数 k より一つ小さい値である。数学的な証明ではないが、なぜ一つ小さいかという、 N_i および p_i には

$$\begin{aligned} N_1 + N_2 + \cdots + N_k &= n \\ p_1 + p_2 + \cdots + p_k &= 1 \end{aligned}$$

という関係があり、 N_i および p_i は完全には「自由」ではないからである。

n がどれくらい大きければカイ 2 乗分布で「近似」できるかが問題である。一般的には、「すべての期待値 $n \cdot p_i$ が 5 以上となること」とされている²³。

また、

$$\alpha = P\{T > x_0\}$$

は「有意水準」と呼ばれる。有意水準の選び方には絶対的なものはないが、5% または 1% が慣例的に用いられている。表 4.3.1 にカイ 2 乗分布の代表的なパーセント点の値を掲載した²⁴。この表は n が十分大きい時に統計量 T が従う分布の近似値であることに注意されたい。また、この表から、例えば、自由度 10 のカイ 2 乗分布からの確率変数の値が 18.31 以下となる確率は 95% である、と読むことができる。有意水準を 5% とすると臨界値 x_0 は 18.31 であり、統計量 T がこの値より大きければ、帰無仮説 H_0 を棄却することになる。

²³ Knuth[1]による。なお、伏見[2]では 10 以上としているが、実務上 100 程度にすることが望ましいとしている。

²⁴ より詳しい値は、例えば、Abramowitz [19]を参照されたい。

自由度	1%	5%	25%	50%	75%	95%	99%
1	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
2	0.02010	0.1026	0.5753	1.386	2.773	5.991	9.210
3	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
4	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
5	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
6	0.8720	1.635	3.455	5.348	7.841	12.59	16.81
7	1.239	2.167	4.255	6.346	9.037	14.07	18.48
8	1.646	2.733	5.071	7.344	10.22	15.51	20.09
9	2.088	3.325	5.899	8.343	11.39	16.92	21.67
10	2.558	3.940	6.737	9.342	12.55	18.31	23.21
11	3.053	4.575	7.584	10.34	13.70	19.68	24.73
12	3.571	5.226	8.438	11.34	14.84	21.03	26.22
15	5.229	7.261	11.04	14.34	18.25	25.00	30.58
20	8.260	10.85	15.45	19.34	23.83	31.41	37.57
30	14.95	18.49	24.48	29.34	34.80	43.77	50.89
50	29.71	34.76	42.94	49.33	56.33	67.50	76.15

4.3.1.1.2 例

さいころを、100 回ふったところ以下の表のような結果を得た。このさいころは正しいか²⁵。

目の数 i	1	2	3	4	5	6
頻度 N_i	19	16	12	18	18	17

さいころが正しいものとするれば、帰無仮説は、

$$H_0 : p_1 = 1/6, p_2 = 1/6, \dots, p_6 = 1/6$$

となる。各目の出る期待値は $100/6$ であるから、

$$T = \sum_{i=1}^6 \frac{(N_i - 100/6)^2}{100/6} = 1.88$$

をえる。これを自由度 $6-1=5$ のカイ 2 乗分布の 95 パーセント点 11.07 と比較すると、 $T = 1.88 < 11.07$ である。 H_0 は採択された(正しくないさいころとは言えない)。

別のさいころを振ったところ、次の結果を得た。

²⁵ 国沢清典[17]より。

目の数 i	1	2	3	4	5	6
頻度 N_i	10	16	18	28	12	16

この場合の統計量は 11.84 であり、 $T = 11.84 > 11.07$ であるから、このさいころが正しいとする帰無仮説は棄却された(正しくない)。◆

4.3.1.1.3 連続型への応用

カイ 2 乗検定は基本的には離散型の分布に対する検定であるが、連続型の分布に対しても、母集団の事象を $E_i (i = 1, 2, \dots, k)$ に区分し、それぞれの事象の発生確率 $p_i (i = 1, 2, \dots, k)$ を求めることにより、まったく同様にして、検定することができる。連続型のいわゆる「裾」部分の発生確率が非常に小さい場合には、区分をまとめるなどして、発生数の期待値 $n \cdot p_i$ が小さくなり過ぎないように設定する必要がある。

4.3.1.1.4 例

$n = 50$ 個の正規分布に従うと思われる乱数列を生成し、以下の表のような区分内にある数値の頻度を数えた。この数列は正規分布からの乱数列であるか。

から		-3.0	-2.4	-1.8	-1.2	-0.6	0.0	0.6	1.2	1.8	2.4	3.0
まで	-3.0	-2.4	-1.8	-1.2	-0.6	0.0	0.6	1.2	1.8	2.4	3.0	
頻度	0	1	1	0	8	12	12	8	6	2	0	0

次の表のとおり、-1.2 以下および 1.2 以上をまとめ直し、6 区分として統計量 T を求めた。自由度 5 のカイ 2 乗分布の 95% 点は 11.07 であるので、 $T = 3.42 < 11.07$ であるから、帰無仮説は採択された(正規分布からの乱数でないといえない)。

から		-1.2	-0.6	0.0	0.6	1.2	
まで	-1.2	-0.6	0	0.6	1.2		合計
N_i	2	8	12	12	8	8	50
p_i	0.11507	0.15918	0.22575	0.22575	0.15918	0.11507	1
$n \cdot p_i$	5.75	7.96	11.29	11.29	7.96	5.75	50
$(N_i - n \cdot p_i)^2 / n \cdot p_i$	2.44873	0.00021	0.04498	0.04498	0.00021	0.87716	3.42

◆

4.3.1.1.5 統計量 T のシミュレーション

統計量 T は、標本数 n が十分大きければカイ 2 乗分布で近似できることは前述のとおりであるが、そうでないとき、その分布を解析的に求めることができるとは限らない。よって正確な $P\{T > x_0\}$ の値(仮に「棄却確率」とよぶ)は求まらない。ここでは「対象となる母集団の分布に従う乱数列」を生成することができたとして、シミュレーションを応用して、 T の分布の棄却確率を求めてみる。手順は以下のとおりである。

1. 対象となる母集団の分布に従う乱数を n 個生成し、統計量 T を求める。
2. これらの乱数と独立な乱数をさらに n 個生成し、同様に統計量 T を求める。
3. 以上を r 回繰り返し、求めた統計量 T を $T(1), T(2), \dots, T(r)$ とする。
4. $\{T(1), T(2), \dots, T(r)$ のうち $> x_0$ となったものの数 $\} / r$ を棄却確率とする。

カイ 2 乗分布に従う確率変数からの乱数は、例えば、 $Z_i (i = 1, 2, \dots, n)$ を標準正規分布に従う確率変数とすると、 $\chi^2 = \sum_{i=1}^n Z_i^2$ が自由度 $n-1$ のカイ 2 乗分布に従う確率変数となる²⁶ことを利用して、生成することができる。標準正規分布に従う乱数列を生成することができること、つまり標準正規分布に従うことが検定によって確認された(棄却されていない)乱数列を使用すること、が前提となる。

4.3.1.2 KS(Kolmogorov-Smirnov)検定

4.3.1.2.1 概要

カイ 2 乗検定が基本的には離散型の分布に対する検定であるのに対して、KS 検定は連続型の分布に対する検定方法である。確率変数 X に従う分布関数(distribution function)を $F(X)$ とすると $P\{X < x\} = F(x)$ という関係がある。この確率変数の変量を n 回観測し観測値 X_1, X_2, \dots, X_n を得たとする。このとき「経験分布関数(empirical distribution)」 $F_n(X)$ を

$$F_n(x) = \frac{1}{n} \{x \text{ 以下の } X_1, X_2, \dots, X_n \text{ の数} \}$$

と定義する。KS 検定は分布 $F(x)$ と分布 $F_n(x)$ の差を用いるものである。乱数列が分布 $F(x)$ によく従うものであるならば、 $F(x)$ と $F_n(x)$ の差は小さなものになるだろう。下の図はこの様子のイメージを表現したものである。図 4.3.1 に比較して図 4.3.2 は $F(x)$ と $F_n(x)$ の差が大きく、 $F_n(x)$ がこれほど大きく乖離することは、乱数列が目的とする分布に従う確率変数からの乱数列ではないのではないかと疑問をもたせる。どの程度乖離しているかを知るのが KS 検定である。

²⁶ 4.1 乱数の生成法を参照されたい。

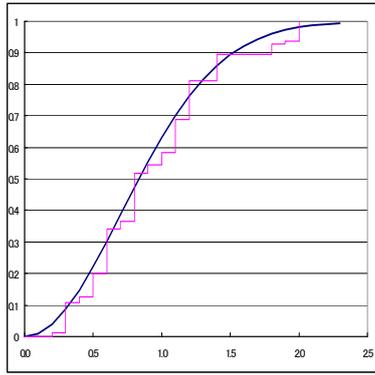


図 4.3.1

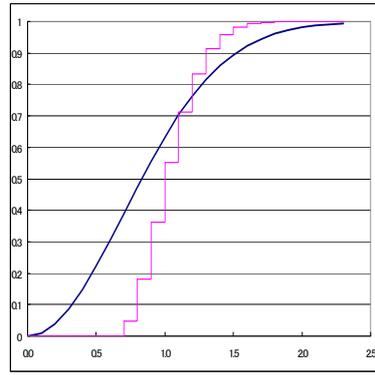


図 4.3.2

KS 検定では次の統計量を使用する。

$$K_n^+ = \sqrt{n} \max_{-\infty < x < \infty} (F_n(x) - F(x))$$

$$K_n^- = \sqrt{n} \max_{-\infty < x < \infty} (F(x) - F_n(x))$$

ここに \sqrt{n} は、 x を固定したとき $F_n(x)$ の期待値が $F(x)$ であり、標準偏差が $1/\sqrt{n}$ に比例するところから、 K_n^+ および K_n^- の標準偏差が標本数 n に依存しないように工夫したものである。これは以下の様に考えることができる。 N_n を

$$N_n = \{x \text{ 以下の } X_1, X_2, \dots, X_n \text{ の数} \}$$

とおくと、 N_n は二項分布に従い、その確率は、 $X_i < x$ となる確率が $F(x)$ であるところから $F(x)$ 、よって平均と標準偏差はそれぞれ $n \cdot F(x)$ および $\sqrt{n \cdot F(x) \cdot (1 - F(x))}$ となる。 $F_n(x) = N_n / n$ であるから、 $F_n(x)$ の平均と標準偏差はそれぞれ $F(x)$ および $\sqrt{F(x) \cdot (1 - F(x))} / \sqrt{n}$ となる。なお、 \sqrt{n} を除いて、

$$D_n^+ = \max_{-\infty < x < \infty} (F_n(x) - F(x))$$

$$D_n^- = \max_{-\infty < x < \infty} (F(x) - F_n(x))$$

とする場合もある。

K_n^+ および K_n^- の従う分布は、分布 $F(x)$ にかかわらず同じものである。これはカイ 2 乗検定で使用する統計量 T が分布 $F(x)$ によって異なることと大きく異なる点である。次の定理が成立する。

4.3.1.2.2 定理

X_1, X_2, \dots, X_n が分布 $F(x)$ に従う確率変数とする。このとき K_n^+ および K_n^- の従う分布は $F(x)$ に依存しない。

証明: K_n^+ で証明する。 K_n^- でも同じ。まず定義より、

$$P\{K_n^+ < d\} = P\left\{\sqrt{n} \max_{-\infty < x < \infty} \left(\frac{\#i: X_i \leq x}{n} - F(x)\right) < d\right\}$$

ここに $(\#i: X_i < x)$ は「 $X_i < x$ となる i の数」を表す。 $F(x)$ は単調関数であるから、

$$= P\left\{\sqrt{n} \max_{-\infty < x < \infty} \left(\frac{\#i: F(X_i) \leq F(x)}{n} - F(x)\right) < d\right\}$$

X_i が確率変数とすると $F(X_i)$ は一様分布に従う確率変数であるから、 U_i を一様分布に従う確率変数として、

$$= P\left\{\sqrt{n} \max_{-\infty < x < \infty} \left(\frac{\#i: U_i \leq F(x)}{n} - F(x)\right) < d\right\}$$

$F(x)$ は $(0,1)$ 内の値をとるので $y = F(x)$ とおくと $-\infty < x < \infty$ のとき $0 < y < 1$ であるので、

$$= P\left\{\sqrt{n} \max_{0 < y < 1} \left(\frac{\#i: U_i \leq y}{n} - y\right) < d\right\}$$

これは K_n^+ の分布が $F(x)$ に依存しないことを示している。◆

K_n^+ および K_n^- は x のあらゆる点を対象として最大値を求めるものであるが、 $F_n(X)$ は観測値 x_1, x_2, \dots, x_n のところだけで値が増加する階段関数であるので、最大値は n 個の観測値の点 x_1, x_2, \dots, x_n のいずれかで取る。よって n 個の観測値を大きさの順に並べ替え $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ とすると、

$$K_n^+ = \sqrt{n} \max_{1 \leq j \leq n} \left(\frac{j}{n} - F(x_{(j)})\right)$$

$$K_n^- = \sqrt{n} \max_{1 \leq j \leq n} \left(F(x_{(j)}) - \frac{j-1}{n}\right)$$

となる。 n が大きくなると観測値の並べ替えには相当の時間と記憶容量を必要とするが、そもそも並べ替えをしない方法など、効率よく計算する方法が提案されている²⁷。

n が十分大きく乱数列が分布 $F(x)$ に従うものとするとき K_n^+ および K_n^- の極限分布は次のようになる。

²⁷ Knuth[1]および伏見[2]を参照されたい。

$$\lim_{n \rightarrow \infty} P\{K_n^+ < d\} = \lim_{n \rightarrow \infty} P\{K_n^- < d\} = 1 - \exp(-2d^2)$$

これから、有意水準 α を与える点 x_α 、すなわち、

$$P\{K_n^+ > x_\alpha\} = P\{K_n^- > x_\alpha\} = \alpha$$

を満たす点 x_α は、

$$x_\alpha \approx \sqrt{-(1/2)\log \alpha}$$

として近似することができる。 n が小さい場合は数表から求める。表 4.3.2 は K_n^+ および K_n^- の分布の代表的なパーセント点である。カイ 2 乗検定にあつては統計量 T の分布をカイ 2 乗分布で近似するが、KS検定で用いるこの表は何らかの近似の分布ではなく K_n^+ または K_n^- の分布の表である²⁸。

表:4.3.2 K_n^+ および K_n^- の分布の代表的なパーセント点

n	1%	5%	25%	50%	75%	95%	99%
1	0.01000	0.05000	0.25000	0.50000	0.75000	0.95000	0.99000
2	0.01400	0.06749	0.29290	0.51760	0.70710	1.0980	1.2728
3	0.01699	0.07919	0.31120	0.51470	0.75390	1.1017	1.3589
4	0.01943	0.08789	0.32020	0.51100	0.76420	1.1304	1.3777
5	0.02152	0.09471	0.32490	0.52450	0.76740	1.1392	1.4024
6	0.02336	0.10020	0.32720	0.53190	0.77030	1.1463	1.4144
7	0.02501	0.10480	0.32800	0.53640	0.77550	1.1537	1.4246
8	0.02650	0.10860	0.32800	0.53920	0.77970	1.1586	1.4327
9	0.02786	0.11190	0.32740	0.54110	0.78250	1.1624	1.4388
10	0.02912	0.11470	0.32970	0.54260	0.78450	1.1658	1.4440
11	0.03028	0.11720	0.33300	0.54390	0.78630	1.1688	1.4484
12	0.03137	0.11930	0.33570	0.54530	0.78800	1.1714	1.4521
15	0.03424	0.12440	0.34120	0.55000	0.79260	1.1773	1.4606
20	0.03807	0.12980	0.34610	0.55470	0.79750	1.1839	1.4698
30	0.04354	0.13510	0.35090	0.56050	0.80360	1.1916	1.4801
∞	0.07089	0.16010	0.37930	0.58870	0.83260	1.2239	1.5174

4.3.1.2.3 例

乱数列を 10 個生成し、これらを小さい順に並べると以下の表になった。

²⁸ Knuth[1]より。

12 14 19 23 78 81 91 130 143 162

この乱数列は平均を 100 とする指数分布 $F(x) = 1 - \exp(-x/100)$ からの乱数列であるか、KS 検定をした。計算は以下の様になる。

j	1	2	3	4	5	6	7	8	9	10
$x_{(j)}$	12	14	19	23	78	81	91	130	143	162
$F(x_{(i)})$	0.113	0.131	0.173	0.205	0.542	0.555	0.597	0.727	0.761	0.802
$\frac{j}{n} - F(x_{(i)})$	-0.013	0.069	0.127	0.195	-0.042	0.045	0.103	0.073	0.139	0.198
$\frac{F(x_{(i)})}{-(j-1)/n}$	0.113	0.031	-0.027	-0.095	0.142	0.055	-0.003	0.027	-0.039	-0.098

この結果、 $K_{10}^+ = \sqrt{10} \cdot (0.198) = 0.626$ 、 $K_{10}^- = \sqrt{10} \cdot (0.142) = 0.449$ を得る。これらの数値は、有意水準を 5% としたときの 95 パーセント点 1.1658 より小さいので、平均を 100 とする指数分布からの乱数列ではないとはいえない。

10 個の乱数列が、

66 72 81 94 112 116 124 140 145 155

である場合は、 $K_{10}^+ = 0.671$ 、 $K_{10}^- = 1.528$ であり、 K_{10}^- が 95 パーセント点 1.1658 より大きいから、この乱数列が平均を 100 とする指数分布からの乱数列であるとする仮定は棄却される。◆

4.3.1.2.4 K_n^+ 分布のシミュレーション

K_n^+ の分布をシミュレーションを用いて生成しよう。 K_n^+ の統計量 x_0 が与えられたとき $K_n^+ < x_0$ となる確率を与えるものである。 K_n^+ (および K_n^-) の分布は前述の様に分布 $F(x)$ に依存せず定まるものであるから、分布 $F(x)$ として一様分布を採用する。一様分布に従う乱数 U_i が生成できるものとして(または、検定の結果、一様分布に従う乱数として採用された乱数列を使用して)、シミュレーションを行う。手順はカイ 2 乗分布のシミュレーションと同じである。

1. 一様分布に従う乱数を n 個生成し、統計量 K_n^+ を求める。
2. これらの乱数と独立な乱数をさらに n 個生成し、同様に統計量 K_n^+ を求める。
3. 以上を r 回繰り返し、求めた統計量 K_n^+ を $K_n^+(1), K_n^+(2), \dots, K_n^+(r)$ とする。
4. $\{K_n^+(1), K_n^+(2), \dots, K_n^+(r)\}$ のうち $< x_0$ となったもの数 $/ r$ を求める確率とする。

4.3.1.2.5 カイ 2 乗検定とKS検定の比較

カイ 2 乗検定と KS 検定を比較すると、カイ 2 乗検定は基本的には離散型の分布を検定する方法ではあるが、事象を区分することによって連続型の分布の検証にも応用できる。これに対して KS 検定は連続型の分布に対する検定であり、事象の区分を必要としない。またカイ 2 乗検定は標本数 n が小さいときその正しい分布を求めることは困難であり、カイ 2 乗分布で近似して検定する。KS 検定では標本数 n が小さいときの近似によらない分布が求められている。Knuth[1]によれば、カイ 2 乗検定と KS 検定の優劣を比較するための理論的な成果はそれほど多くないという。また、カイ 2 乗検定は比較的多くの標本数 n を必要とするので、連続型の分布を検定する限りは KS 検定がやや優れているという。

4.3.2 一様乱数の検定

乱数列の統計的検定は、一様乱数に対してだけでなく、一般的な乱数列に対しても考えることができる。つまり、シミュレーションで用いる一般的な分布からの乱数列を、一様分布に従う乱数に変換するなどして生成し、この乱数列の対して検定することを考えるわけである。例えば、正規分布に従うと思われる乱数列に対して「正規性の検定」と呼ばれる手法が知られている。しかし、このようにして生成された一般の乱数列に対して検定が行われることは、ほとんどないようである。その理由として伏見[2]では、次のように述べている。

「乱数の変換に使われる手法は、正確なものであるか、または近似を用いるものであっても誤差が実用上無視しうる程度に小さいものである場合がほとんどである。したがって変換前の数列に対する検定を充分に行っておけば、変換後の数列に対する検定は行わなくてもよい。また、一様乱数は必要に応じていろいろな乱数に変換して使われるので、そのつど検定するのは大変にわずらわしい」。

以上の理由付けの正当性の判断はおくとして、本稿ではKnuth[1]に従って、一様乱数列の検定方法をいくつか紹介する。検定方法は文字通り山のように提案されている。どの検定方法を選択するかは、乱数を使用する目的、つまり、シミュレーションの目的に従って選択すべきとされている。しかし、どの目的にどの検定方法が適しているかを判断することが、また困難である。Knuth[1]は、

「検定の目的は、適した乱数生成法を選択することではなく、適さないものを選択すること

としている。いくつかの検定を実施し、消去法で残った乱数生成法を採用するわけである。

例えば、線形合同法またはM系列法等によって生成された数列²⁹は、その周期を如何に長くするか、および如何に均等分布を可能にするかを工夫したものである。M系列乱数にいたっても、真の乱数の満たす条件の一部を近似的に満足するのみである。生成された数列が如何に「ばらばら」であるかを検定するため、様々な角度からの検定が提案されている。

4.3.2.1 一様分布検定(または頻度検定)

この検定は、数列が一様分布に従うかを検定するものであり、カイ2乗検定またはKS検定をそのまま、一様分布に応用するものである。カイ2乗検定をおこなう場合は、数列の出現区間を均等に分割し、その分割区間内にある頻度を観察する。KS検定の場合は、数列を(0,1)区間の数列に変換し、一様分布の分布関数 $F(x) = x$ と比較することになる。

Knuth[1]は、カイ2乗検定を使って一様乱数の検定を行うときの注意事項を述べている。要約する

²⁹ 「4.4 一様乱数の生成」を参照されたい。

と以下のとおりである。

1. 試行回数 n は十分大きく取るほうがよい(前述のようにすべての i に対して $n \cdot p_i \geq 5$ となるぐらいの n)。但し n が余りに大きいと乱数列の「局所的」な偏り・規則性がならされてしまう可能性がある。つまりある部分の傾向が、その後の逆の傾向でならされてしまい、判断を誤る可能性がある。計算機で生成する乱数列には、このような局所的規則性は大きいとありえる問題である。いろいろな n に対してカイ2乗検定を行うべきである。
2. KS 検定にも同じ問題が存在し、KS 検定の場合は妥協点として $n=1000$ を提唱している。その代わりにこのような KS 検定量(例えば K^+)を多数もとめ、これらの結果に対してもう一度 KS 検定を行う。こうすることによって局所的な規則性も大域的な規則性も検定することができる。
3. カイ2乗検定および KS 検定ともに、その統計量が大きいことは乱数列の偏りを意味するので、帰無仮説は否定されるのであるが、逆に統計量が小さすぎても不自然である。計算機を利用して生成された乱数列を検定する場合には、特に注意が必要である。

4.3.2.2 系列検定(serial test)

系列検定はカイ2乗検定の考え方を2次元以上に拡張したものである。例えば2次元の場合、 $(0,1)$ 上の一様分布に従う乱数(と思われる)数列から、2次元配列 $(U_{2i}, U_{2i+1}), i=0,1,\dots,n-1$ をとり、単位正方形内に一様にばらまかれているかを検証する。一般に k 次元の場合、区分単位 d (正の整数)を適当にとり、一方、単位 k 次元立方体を d^k 個に分割し、このセル内にいくつの配列が入ったかを数える。 n 個の配列を観察するとすると、各セルに入っている配列の個数の期待値は n/d^k であるので、

$$\chi^2 = \sum_{i=1}^{d^k} \frac{(\text{セル}i\text{内の観測個数} - n/d^k)^2}{n/d^k}$$

を検定することになる。

この検定は、次元が高くなるにつれて計算時間が大幅に長くなるので、高次元の検定は実務上困難である。これを補うためにその他の様々な検定方式が提案されている。

4.3.2.3 間隔検定(gap test)

間隔検定は、ある特定の範囲内の数が出現する間隔を検定するものである。例えば $(0,1)$ 上の一様分布に従う(と思われる)数列 $U_j, j=0,1,\dots,n-1$ に対して、適当に α 及び β ($0 \leq \alpha < \beta \leq 1$) を選んで、 $U_j, U_{j+1}, \dots, U_{j+r-1}$ は (α, β) 内にはないが、 U_{j+r} がこの範囲内にあるとき、間隔を $r+1$ とする。例えば $(\alpha, \beta) = (0.3, 0.7)$ とし、次の数列が得られたとする。この範囲内の数値の次に区切り線(|)を入れた。

0.12 0.36 | 0.02 0.89 0.64 | 0.46 | 0.11 0.08 0.94 0.55 | 0.19 0.76

間隔はそれぞれ 2、3、1、4、2 である。 $p = \beta - \alpha$ とすると、間隔が r である確率 $P[r]$ および間隔が t 以上である確率 $P[t]$ は、

$$P[r] = (1-p)^{r-1} p, r = 1, 2, \dots, t-1$$

$$P[t] = (1-p)^t - 1$$

である。 r が大きくなるとその確率は極めて小さくなると考えられるので t を適当に定め、これ以降をまとめる。検定は、

$$\chi^2 = \sum_{r=1}^t \frac{(N[r] - n \cdot P[r])^2}{n \cdot P[r]}$$

として、カイ 2 乗検定を行う。 $N[r]$ ($r = 1, 2, \dots, t-1$) は間隔が r であった場合の観測値、 $N[t]$ は間隔が t 以上であった場合の観測値とする。

$(\alpha, \beta) = (0, 0.5)$ または $(0.5, 1)$ と定めると、それぞれ「中心から下の数列の長さの検定」および「中心から上の数列の長さの検定」となる。

4.3.2.4 ポーカー検定(poker test)または分割検定(partition test)

古典的なポーカー検定と呼ばれるものは、相次ぐ 5 つの数値 $Y_{5j}, Y_{5j+1}, \dots, Y_{5j+4}$ ($j = 0, 1, \dots$) に対

し、各組が次の 7 つの型のどれになるかを考えるものである。

- ケース1: すべて異なる(abcde)
- ケース2: 2個同種類が1組(ワン・ペア:aabcd)
- ケース3: 2個同種類が2組(ツー・ペア:aabbc)
- ケース4: 3個同種類が1組(スリー・オブ・ザ・キング:aaabc)
- ケース5: 2個同種類が1組と3個同種類が1組(フルハウス:aabbb)
- ケース6: 4個同種類が1組(フォー・オブ・ザ・キング:aaaab)
- ケース7: 5個同種類が1組(ファイブ・オブ・ザ・キング:aaaaa)³⁰

ところが、このような区分の方法を取扱うことは困難であるため、以上の区分を簡単にし、以下のような相異なる数何種類あるかを考えることにする。

- ケース1: 1種類の数値(5個ともおなじ)
- ケース2: 2種類の数値(フルハウスかフォー・オブ・ザ・キング)
- ケース3: 3種類の数値(ツー・ペアかスリー・オブ・ザ・キング)

³⁰ 実際のポーカーでは、このような組み合わせは出現しない。

ケース4: 4種類の数値(ワン・ペア)

ケース5: 5種類の数値(すべて異なる)

例えば0, ..., 9の10個の数値から5個の数値をランダムに選択し、12237であったとすると、数値の種類は4である。これをあと5回繰り返し、82328、33453、66662、19378、21889を得たとすると、数値の種類はそれぞれ3、3、2、5、4であるので、都合、ケース1=0、ケース2=1、ケース3=2、ケース4=2、ケース5=1とおりである(合計6回の試行)。

分割検定では、1から d までの d 個の整数から乱数を生成し、相次ぐ k 個の組(1, 2, ..., d のいずれかで構成される)を n 組観測し、 r 種類($1 \leq r \leq k$)の数からなる k 個組の数を数える(この数を $N[r]$ とする)。 k 個の数の中に r 種類の数がある確率 $P[r]$ は

$$P[r] = \frac{d(d-1)\cdots(d-r+1)}{d^k} \begin{Bmatrix} k \\ r \end{Bmatrix} \quad (1)$$

であるので、

$$\chi^2 = \sum_{r=1}^k \frac{(N[r] - n \cdot P[r])^2}{n \cdot P[r]}$$

によって、カイ2乗検定を行う。

ここに、

$$\begin{Bmatrix} k \\ r \end{Bmatrix}$$

は、 k 個の数値(の位置)を丁度 r 個に分割する分割の方法の数(「第2種スターリング数」と呼ばれる)であり、その各々に対して、 d 個から r 個の異なる数を取り出す方法が $d(d-1)\cdots(d-r+1)$ とおりにある。これを d 個の数値から k 個の数値を取り出す組み合わせ d^k 通りで割ることで、 k 個の相続く数の中に r 種類の数値がある確率 $P[r]$ となる。

「第2種スターリング数」には、次の関係がある。

$$\begin{Bmatrix} k \\ r \end{Bmatrix} = r \begin{Bmatrix} k-1 \\ r \end{Bmatrix} + \begin{Bmatrix} k-1 \\ r-1 \end{Bmatrix} \quad (2)$$

この関係は、 k 個のものを r 組に分けるには、1個を取り除いておいて r 組に分け、取り除いた1個を r 組のいずれかに入れる(第1項)か、1個だけでひとつの組を作成し、残りの $k-1$ 個を $r-1$ 組に分けるか(第2項)、のいずれかをすればよいところから理解できる。

$$\begin{Bmatrix} k \\ r \end{Bmatrix} = \frac{1}{r} \sum_{s=0}^r (-1)^{r-s} \binom{r}{s} s^k$$

という関係もある³¹。以下に $k = 1$ から 10 までの第 2 種スターリング数を掲載した。 $\left\{ \begin{matrix} k \\ 1 \end{matrix} \right\} = 1$ 、

$\left\{ \begin{matrix} k \\ k \end{matrix} \right\} = 1$ であることと、関係式 (2) から容易に一般の「第 2 種スターリング数」が求まる。

表 4.3.3: 第 2 種スターリング数

$k \setminus r$	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	1	3	1							
4	1	7	6	1						
5	1	15	25	10	1					
6	1	31	90	65	15	1				
7	1	63	301	350	140	21	1			
8	1	127	966	1,701	1,050	266	28	1		
9	1	255	3,025	7,770	6,951	2,646	462	36	1	
10	1	511	9,330	34,105	42,525	22,827	5,880	750	45	1

例えば、 $\{B1, B2, \dots, B5\}$ の 5 個の位置を 2 種類に区分する方法は

$\{B1 | B2, B3, B4, B5\}, \{B2 | B1, B3, B4, B5\}, \{B3 | B1, B2, B4, B5\},$
 $\{B4 | B1, B2, B3, B5\}, \{B5 | B1, B2, B3, B4\}, \{B1, B2 | B3, B4, B5\},$
 $\{B1, B3 | B2, B4, B5\}, \{B1, B4 | B2, B3, B5\}, \{B1, B5 | B2, B3, B4\},$
 $\{B2, B3 | B1, B4, B5\}, \{B2, B4 | B1, B3, B5\}, \{B2, B5 | B1, B3, B4\},$
 $\{B3, B4 | B1, B2, B5\}, \{B3, B5 | B1, B2, B4\}, \{B4, B5 | B1, B2, B3\}$

の 15 通りある。| は区分の位置を表す。

4.3.2.5 札集め検定(coupon collector's test)

札集め検定は、例えば、お菓子のおまけをすべて買い揃えるまでにお菓子をいくつ買わなければならないか、という問題に関連する。シミュレーションでは Y_1, Y_2, \dots を 1 から d までの整数をとる乱数とし、1 から d までの整数がそろった部分列 $Y_{j+1}, Y_{j+2}, \dots, Y_{j+r}$ の長さを観測する。たとえば $d = 3$ とし、 Y_1, Y_2, \dots, Y_7 とし、

1 2 2 1 2 3 1 3 2 2 3 3 2 1 2 3 1

³¹ Abramowitz[19]より。

が得られたとする。1 から 3 までがひとつおとり得られたところで区分すると(| で区分した位置をあらわす)、

$$1\ 2\ 2\ 1\ 2\ 3\ | \ 1\ 3\ 2\ | \ 2\ 3\ 3\ 2\ 1\ | \ 2\ 3\ 1$$

となる。ひとつおりの数がえられるまでの部分列の長さ(r)は6、3、5、3である($r \geq d = 3$)。長さが3、5、6の部分列の数は、各々、2、1、1である。長さが4の部分列、長さが7以上の部分列の数はどちらも0である。

部分列の長さが r である確率 $P[r]$ 、部分列の長さが t 以上である確率 $P[t]$ は、

$$\begin{aligned} P[r] &= \frac{d!}{d^r} \binom{r-1}{d-1}, r = d, d+1, \dots, t-1 \\ P[t] &= 1 - \frac{d!}{d^{t-1}} \binom{t-1}{d} \end{aligned} \quad (1)$$

であるので、

$$\chi^2 = \sum_{r=d}^t \frac{(N[r] - n \cdot P[r])^2}{n \cdot P[r]}$$

としてカイ2乗検定をおこなう。ここに $N[r]$ ($r = d, d+1, \dots, t-1$)は長さが r の部分列の数の観測値、 $N[t]$ は長さが t 以上の部分列の観測値である。 r が大きい場合の確率は非常に少なくなると考えられるから、 t を適当に選択し、これらをまとめる。

(1)式は次のように求める。まず、 r 個の数列の中に d 個すべての数が(丁度 r 番目に全部そろったわけではないが、ともかく)そろっている確率は、 r 個の数の中に d 種類の数がある確率に等しいから、「ポーカー検定」の(1)式より、

$$\frac{d!}{d^r} \binom{r}{d}$$

である。そこで、部分列の長さが r である確率は、

$$\begin{aligned} P[r] &= P\{r\text{個めに丁度そろう}\} \\ &= P\{r\text{個中にすべてそろっている}\} \\ &\quad - P\{r-1\text{個中にすべてそろっている}\} \end{aligned}$$

$$\begin{aligned}
&= \frac{d!}{d^r} \left\{ \begin{matrix} r \\ d \end{matrix} \right\} - \frac{d!}{d^{r-1}} \left\{ \begin{matrix} r-1 \\ d \end{matrix} \right\} \\
&= \frac{d!}{d^r} \left(d \left\{ \begin{matrix} r-1 \\ d \end{matrix} \right\} + \left\{ \begin{matrix} r-1 \\ d-1 \end{matrix} \right\} \right) - \frac{d!}{d^{r-1}} \left\{ \begin{matrix} r-1 \\ d \end{matrix} \right\} \\
&= \frac{d!}{d^r} \left\{ \begin{matrix} r-1 \\ d-1 \end{matrix} \right\}
\end{aligned}$$

となる。最後から 2 番目の等号は、「ポーカー検定」の(2)式から得られる。また、部分列の長さが t 以上となる確立は、

$$\begin{aligned}
P[t] &= P\{t\text{個以上めに丁度そろう}\} \\
&= 1 - [P\{d\text{個めに丁度そろう}\} + \dots + P\{t-1\text{個めに丁度そろう}\}] \\
&= 1 - \left[\begin{array}{l} P\{d\text{個そろっている}\} \\ + P\{d+1\text{個そろっている}\} - P\{d\text{個そろっている}\} \\ + \dots \\ + P\{t-1\text{個そろっている}\} - P\{t-2\text{個そろっている}\} \end{array} \right]
\end{aligned}$$

であることから、

$$\begin{aligned}
P[t] &= 1 - P\{t-1\text{個そろっている}\} \\
&= 1 - \frac{d!}{d^{t-1}} \left\{ \begin{matrix} t-1 \\ d \end{matrix} \right\}
\end{aligned}$$

となる。

4.3.2.6 順列検定

順列検定は、乱数列の大小関係の場合の数を数えることによって行う検定である。まず、乱数列を t 個 n 組生成する。 t 個の数の大小関係は $t!$ 個の場合がある。 n 組の中にこれらの大小関係がいくつずつ出現したかを数える。例えば、 $t = 3$ とし乱数列を U_1, U_2, U_3 とするとこれらの大小関係は以下の $3! = 6$ とおりある。

$$\begin{aligned}
&\text{ケース1: } U_1 < U_2 < U_3, \quad \text{ケース2: } U_1 < U_3 < U_2 \\
&\text{ケース3: } U_2 < U_1 < U_3, \quad \text{ケース4: } U_2 < U_3 < U_1 \\
&\text{ケース5: } U_3 < U_1 < U_2, \quad \text{ケース6: } U_3 < U_2 < U_1
\end{aligned}$$

たとえば、(0,1) 上の一様乱数として、以下の 3 個 4 組 ($t = 3, n = 4$) が生成されたとする。| は 3 個ずつの区分の位置である。

$$0.12 \ 0.36 \ 0.02 \ | \ 0.89 \ 0.64 \ 0.46 \ | \ 0.11 \ 0.08 \ 0.94 \ | \ 0.80 \ 0.89 \ 0.76$$

それぞれケース 5、ケース 6、ケース 3、ケース 5 に相当する。よって、 $N[r]$ をケース r の出現度頻度とすると、 $N[1]=0$ 、 $N[2]=0$ 、 $N[3]=1$ 、 $N[4]=0$ 、 $N[5]=2$ 、 $N[6]=1$ である。ケース 1 からケース $t!$ の確率はどれも $1/t!$ であるから

$$\chi^2 = \sum_{r=1}^{t!} \frac{\left(N[r] - n \cdot \frac{1}{t!} \right)^2}{n \cdot \frac{1}{t!}}$$

として、カイ 2 乗検定を行う。この検定では乱数に等しいものはないと仮定する。

4.3.2.7 連の検定

連の検定は、生成した乱数が単調減少(または単調増加)している部分の長さを観測し、理論値と比較する。この検定の場合も乱数に等しいものはないと仮定する。目的は乱数列の傾向を検定することである。大きな数の後、単調減少する数が続く傾向があれば、良い検定結果は得られない。例えば、次の 11 個の乱数が生成されたとする。両端と単調減少ではなくなった位置を | で示した。

| 0.12 | 0.36 0.02 | 0.89 0.64 0.46 0.11 0.08 | 0.55 | 0.89 0.76 |

単調減少している部分の連の長さは、1、2、5、1、2 である。ここで注意すべきは、あい隣り合う「連」が独立ではないことである。第一の連の長さは 1 である。これは 2 つ目の数が 0.36 であったためであるが、2 つ目の連の長さもこの 0.36 に依存する。連のあいだの独立性が満たされていないため、このままではカイ 2 乗検定が応用できない³²。そこで、連を区切る乱数を捨てて、数列を区切りなおす。次のような数列を得ることになる。

| 0.12 | 0.02 | 0.64 0.46 0.11 0.08 | 0.89 0.76 |

単調減少している部分の連の長さは、1、1、4、2 となった。一般に長さ r の連、 $U_{j+1}, U_{j+2}, \dots, U_{j+r}$ が得られたら、 U_{j+r+1} を捨て、 U_{j+r+2} から新たに連の長さを数え始めることとする。こうすることによって、絶えず新たな数列の連の長さを観測することになるので、連のあいだの独立性が得られる。連の長さが r である確率 $P[r]$ および連の長さが t 以上である確率 $P[t]$ は、単調減少、単調増加にかかわらず次のとおりである。

$$P[r] = \frac{1}{r!} - \frac{1}{(r+1)!}, r = 1, 2, \dots, t-1$$

$$P[t] = \frac{1}{t!}$$

³² 実は工夫をしてこのままの連の定義に対してカイ 2 乗検定を定義できる。詳細は Knuth[1]を参照されたい。

例えば、連の長さが2となる確率は、 X, Y, Z を $(0,1)$ 上の一様分布の確率変数として、

$$\begin{aligned}
 P[2] &= P\{X > Y < Z\} \\
 &= \int_0^1 \int_0^x \int_y^1 dz dy dx \\
 &= \int_0^1 \int_0^x (1-y) dy dx \\
 &= \int_0^1 \left(x - \frac{1}{2}x^2\right) dx \\
 &= \frac{1}{2!} - \frac{1}{3!}
 \end{aligned}$$

である。一般の場合も同様に、連の長さが r となる確率は、 X_0, X_1, \dots, X_r を $(0,1)$ 上の一様分布の確率変数として、

$$\begin{aligned}
 P[r] &= \int_0^1 \int_0^{x_1} \cdots \int_0^{x_{r-2}} \int_{x_{r-1}}^1 dx_r dx_{r-1} \cdots dx_0 \\
 &= \frac{1}{r!} x_0^r - \frac{1}{(r+1)!} x_0^{r+1} \Big|_0^1 \\
 &= \frac{1}{r!} - \frac{1}{(r+1)!}
 \end{aligned}$$

である。連の長さが t 以上となる確率は、

$$P[t] = \sum_{r=t}^{\infty} \left(\frac{1}{r!} - \frac{1}{(r+1)!} \right) = \frac{1}{t!}$$

である。そこで、連を n 個生成し

$$\chi^2 = \sum_{r=1}^t \frac{(N[r] - n \cdot P[r])^2}{n \cdot P[r]}$$

としてカイ2乗検定を行う。ここに $N[r]$ ($r=1, 2, \dots, t-1$)は、連の長さが r であった場合の数の観測値、 $N[t]$ は連の長さが t 以上であった場合の観測値とする。

4.3.2.8 t個の最大値検定

この検定は、乱数を一定数ずつ生成し、その中の最大値を求めて検定するものである。例えば以下の3個4組の乱数列を得たとする。組の区切りを|で示している。

0.12 0.36 0.02 | 0.89 0.64 0.46 | 0.11 0.08 0.94 | 0.80 0.89 0.76

各々の組中の最大値、

0.36 0.89 0.94 0.89

に対して検定を行う。一般には、 $(0,1)$ 上の一様分布に従う t 個 n 組の乱数列、

$$j+1\text{組みめの乱数列} = \{U_{ij}, U_{ij+1}, \dots, U_{ij+t-1}\} (j = 0, 1, \dots, n-1)$$

を生成し、

$$V_j = \max \{U_{ij}, U_{ij+1}, \dots, U_{ij+t-1}\} (j = 0, 1, \dots, n-1)$$

とおき、数列 V_0, V_1, \dots, V_{n-1} に対してKS検定を行う。KS検定を行うのは、

$$\begin{aligned} & P \{ \max(U_0, U_1, \dots, U_{t-1}) \leq x \} \\ &= P \{ U_0 \leq x, U_1 \leq x, \dots, U_{t-1} \leq x \} \\ &= P \{ U_0 \leq x \} \cdot P \{ U_1 \leq x \} \cdots P \{ U_{t-1} \leq x \} \\ &= x^t \end{aligned}$$

であるように、 $V_j (j = 0, 1, \dots, n-1)$ が連続分布 $F(x) = x^t$ に従うからである。または、 $V_j^t (j = 0, 1, \dots, n-1)$ が一様分布に従うとして検定することもできる。

4.3.2.9 衝突検定

衝突検定も、乱数列の一様性を検定するものである。乱数列が適度にばらついていれば、特定の部分に偏らないであろう。これはたくさんの壺にいくつかの玉を投げ入れ、2個以上玉の入っている壺の数を検定することに似ている。ただし、衝突検定では、衝突した回数の合計を数える。例えば、壺の中に3個の玉が入っていれば衝突を2回と数える。一般には壺を多次元に配置し、乱数列の多次元における一様性を検定する。例えば0から9までの10種類の整数を取る乱数の2次元の一様性を検定したいとする。 $10^2 = 100$ 個の壺、 $\{壺_{ij} : i = 0, 1, \dots, 9; j = 0, 1, \dots, 9\}$ を用意する。以下のような16個の乱数列が生成されたとする。|で2個ずつの区切りを示した。

$$18 | 36 | 47 | 92 | 15 | 60 | 47 | 55$$

それぞれ壺₁₈、壺₃₆等などに配置される。壺₄₇で衝突が1回が発生しているので、結局この試行における衝突回数の合計は1回である。

一般には、0から $d-1$ までの d 個の整数を取る乱数列に対して、 k 次元の一様性を検定するとき、 $m = d^k$ 個の位置(壺)を用意し、 k 個 n 組の整数の組、

$$j+1\text{組みめの乱数列} = \{Y_{kj}, Y_{kj+1}, \dots, Y_{kj+k-1}\} (j = 0, 1, \dots, n-1)$$

ただし、 Y は0から d までの整数

を生成する。これらと m 個の位置との衝突回数の合計を計測する。衝突回数の合計が r である確

率 $P[r]$ は、

$$P[r] = \frac{m(m-1)\cdots(m-(n-r)+1)}{m^n} \binom{n}{n-r}$$

となる。これは、衝突の回数の合計が r であるとき、 n 個の玉は丁度 $n-r$ 個の位置(壺)に入っている。 n 個を丁度 $n-r$ 個に分割する場合は $\binom{n}{n-r}$ である。この各々の場合に対して、 m 個の壺から $n-r$ 個の玉の入っている壺を選択する場合は $m(m-1)\cdots(m-(n-r)+1)$ である。これを、 m 個の壺に n 個の玉を入れる組み合わせである m^n とおりで割れば、衝突回数の合計が r である確率が求まる。

先の例では、 $d=2$ 、 $k=2$ 、 $m=10^k=100$ 、 $n=8$ である。この条件のもと、1 回の衝突が発生する確率はつぎのとおりである。まず、8 個の玉を 7 つに分割する。分割の方法は $\binom{8}{7}=28$ とおりある。この各々に対して 7 個の壺を割り振る場合の数は $100 \times 99 \times \cdots \times 94$ である。8 個の玉を 100 個の壺に入れる入れかたは 100^8 とおりある。よってこの確率は、

$$\frac{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96 \cdot 95 \cdot 94}{100^8} 28 = 0.226$$

となる。 $r=0$ から 7 までの確率と累積確率は以下のとおりである。

r	0	1	2	3	4	5	6	7
確率	75.0%	22.6%	2.3%	0.1%	0.0%	0.0%	0.0%	0.0%
累積確率	75.0%	97.6%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%

この率は、壺の数 m を固定し、 n 個の玉が s 個の壺に入っている場合の確率 $q_m(n, s)$ の漸化式からも求めることができる。

$$q_m(n, s) = \frac{s}{m} q_m(n-1, s) + \frac{m-(s-1)}{m} q_m(n-1, s-1),$$

ここに

$$q_m(1, s) = \begin{cases} 1, & s=1 \\ 0, & s \neq 1 \end{cases}$$

$$q_m(n, 0) = 0$$

これは、 n 個の玉が s 個の壺に入る確率は、 $n-1$ 個の玉が s 個の壺に入っており、もう 1 個玉を加えたときにその s 個の壺にはいる確率(第 1 項)と、 $n-1$ 個の玉が $s-1$ 個の壺に入っており、もう 1 個玉を加えたときに、空いている壺に入る確率(第 2 項)を加えたものである、というところから理解できる。

$$P[r] = q_m(n, n-r)$$

として衝突回数の確率が求まる。先の例の場合の $q_m(n, s)$ は以下のとおりである。

$s \setminus n$	1	2	3	4	5	6	7	8
1	100.0%	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2		99.0%	3.0%	0.1%	0.0%	0.0%	0.0%	0.0%
3			97.0%	5.8%	0.2%	0.0%	0.0%	0.0%
4				94.1%	9.4%	0.6%	0.0%	0.0%
5					90.3%	13.6%	1.3%	0.1%
6						85.8%	18.0%	2.3%
7							80.7%	22.6%
8								75.0%

4.4 一様乱数の生成

4.4.1 はじめに

この節では、一様乱数の生成方法について述べる。一様乱数の生成方法は古くから研究され多くの提案がある。ここで紹介する方法は、コンピュータで計算させて生成する乱数であり、算術乱数とも呼ばれるものである。モンテ・カルロ・シミュレーションでは、非常に多くの一様乱数を必要とするので、以下に紹介する一様乱数の生成方法の多くのものが、コンピュータのオペレーション・システムに密着している。歴史的に線形合同法から発展した分野であり、生成法の多くは代数学を基礎とする。代数学への深入りは本稿の目的とするところではないので、本稿では基本となる定理を紹介するにとどめる。参考文献に代数学の基本書を紹介しているので、参考にされたい。

一様乱数の生成方法として「物理乱数源」を用いることもできる。これは電子管などの雑音や放射性物質からの雑音として検出される物理現象、例えば放射性同位元素のガンマ崩壊を利用するものである。これらの装置をコンピュータのシステムの一部として組み込み、利用する方法である。

この方法に対して、コンピュータにあらかじめ組み込んだ算法(ロジックまたは手順)に従い、生成する乱数を「算術乱数」と呼ぶ。これは、算法に従い乱数が生成されるので、どのような乱数が生成されるかをあらかじめ予測することができる。算術乱数はこの意味で「真の意味の乱数」ではなく「擬似乱数(pseudorandom numbers)」と呼ばれる³³。まず擬似乱数に求められる条件を考察する(参考資料 津田[8])。

1. 多数の乱数を速やかに生成できること。前述のようにモンテ・カルロ・シミュレーションでは多くの乱数を必要とする。場合によっては数億個以上の乱数を必要とするので、これらの乱数を高速で生成する必要がある。
2. 生成する乱数に周期があるならば、その周期は十分長いものであること。多くの乱数を必要とするので周期が短いと、生成した乱数全体が「ばらついて」いないことになる。
3. 再現性があること。乱数はその一様性などの統計的性質を検証したのち使用するが、検証で使用した乱数列を再現できなければ、検証の意味がない。また、研究対象のモデルを変更し、モデルの変更がもたらす影響を評価する際、変更前に使用した乱数列と同じ乱数列を使用して評価したい。このためには乱数列に再現性がなければならない。以下に紹介する乱数生成法は「初期値」を与えることによって、生成する乱数列を制御することができる。
4. 生成した乱数列が「良好な統計的性質」を持つこと。統計的検定を行いパスしたものでなければならない。

³³ 「真の意味の乱数」は何かという問題は、実は哲学的な問題ですらある。ここでは単に「予想のできない数値の出現」を意味する程度に理解されたい。

以下では、算術乱数の内、その性質がよく研究されている「線形合同法(Linear Congruential Method)」を紹介し、線形合同法の欠点と指摘される「多次元結晶構造」の説明と「スペクトル検定」を紹介する。最後に「M 系列乱数」を紹介する。

4.4.2 線形合同法(Linear Congruential Method)

現在使われている最もよく知られた一様乱数生成法である。1948 年ごろ、D.H. Lehmer が発表したものとされている(Knuth[1])。算法は以下のとおり。

$$X_{n+1} = (aX_n + c) \bmod M, \quad n \geq 0$$

ここに

- M : 法 (Modulus), $0 < M$ である整数
- a : 乗数(Multiplier), $0 \leq a < M$ である整数
- c : 増分(Increment), $0 \leq c < M$ である整数
- X_0 : 初期値(Seed), $0 \leq X_0 < M$ である整数

である。 $c = 0$ のときを「乗算合同法(Multiplicative Congruential Method)」、 $c \neq 0$ のときを「混合合同法(Mixed Congruential Method)」という。 $\bmod M$ は「計算結果を M で割り、余りを取る」の意味である。初期値 X_0 から始め順次 X_1, X_2, \dots と求まる。 M で割った余りは $\{0, 1, 2, \dots, M-1\}$ までの M 通りあるので、上手く M, a, c, X_0 を選択すれば最大 M 個の数値が得られる。一旦同じ数値が出現したらそこから先は以前の繰り返しとなる。この意味で線形合同法の最長周期は M である。最長周期が得られなくとも十分大きな周期が得られれば実用的である。十分大きな周期を得るための M, a, c, X_0 の条件は、代数学から与えられる。この $\{X_n\}$ から

$$U_n = X_n / M, n \geq 0$$

として数値列 $\{U_n\}$ を求めると、これは $[0, 1)$ 内の実数値をとる数値列となる。

4.4.2.1 線形合同法の性質のまとめ

1. いったん $X_m = 0$ となるとそれ以降の $X_n (n > m)$ はすべてゼロとなる。
2. 数列のどこかで以前出現した数値が出ると、それ以降は以前の数値の繰り返しとなる。繰り返し部分を「循環部分」、循環部分の長さを「周期」という。
3. M, a, c, X_0 のとり方によって、「ばらつき」た数列が得られるとは限らない。
4. これらが如何に「ばらつき」ているかは、統計的検定による評価を行わなければならない。
5. 数列の最長周期は M である。

4.4.2.2 例

$M = 10, a = c = X_0 = 7$ のとき、 X_0, X_1, \dots は、

7, 6, 9, 0, 7, 6, 9, 0, ...

であり、周期は 4 となる。最長周期 10 を得られていない。◆

線形合同法による数列は、「再現性」があることは明らかである。「高速に生成できること」および「良好な統計的性質を有すること」は、後に議論することとして、まず十分長い周期を持つための条件を考察する。最長周期は M であるから、周期が M となるための a および c の条件を考える。これは次の定理によって与えられる。

4.4.2.3 定理

線形合同法による数列が最長周期 M を持つための必要十分条件は、

1. c と M が互いに素であり、かつ、
2. $a-1$ が M を割り切るすべての素数 p の倍数であり、かつ、
3. M が 4 の倍数であれば、 $a-1$ も 4 の倍数であること、

である。

証明は Knuth[1] (3rd Ed 日本語版 pp15)、津田[8](pp19)を参照されたい。◆

この定理から次の定理が導かれる。

4.4.2.4 定理

$M = 2^\beta$ ($\beta \geq 2$) のとき、線形合同法による数列が最長周期 M を持つための必要十分条件は、

1. c と M が互いに素であり、かつ、
2. $a-1$ が 4 の倍数である。◆

4.4.2.5 例

1. $M=8, a=9, c=1$ のとき(「4.4.2.3 定理」の条件1, 2および3を満足する)。 $X_0=0$ とすると、数列は $\{0, 1, 2, \dots, 7\}$ である。
2. $M=18, a=7, c=1$ のとき(「4.4.2.3 定理」の条件1および2のみ満足する)。 $X_0=0$ とすると、数列は $\{0, 1, 8, 3, 4, 11, 6, 7, 14, 9, 10, 17, 12, 13, 2, 15, 5\}$ である。
3. $M=2^8=256, a=9, c=1$ のとき(「4.4.2.4 定理」の条件1および2を満足する)。
4. $M=2^{32}, a=69069, c=1$ のとき(「4.4.2.4 定理」の条件1および2を満足する)。

◆

これらの定理によると、乗算合同法では $c = 0$ であるので、最長周期 M を得ることはできない。しかし c を加える部分の計算が省略できるので、計算時間を短縮できるというメリットがある。また、本稿では詳細には言及できないが、 c は「ばらつき」に大きな影響を与えない、とされており、乗算合同法はよく研究されている。

乗算合同法 ($c = 0$) では、 X_n がいったんゼロとなるとそれ以降の数列はすべてゼロとなってしまうため、法 M のもとでは X_n の取りえる範囲は $\{1, 2, \dots, M - 1\}$ でなければならない。よって乗算合同法の最長周期は $M - 1$ である。次の定理がある。

4.4.2.6 定理

乗算合同法 ($c = 0$) の可能な最長周期は、 a の M に関する位数 $\lambda(a, M)$ であり、この最長周期は、

1. X_0 と M が互いに素であり、
2. a が M を法とする原始根である

とき得ることができる。ここに、位数および原始根の定義は以下のとおり。

「 a の M に関する位数 $\lambda(a, M)$ 」とは、

$a^\lambda = 1 \pmod{M}$ とする最小の整数 λ のことである。位数の最大値は $M - 1$ である。

「 a が M を法とする原始根である」とは、

位数をその最大値である $M - 1$ とする数 a のことである。

証明は Knuth[1] (3rd Ed 日本語版 pp19) を参照されたい。◆

4.4.2.7 例

1. $a = 2, M = 7$ のとき $2^1=2, 2^2=4, 2^3=1$ であるので、位数 $\lambda(2, 7) = 3$ である。位数がその最大値 6 ではないので、 $a = 2$ は原始根ではない。
2. $a = 5, M = 7$ のとき $5^1=5, 5^2=4, 5^3=6, 5^4=2, 5^5=3, 5^6=1$ であるので、位数 $\lambda(5, 7) = 6$ である。位数がその最大値 6 であるので、 $a = 5$ は原始根である。実際、 $M = 7$ と素である $X_0 = 4$ をとると、 X_0, X_1, \dots, X_5 は $\{4, 6, 2, 3, 1, 5\}$ であり、最長周期 6 がえられている。



Knuth[1]ではさらにこの一般形の定理をうけ、素数とは限らない一般の M に対して、原始根 a を求めるための必要十分条件を与えている。実は、 M が素数である場合には、原始根が必ず存在することが知られており、その場合の最長周期と最長周期を得るための条件は、次の定理のとおりである。この定理は整数論の基本的定理から理解することができる。

4.4.2.8 定理

乗算合同法($c=0$)の最長周期($M-1$)は次のとき得ることができる。

1. 法 M を素数とする(これを p とおく)。
2. p の原始根の一つを a とする(p の原始根は必ずひとつ存在する)。
3. $X_0 \neq 0$ とする。

証明の概要

整数論の結果によれば、素数の既約剰余系は巡回群であり、その巡回群の生成元が原始根となることから、この定理は明らかであるが、ここではもう少し整数論を復習することとする³⁴。

定義(剰余類)

整数 a, b について $a-b$ が M で割り切れるとき、 a と b は法 M によって合同であるという。ある a と合同なすべての整数の集合を剰余類という。

定義(完全剰余系)

法 M に対して、各剰余類から 1 つずつ整数を取り出して並べた集合を、完全剰余系という。

定義(既約剰余系)

法 M の完全剰余系のうち、 M と互いに素の要素だけを取り出したものを既約剰余系という。

定義(群)

集合 G において、その任意の2元 x, y に対して $z = x \bullet y, z \in G$ となるような演算 \bullet を考える。この演算が次の性質をもつとき、 G を群という。

1. 任意の3元 x, y, z に対して、 $(x \bullet y) \bullet z = x \bullet (y \bullet z)$
2. 元 e が存在して、任意の元 x に対して $e \bullet x = x \bullet e = x$
3. 任意の元 x に対して、対応する元 x^{-1} が存在し、 $x \bullet x^{-1} = x^{-1} \bullet x = e$

補題

既約剰余系は乗法に関して群をなす。

定義(巡回群)

有限の群で、すべての要素が1つの要素 a の累乗となる群を巡回群という。

補題

素数の既約剰余系は巡回群である。

以上の証明は例えば「遠山[14]155 ページ」を参照されたい。

M を素数とすると、補題から、 $\{1, 2, \dots, M-1\}$ は既約剰余系であり、乗法に関して群をなし、かつ

³⁴ 書きぶりは湯前[10]を参考にした。

巡回群である。巡回群には生成元が存在し、これを a とすると $a^1 = a, a^2, a^3, \dots, a^{M-1} = 1$ はすべて異なり、各々は $\{1, 2, \dots, M-1\}$ のいずれかである。すなわち最長周期が得られた。 a の M に関する位数は $M-1$ となり、これは位数の最大値であるので、 a は原始根である。◆

4.4.2.9 例

1. 1 と 8 は法 7 によって合同である(7 で割った余りがどちらも 1 である)。よって同じ剰余類に入る。
2. $\{0, 1, 2, 3, 4, 5, 6\}$ は法 7 の完全剰余系の一つである。
3. $\{1, 2, 3, 4, 5, 6\}$ は法 7 の既約剰余系である。 $\{1, 5\}$ は法 6 の既約剰余系である。
4. $\{1, 2, 3, 4, 5, 6\}$ は法 7 (と乗法) に関して巡回群である。その生成元の一つは 5 である。すなわち 5 は素数 7 の原始根の一つである。◆

素数 p が与えられたとき、 p を法とする原始根は必ず存在する。いくつ原始根が存在するかは、オイレル(Euler)の関数 φ によって計算できる。

4.4.2.10 補題

素数 p の原始根は $\varphi(p-1)$ 個存在する。ここに φ はオイレルの関数であり、

$$\varphi(n) = n \text{ より小さく } n \text{ と互いに素である正整数の数}$$

である。これは、

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right), \text{ 積は } n \text{ を割るすべての素数 } p \text{ にわたる。}$$

によって計算できる。

証明は遠山[14]を参照されたい。◆

4.4.2.11 例

M=7 M-1=6=2×3 であるので $\varphi(6)=6(1-1/2)(1-1/3)=2$ となる。実際に 6 より小さく 6 と互いに素である正整数は 1 と 5 の 2 個である。また、試行錯誤の結果、実際に原始根は 3 と 5 の 2 個である事がわかる。

M=11 M-1=10=2×5 であるので $\varphi(10)=10(1-1/2)(1-1/5)=4$ となる。実際に 10 より小さく 10 と互いに素である正整数は 1, 3, 7 および 9 の 4 個である。また、試行錯誤の結果、実際に原始根は 2, 6, 7 および 8 の 4 個である事がわかる。

M=13 M-1=12=2×2×3 であるので $\varphi(12)=12(1-1/2)(1-1/3)=4$ となる。実際に 12 より小さく 12 と互いに素である正整数は 1, 5, 7 および 11 の 4 個である。また、試行錯誤の結果、実

際に原始根は 2、6、7 および 11 の 4 個である事がわかる。

$M=2^{31}-1=2,147,483,647$ (メルセンヌ素数)

$M-1=2^{31}-2=2 \cdot 3^2 \cdot 7 \cdot 11 \cdot 31 \cdot 151 \cdot 331$ であるので $\varphi(2^{31}-2)=534,600,000$ 個の原始根が存在する。一般に 2^n-1 の形で表すことのできる素数を「メルセンヌ素数」という。◆

原始根が存在すること、また、いくつ存在するかを知ることはできたが、実際、何が原始根であるかを発見することは容易ではない。しかし一つ発見できれば、その他の原始根を発見する方法がある。

4.4.2.12 補題

a を法 M の原始根とすると、 $M-1$ と互いに素である任意の整数 K に対して、 $B = a^K \pmod{M}$ も原始根である。◆

4.4.2.13 例

$M=13$ とする。 $M-1=12$ と互いに素である整数のうち $M-1$ より小さいものは $\{1,5,7,11\}$ である。2 は原始根であるが、 $2^1=2$ 、 $2^5=6$ 、 $2^7=11$ 、 $2^{11}=7$ も原始根である。◆

「4.4.2.8 定理」と同等の定理として次の定理がある。

4.4.2.14 定理

乗算合同法 ($c=0$) の最長周期 ($M-1$) は次のとき得ることができる。

1. 法 M を素数とする(これを p とおく)。
2. $p-1$ の任意の素因数 q に対して、

$$a^{\frac{p-1}{q}} \not\equiv 1 \pmod{p}$$

3. $X_0 \neq 0$ とする。◆

「4.4.2.8 定理」の条件 2 と「4.4.2.14 定理」の条件 2 が同等の条件であるところから理解できよう。「4.4.2.14 定理」によれば、「原始根」を意識せず $p-1$ の素因数に対して条件 2 を検証すれば、最長周期を与える a を求めることはできる。ただし M が大きければ、その検証が大変な計算ではある。

以上、乗算合同法における最長周期 ($M-1$) を得るための条件として、 M が素数である場合を紹介した。乗算合同法において最長周期を得られなくとも、十分大きな周期が得られればよい。次の定理がある。

4.4.2.15 定理

乗算合同法($c=0$)で、 $M=2^\beta$ ($\beta \geq 4$) の場合、可能な最長周期は $M/4$ であり、次の場合にのみ可能である。

1. $a=3(\text{mod } 8)$ または $a=5(\text{mod } 8)$
2. X_0 が奇数。◆

4.4.2.16 例

$M=16$ 、 $a=3$ 、 $c=0$ 、 $X_0=1$ のとき、数列は{1,3,9,11}であり、周期は $16/4=4$ である。◆

以上は線形合同法における可能な最長周期と、その最長周期を得るための条件に関する定理である。これらの定理によれば、法 M として素数を選択するとその他のパラメータの選択肢が多そうである。法 M として 2 の累乗を考えるなど、計算機のオペレーション・システムを意識し、計算時間の高速化を目的とした提案もあるが、これらの定理の主な目的は、機械上の制限のもと、如何に長期の周期を得るかである。

しかし、以上の定理は、得られた数列の「ばらつき」に関しては何も保証していない点に注意が必要である。例えば「4.4.2.5 例 1.」の数列は「ばらついている」とはいいがたい。「4.4.2.1 線形合同法の性質のまとめ」でも述べたように、線形合同法による数列を「擬似乱数」として利用する場合には、その「ばらつき」に関して統計的検定による評価を行わなければならない。

実は線形合同法による数列には、その「ばらつき」に関して重大な「欠点と称される点」が指摘されている。これが「多次元結晶構造」といわれるものである。

4.4.3 多次元結晶構造

線形合同法による数列は、以前の数の一次式により、順次生成される。この性質から、線形合同法による数列 $\{X_n\}$ から、相隣り合う 2 数を並べた (X_n, X_{n+1}) を xy 座標軸平面上にプロットすると、プロットされた点が「結晶構造」をなす。例えば $M = 11$ 、 $a = 2$ 、 $c = 0$ 、 $X_0 = 1$ として、一周期中の 10 個の「2 次の連点 (2 Tuples)」、 $(X_0, X_1), (X_1, X_2), \dots, (X_9, X_{10})$ をプロットすると図 4.4.1 のとおりである。

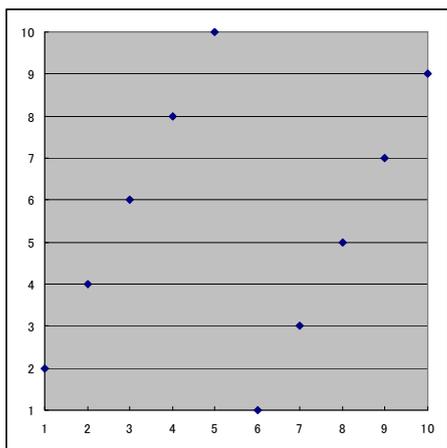


図 4.4.1

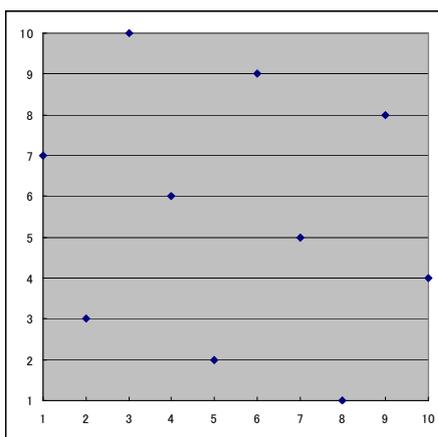


図 4.4.2

プロットされた点が「規則正しく」、「さまざま」な直線上に並んでいるように見える。例えば点 $(1,2)$ と点 $(2,4)$ は直線 $y = 2x$ 上の点であり、この直線に平行な直線群上にその他の点もプロットされている。例えば $(4,8)$ はこの直線上の点であり、 $(7,3)$ 、 $(8,5)$ は、この直線に平行な直線である $y = 2x - 11$ 上にある。また、点 $(1,2)$ と点 $(6,1)$ は直線 $y = (-1/15)x + 11/5$ 上の点であり、その他の点もこの直線に平行な直線群上にある。 $(1,2)$ を通る直線は 6 本見て取れる。

10 の 2 乗個の格子点中に 10 個の点しかなく、 x 座標に重複はなく、 y 座標にも重複がないから、この結晶構造は線形合同法にあっては当然のものではある。また、次のように考えることもできる。すなわち、 x 座標は線形合同法により生成された重複のない点群であるが、法 M による剰余を取ることを行なえば、 y 座標はすべて $y = 2x$ 上の点であり、 M による法を取ることによって、この直線が 2 分割され、 $0 \leq y \leq 10$ の範囲に平行移動した、と見れば、すくなくとも右上がりの平行な傾きが 2 の直線群上にすべての点がプロットされることが理解できる。

(X_0, X_1, X_2) 、 (X_1, X_2, X_3) などの 3 次の連点の場合、および、4 次以上の連点の場合でも同様に、連点が平面 (4 次以上の場合は超平面) 上に並ぶ。

このような結晶構造は、線形合同法では避けられない現象ではあるが、結晶構造のうちでも、プロット平面上により均等に分散した点を与える方法 (パラメータ) が望ましい。例えば図 4.4.2 には、 $M = 11$ 、 $a = 7$ 、 $c = 0$ 、 $X_0 = 1$ として、一周期中の 10 個の点をプロットした。図 4.4.1 と比較して、「より均等にばらついている」と見て取れる。

さらにこういった結晶構造は、線形合同法による数列を最長周期までとったときのものであり、最長周期に至らない時点では、一見、ばらついているように見える。図 4.4.3、図 4.4.4 および図 4.4.5 は、それぞれ、 $M = 256$ 、 $a = 9$ 、 $c = 1$ 、 $X_0 = 1$ の場合の最初の 50 点、100 点および 256 点である。当初のプロットは、一見、ばらついているように見えるが、数列の生成数が多くなるに従い、結晶構造が見えてくる。

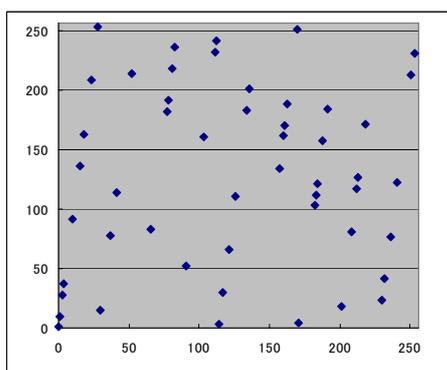


図 4.4.3

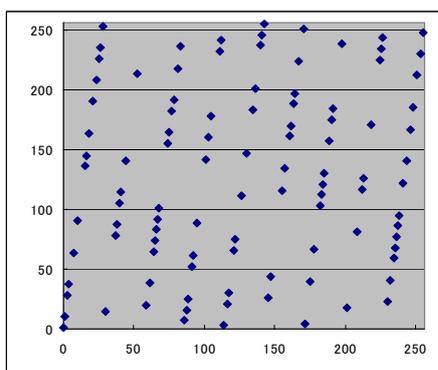


図 4.4.4

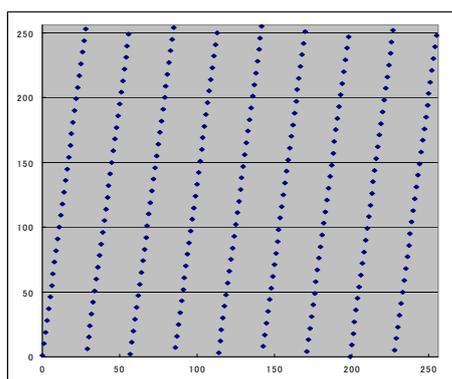


図 4.4.5

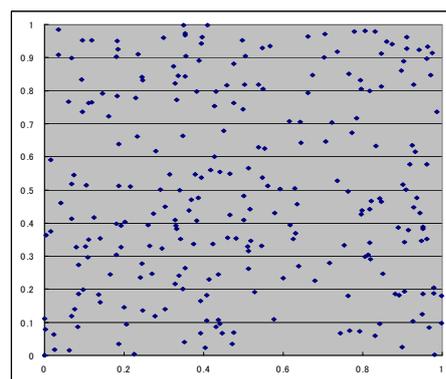


図 4.4.6

図 4.4.6 は、 $M = 2^{32}$ 、 $a = 69069$ 、 $c = 1$ 、 $X_0 = 0$ の場合の最初の 256 点である (座標を M で割り、 $[0, 1) \times [0, 1)$ 区間にプロットした)。全部で約 40 億個の点のうちの 256 個であるから、この時点では、適度に分散しているように見える。同じ 256 個の点であるが図 4.4.5 と比較すると「ばらつき」が異なって見える。この場合でも、生成数が M に近づくと結晶構造が見えてくる。

4.4.4 スペクトル検定

4.4.4.1 スペクトル検定

線形合同法における結晶構造は、その線形性に依存し、避けられない性質ではあるが、結晶構造のうちでも、前述のように、より均等に分布している数列、よってその数列を生成するパラメータ (M, a, c, X_0) を選択したい。その選択の指標として提唱されている方法が「スペクトル検定」である。

この方法は Knuth[1]によれば、「現在のところ最も強力な検証ツール」であり、「カイ2乗検定法およびKS検定法より強力な方法」とされている。

考え方は、結晶構造中の平行超平面群の距離を測定し、その距離の最大値を指標とするものである。たとえば二次元の場合、先に見たように連数のプロットは、いくつかの平行直線群上に位置している。ある平行直線群を見ると、その中の直線群は均等な距離で並んでいる。この距離は、平行直線群間で異なる。この距離の最大値をこの数列のばらつきの指標とする、という考え方である。結局、間隔の最大値が小さい数列に高い評価を与えることになる。この指標は、次元数によって異なる³⁵。

法 M が素数で乗数 a がその法 M の原始根であるとき、次の定理がある。

4.4.4.2 定理

法 M が素数で乗数 a がその法 M の原始根であるとき、乗算合同法 $(c = 0)$ による数列から作られる k 次元連数による超平面群の最大間隔 $d_{\max}(k)$ は次で与えられる。

$$d_{\max}(k) = \frac{M}{|v|_{\min}}$$

ここに、 $v = (v_1, v_2, \dots, v_k)$ はすべてがゼロではない整数ばかりからなるベクトルであり、

$$|v|_{\min} = \min \{ \sqrt{v_1^2 + v_2^2 + \dots + v_k^2} : v_1 + v_2 a + v_3 a^2 + \dots + v_k a^{k-1} \equiv 0 \pmod{M} \}$$

とする。◆

詳細は Knuth[1] (3rd Ed 日本語版 pp85) および伏見[2](pp8)を参照されたい。

³⁵ 本稿では、簡便のため、距離の最大値を指標として紹介しているが、スペクトル検定にあつては、正しくは、距離の最大値の「逆数」を指標とする。この場合指標の大きなものに高い評価を与えることになる。

4.4.4.3 例

$M = 11$ 、 $a = 2$ (原始根)、 $c = 0$ 、 $X_0 = 1$ の場合の $d_{\max}(2)$ を計算する。 $v_1^2 + v_2^2$ の最小値を見つければよいから、 v_1 および v_2 の絶対値の小さいところ、例えば、絶対値が 2 以下のあたりで $v_1 + 2v_2 = 0 \pmod{11}$ を満足する組み合わせはないか、調べる。結果は以下のとおり。

		$v_1 + 2v_2 \pmod{11}$				
$v_2 \setminus v_1$		-2	-1	0	1	2
-2		5	6	7	8	9
-1		7	8	9	10	0
0		9	10	0	1	2
1		0	1	2	3	4
2		2	3	4	5	6

結局 $(0, 0)$ 以外の組み合わせとして $(2, -1)$ 、 $(-2, 1)$ が該当する。いずれの場合も $|v| = \sqrt{5}$ 、よって $d_{\max}(2) = 11/\sqrt{5} = 4.92$ である。図 4.4.7 で確認してみよう。

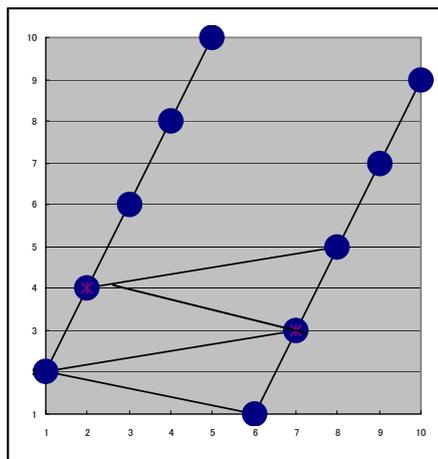


図 4.4.7

2次元にプロットしたものが上図である。いくつかの直線群間の平行線を書き入れてある。見たところ、 $y = 2x$ と $y = 2x - 11$ の直線群の間隔が一番大きそうである。直線 $ax + by = s$ と $ax + by = t$ 間の距離は $|s - t| / \sqrt{a^2 + b^2}$ であるから、この 2 直線間の距離は $11/\sqrt{5}$ である。図でも確認できた。

$M = 11$ 、 $a = 7$ (原始根)、 $c = 0$ 、 $X_0 = 1$ の場合は、 $d_{\max}(2) = 11/\sqrt{10} = 3.48$ となり、乗数が 2 の場合より小さい。よって、スペクトル検定によれば乗数 7 は乗数 2 よりも適切な乗数といえる。これは先に図 4.4.1 および図 4.4.2 を比較した時の分布のイメージに適合する。◆

数学的な定式化が得られたからといって、実際に計算値が求められるものではない。M が非常に

大きな数(例えば 2^{32})の場合、数学的な定式そのままでは、計算時間が膨大にかかり、実際には使えない。そこで数学的な定式をもとに「さまざま」な工夫を加えた算法(プログラム・ロジック)が提唱されている。実際の算法は、Knuth[1]、伏見[2]、Ripley[6]、Fishman[7]を参照されたい。

4.4.4.4 評価

スペクトル検定による線形合同法のパラメータの評価は、例えば、Knuth[1]、伏見[2]、Ripley[6]、Fishman[7]にある。各著者独自の方法を加え評価を行っているので、最も優秀な線形合同法を特定することは困難である。Knuth[1]は、以下の注意を促している。スペクトル検定にも限界があることに注意しなければならない。

1. パラメータの選択は研究の目的によって異なる。
2. 良好なスペクトル検定の結果が得られたとしても、 M が十分大きくない場合は注意を要する。
3. 検定は最良な方式を選定するためのものではなく、使用してはいけない方式を検出する目的で行うべきである。
4. その他の検定もあわせて行い、評価すべきである。

FishmanとMooreは、法 $M=2^{31}-1$ (メルセンヌ素数)である乗算合同法($c=0$)について、2次元から6次元のスペクトル検定を行い、最長周期を与える乗数(a)の評価を行った。 $\varphi(2^{31}-2)=534,600,000$ 個の原始根に対する評価である(「4.4.2.11例」参照)。「超平面の距離の最大値の理論的最小値」の125%以内に収まる距離を算出する乗数を探し出すと、約400個の原始根が合格したという。合格率は $7.5 \times 10^{-5}\%$ である。原始根をでたらめに採用してもとても高性能な乗数にはあたらないことを示している。第一位は950,706,376等であったという³⁶。

³⁶ Fishman[7]およびFishman[20]

4.4.5 M系列乱数

線形合同法による数列が多次元結晶構造を有するのは、その生成方法が次数 1 の漸化式によっているからである。そこで、これを避けるため、高次の漸化式を使って乱数を生成する方法が提案されている。そのうち現在までに最もよく研究されている方法が、「M 系列」を用いて乱数列を構成する方法である。M 系列を用いて構成される乱数を「M 系列乱数」という。M 系列は計算機の論理和、特に排他的論理和の機能を使って生成する。論理和は乗算などより演算処理が高速であるので、上手く設計すれば非常に高速な乱数列の生成算法を組むことができる。また、周期を計算機の 1 語(ワード)長以上とすることもできる。線形合同法では、同じ数値が出現するとそこで一周期となるが、M 系列乱数では必ずしもこのようなことはない。線形合同法によれば、32 ビットの計算機では、一周期の最大長は $2^{32} - 1$ であるが、M 系列乱数では、例えば、 $2^{521} - 1$ 等の周期の数列を設計することができる。また、ある条件を満たせば、高次元で均等分布する(高次元超平面を「埋め尽くす」)数列を構成できる。

ただし、これらの数列が如何に「ばらばらである」かは、別途検証しなければならない。M 系列乱数にあっても如何にばらばらかを理論的に保証するものではないことに注意する必要がある。

4.4.5.1 原始多項式

M 系列では、次の p 次の線形漸化式を考える。

$$x_n = a_1 \cdot x_{n-1} + a_2 \cdot x_{n-2} + \cdots + a_p \cdot x_{n-p} \pmod{2}, a_p = 1 \quad (1)$$

この漸化式は法 2 で考えるので、係数 a_i および変数 x_i は、すべて 0 か 1 である。また、 p 個の初期値 $(x_0, x_1, \dots, x_{p-1})$ を必要とするが、これらは「すべてがゼロ」ではないように選択する。漸化式 (1) で生成される数列(といっても 0 か 1 のいずれかしか取らない) $\{x_i\}$ を考える。この数列の周期は何であろうか。たとえば、次のような数列が得られたとしよう。

011010110101101...

これを

01101 01101 01101...

と区切りこのパターンが続くようであれば周期は 5 である。 $\{x_i\}$ は無限に数値の連続する数列、無限数列であるが、これを p 個ずつ区切って以下のようにならべる。

$(x_0, x_1, \dots, x_{p-1})$
 (x_1, x_2, \dots, x_p)
...

このような数値の組み合わせは、 x_i が 0 か 1 の 2 値しか取らないので 2^p とおりある。いったん以前に出現した数値の組み合わせと同じ組み合わせが生成されれば、そこからは同じ組み合わせが周回する。初期値の p 個をすべてはゼロではないとしているので、すべてゼロの場合を除くと、 $\{x_i\}$ の周期は $2^p - 1$ を超えない。よって次の定理が成立する。

4.4.5.2 定理

漸化式(1)によって生成される数列 $\{x_i\}$ は周期的であり、周期 $T \leq 2^p - 1$ である。◆

ここに次の定理が知られている。詳細は Golomb[23]を参照されたい。

4.4.5.3 定理

漸化式(1)の周期が最大値 $2^p - 1$ を取るための必要十分条件は、漸化式(1)に従って定義される多項式、

$$f(x) = 1 + a_1x + a_2x^2 + \cdots + a_px^p, a_p = 1 \quad (2)$$

が、素数 2 上のガロア体 (GF(2)) 上で原始多項式であることである。◆

この条件が成立するとき数列 $\{x_i\}$ を「 p 次の線形最大周期列 (Maximum-length linearly recurring sequence)」、略して「M 系列」とよび、多項式 $f(x)$ を漸化式(1)あるいは M 系列 $\{x_i\}$ の「特性多項式」と呼ぶ。

M 系列は、フィード・バック・シフトレジスタ (Feedback Shift Register: FSR) 系列とも呼ばれる。この命名は M 系列を計算機の FSR によって生成することができることからきている。FSR は、例えば p 個の記憶領域を持っており、次々と領域内のデータを(例えば右に)移転していくのだが、空いた領域(例えば一番左の領域には)、(1)式などの計算をした結果を入れる装置のことである。

ガロア体と原始多項式について少し復習する。「体」とは加減乗除の定義された集合のことである。GF(2)というのは $\{0, 1\}$ の要素しかない、加減算および乗算が表 4.4.1 のように定義された集合のことである。

表 4.4.1 GF(2)上の加減乗算

+	0	1	-	0	1	×	0	1
0	0	1	0	0	1	0	0	0
1	1	0	1	1	0	1	0	1

GF(2)上では加算と減算の結果が同じであり-1 と+1 の区別がない。

GF(2)上の加算はコンピューター上の「排他的論理和 (EOR)」という機能で計算できる。これは 2 進数の計算を行い、桁上りを見捨てて一桁目のみを取る計算である。

p 次多項式 $f(x)$ が GF(2) 上で「原始多項式」であるとは、GF(2) 上で既約多項式であり、その指数が $2^p - 1$ であることをいう。多項式が「既約多項式」であるとは、係数が定義されている「体」上で根をもたないことをいう。GF(2) 上で既約多項式というのは 0 と 1 を根に持たないことをいう。多項式の「指数」とは、 $x^n - 1$ という形の多項式のうち、 $f(x)$ で「割り切れる」ものの最低の次数 n のことである。GF(2) 上の多項式どおしの演算を行なう際には、係数の演算を GF(2) の上で行なう。例えば、

$$(1 + x + x^3) - (x + x^4) = 1 + x^3 + x^4$$

$$-1 + x^{15} = 1 + x^{15} = (1 + x + x^4)(1 + x + x^2 + x^3 + x^5 + x^7 + x^8 + x^{11})$$

等である。 $f(x) = 1 + x + x^4$ とすると、 $f(0) = f(1) = 1$ であるので GF(2) 上で根を持たず、GF(2) 上で既約多項式である。また $n = 15$ 以下の次数の $1 + x^n (= x^n - 1)$ という形の多項式を割り切らなければ、 $f(x)$ の指数は $15 = 2^4 - 1$ であり原始多項式であるということになる。

原始多項式は計算機を使って膨大な数のものが調べられている。表 4.4.2 にその例を記載した³⁷。 $1 + x^q + x^p$ という形をした原始多項式を M 系列の特性多項式として使用すると、漸化式(1)が 2 項となり計算効率が良いのでよく研究されている。

4.4.5.3.1 M 系列の例

表 4.4.2 から $1 + x^3 + x^5$ は GF(2) 上の原始多項式である。これを特性多項式とする M 系列は漸化

表 4.4.2			
GF(2) 上の原始多項式 (p ごとに一つずつ掲載した。また、 $p \geq 31$ に対しては $2^p - 1$ が素数となっている)			
p	原始多項式	p	原始多項式
1	$1 + x$	9	$1 + x^4 + x^9$
2	$1 + x + x^2$	10	$1 + x^3 + x^{10}$
3	$1 + x + x^3$	31	$1 + x^3 + x^{31}$
4	$1 + x + x^4$	89	$1 + x^{38} + x^{89}$
5	$1 + x^3 + x^5$	127	$1 + x + x^{127}$
6	$1 + x + x^6$	521	$1 + x^{32} + x^{521}$
7	$1 + x + x^7$	2281	$1 + x^{715} + x^{2281}$
8	$1 + x^2 + x^3 + x^4 + x^8$	19937	$1 + x^{881} + x^{19937}$

³⁷ Tezuka, Syu [3], Golomb[23]より。

式

$$x_n = x_{n-3} + x_{n-5} \pmod{2}$$

で与えられる。最初の 5 つ(初期値)を 11111 とすると($x_0 = x_1 = x_2 = x_3 = x_4 = 1$)、生成される系列は次のとおり。

$$\{x_n\} = \{1111100011011101010000100101100\dots\}$$

この系列の周期が $2^5 - 1 = 31$ であることを確認しよう。5 つずつの組を取り、同一の組が出現するまでは一周期である。頭から 5 つずつ数値をとって

$$(x_0, x_1, \dots, x_4)$$

$$(x_1, x_2, \dots, x_5)$$

...

の様に並べたものが表 4.4.3 である。

表 4.4.3					
n	$\{x_n x_{n+1} \dots x_{n+4}\}$	10 進 表記	n	$\{x_n x_{n+1} \dots x_{n+4}\}$	10 進 表記
0	11111	31	16	01000	8
1	11110	30	17	10000	16
2	11100	28	18	00001	1
3	11000	24	19	00010	2
4	10001	17	20	00100	4
5	00011	3	21	01001	9
6	00110	6	22	10010	18
7	01101	13	23	00101	5
8	11011	27	24	01011	11
9	10111	23	25	10110	22
10	01110	14	26	01100	12
11	11101	29	27	11001	25
12	11010	26	28	10011	19
13	10101	21	29	00111	7
14	01010	10	30	01111	15
15	10100	20	31	11111	31

この場合、

$$(1,1,1,1,1)$$

$$(1,1,1,1,0)$$

...

の順に並んでいる。これらは0と1から5つ選んで並べたあらゆる組み合わせのうち、すべてゼロとなるもの以外の31通りをすべて尽くしている。これら5組の数値を2進数表記とみて10進数表記に変換した数値を横に並べた。0を除く1から31までの数値が、 $n=0$ から $n=30$ までのあいだに一通り出現していることがわかる。

このことからM系列は、その初期値を別のもの(例えば11110等)としても、数値が上記のパターンで巡回することがわかる。M系列は初期値に係わらず本質的に同じ系列となる。◆

4.4.5.4 M系列の性質

以上のM系列の性質をまとめる。その他の性質を含め、詳細は伏見[2]およびGolomb[23]を参照されたい。

1. 周期 $T = 2^p - 1$ 。
2. $(x_n, x_{n+1}, \dots, x_{n+p-1})$ 等の p 個組は、一周期において、すべてがゼロである場合を除く、あらゆるパターンが一回ずつ出現する。
3. 特性多項式が同じであれば、初期値が異なることにより異なった出現パターンとはなるものの、生成されるM系列は「ずれ」を除いて一致する。
4. 一周期中に、1は 2^{p-1} 個、0は $2^{p-1} - 1$ 個出現する。0の出現回数は1回少ない。
5. $f(x)$ が p 次の原始多項式ならば、 $f(x^{-1})x^p$ も p 次の原始多項式となることが知られている。よって $f(x) = 1 + x^q + x^p$ が原始多項式なら、 $f(x^{-1})x^p = 1 + x^{p-q} + x^p$ も原始多項式である。

4.4.5.5 M系列乱数

M系列 $\{x_n\}$ から $d (d \geq 2)$ ビット(桁)の2進数系列 $\{W_n\}$ を次のように構成する。

$$W_n = 0.x_{n+\tau_1} x_{n+\tau_2} \cdots x_{n+\tau_d} \text{ (2進表記), } (n = 0, 1, 2, \dots)$$

ここに $\tau_1, \tau_2, \dots, \tau_d$ は互いに異なり n に無関係な定数とする。最初のいくつかを列挙すると以下のとおり。

$$W_0 = 0.x_{\tau_1} x_{\tau_2} \cdots x_{\tau_d}$$

$$W_1 = 0.x_{1+\tau_1} x_{1+\tau_2} \cdots x_{1+\tau_d}$$

$$W_2 = 0.x_{2+\tau_1} x_{2+\tau_2} \cdots x_{2+\tau_d}$$

先の例で見たように数列 $\{x_{\tau_1}, x_{1+\tau_1}, x_{2+\tau_1}, \dots\}$ 、 $\{x_{\tau_2}, x_{1+\tau_2}, x_{2+\tau_2}, \dots\}$ 等は本質的に同じ数列であり、初期値が異なるだけである。数列 $\{x_{n+\tau_d}, x_{n+\tau_d+1}, \dots\}$ は数列 $\{x_n, x_{n+1}, \dots\}$ を τ_d だけ「ずらした」数列である。

このように構成される数列を「M 系列乱数」という。これらが如何に乱数の性質を満たしているかが課題である。

M 系列 $\{x_n\}$ の特性多項式が $f(x) = 1 + x^q + x^p$ という形の場合、この系列の漸化式は、 $x_n = x_{n-q} + x_{n-p}$ という形である。この漸化式に従って W_{n-q} と W_{n-p} のビット(桁)ごとの排他的論理和を計算すれば、 W_n を求めることができる。この「ビット(桁)ごとの排他的論理和」を求めることを演算式 \oplus であらわせば、

$$W_n = W_{n-q} \oplus W_{n-p}$$

と表記できる。2つの数値からそのビットごとの排他的論理和を求めることにより次の数値を求めることになる。コンピュータは排他的論理和を高速に計算するので、この方法によれば数列を高速に生成することができる。

4.4.5.1 M系列乱数の例 1

「4.4.5.3.1 M系列の例」から $d = 2$ ビット(桁)の 2 進数を、 $\tau_1 = 0$ 、 $\tau_2 = 2$ として構成したものが表 4.4.4 である。数列の 10 進表記を見ると、31 個の数値中、0、75、0.5 および 0.25 が 8 回ずつ、0 が 7 回出現していることがわかる。また 10 進表記を 4 倍した数列を見ると、同じく 31 個の数値中、3、2 および 1 が 8 回ずつ、0 が 7 回出現している。

4.4.5.2 M系列乱数の例 2

「4.4.5.3.1 M系列の例」から $d = 5$ ビット(桁)の 2 進数を $\tau_1 = 0$ 、 $\tau_2 = 1$ 、 $\tau_3 = 2$ 、 $\tau_4 = 3$ 、 $\tau_5 = 4$ として構成すると、この例中の表 4.4.3 と同様の結果となる(頭に 0. を加えて、例えば 0.11111 などとするだけである)。10 進表記の列には 1 から 31 までの数値が一通り出現している。ゼロは出現していない。このように構成された M 系列乱数では $1/32$ から $31/32$ まで $1/32$ 刻みで一通りの数値が出現する。

4.4.5.3 M系列乱数の例 3

「4.4.5.3.1 M系列の例」から、今度は $d = 5$ ビット(桁)の 2 進数を $\tau_1 = 0$ 、 $\tau_2 = 1$ 、 $\tau_3 = 2$ 、 $\tau_4 = 3$ 、 $\tau_5 = 20$ として構成したものが表 4.4.5 である。32 倍したのを見ると、一周期中に

3 5 6 8 11 13 14 16 19 21 22 24 27 29 30

の 15 個が 2 回ずつ出現し、0 が一回出現している。

n	$W_n = 0.x_n x_{n+1}$	10進表記	×4	n	$W_n = 0.x_n x_{n+1}$	10進表記	×4
0	0.11	0.75	3	16	0.00	0.00	0
1	0.11	0.75	3	17	0.10	0.50	2
2	0.11	0.75	3	18	0.00	0.00	0
3	0.10	0.50	2	19	0.00	0.00	0
4	0.10	0.50	2	20	0.01	0.25	1
5	0.00	0.00	0	21	0.00	0.00	0
6	0.01	0.25	1	22	0.10	0.5	2
7	0.01	0.25	1	23	0.01	0.25	1
8	0.10	0.50	2	24	0.00	0.00	0
9	0.11	0.75	3	25	0.11	0.75	3
10	0.01	0.25	1	26	0.01	0.25	1
11	0.11	0.75	3	27	0.10	0.50	2
12	0.10	0.50	2	28	0.10	0.50	2
13	0.11	0.75	3	29	0.01	0.25	1
14	0.00	0.00	0	30	0.01	0.25	1
15	0.11	0.75	3	31	0.11	0.75	3

注) 例えば、 $0.11(2進数) = \frac{1}{2} + \frac{1}{2^2}$ (10進数) $= 0.75(10進数)$

n	W_n	10進表記	×32	N	W_n	10進表記	×32
0	0.11110	0.93750	30	16	0.01000	0.25000	8
1	0.11110	0.93750	30	17	0.10000	0.50000	16
2	0.11101	0.90625	29	18	0.00000	0.00000	0
3	0.11000	0.75000	24	19	0.00011	0.09375	3
4	0.10000	0.50000	16	20	0.00101	0.15625	5
5	0.00011	0.09375	3	21	0.01000	0.25000	8
6	0.00110	0.18750	6	22	0.10011	0.59375	19
7	0.01101	0.40625	13	23	0.00101	0.15625	5
8	0.11011	0.84375	27	24	0.01011	0.34375	11
9	0.10110	0.68750	22	25	0.10110	0.68750	22
10	0.01110	0.43750	14	26	0.01101	0.40625	13
11	0.11101	0.90625	29	27	0.11000	0.75000	24
12	0.11011	0.84375	27	28	0.10011	0.59375	19
13	0.10101	0.65625	21	29	0.00110	0.18750	6
14	0.01011	0.34375	11	30	0.01110	0.43750	14
15	0.10101	0.65625	21	31	0.11110	0.93750	30

注) 例えば、 $0.11110(2進数) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}$ (10進数) $= 0.93750(10進数)$

4.4.5.6 高次元均等分布

前述の「4.4.5.5.1 M系列乱数の例 1」では、3 通りの数値、0.25、0.50 および 0.75 が 8 回、0 が 7 回出現している。そもそも 2 進数 2 桁の小数であるので、一番細かい刻みで 1/4 刻みの数値しか出現しないが、これらが一通り出現している。つまり 1/4 刻みで考えて一次元均等分布している。

また、この M 系列乱数の 2 連数を (W_0, W_1) 、 (W_1, W_2) 、 (W_2, W_3) などのようにとり、 xy 平面上にプロットしたものが図 4.4.8 である。4×4=16 個の格子点を埋め尽くし、2 次元でも均等分布することがわかる。

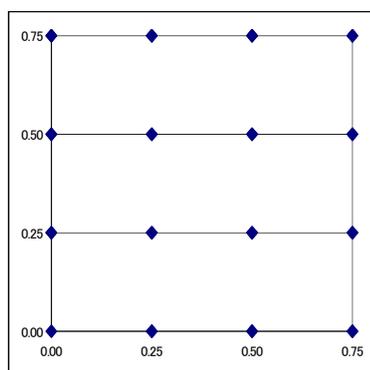


図 4.4.8

同様に、「4.4.5.5.2 M系列乱数の例 2」では、31 通りの数値が 1/32 の刻みで一通り出現し、1次元では均等に分布していることがわかる。この系列の 2 連数を (W_0, W_1) 、 (W_1, W_2) などのように組み合わせ、同じく xy 平面上にプロットしても、プロットする点が高々 31 種類であるので、平面上の 32×32 個の点を埋め尽くすことはできない。つまり 2 次元では均等分布しない。

一方、「4.4.5.5.3 M系列乱数の例 3」では、15 個の数値が 2 度（ゼロは 1 度）出現し、1次元でも均等に分布しない。

この違いはどこからくるのであろうか。実は、M 系列乱数が k 次元均等分布するための必要十分条件が知られている。これが次の「伏見・手塚の定理」である。

4.4.5.6.1 伏見・手塚の定理

p 次特性多項式から d ビットで構成された M 系列乱数列 $\{W_n\}$ が k 次均等分布するための必要十分条件は、次の $(p \times dk)$ 行列における dk 個の p 次列ベクトルが一次独立であることである。

$$\begin{bmatrix} \overrightarrow{W_0} & \overrightarrow{W_1} & \cdots & \overrightarrow{W_{k-1}} \\ \overrightarrow{W_1} & \overrightarrow{W_2} & \cdots & \overrightarrow{W_k} \\ \vdots & \vdots & \cdots & \vdots \\ \overrightarrow{W_{p-1}} & \overrightarrow{W_p} & \cdots & \overrightarrow{W_{p+k-2}} \end{bmatrix}$$

ここに、 $\overrightarrow{W_n}$ は W_n の d 個の構成要素である $x_{n+\tau 1}$ 、 $x_{n+\tau 2} \cdots x_{n+\tau d}$ を並べた次のようなベクトルとする。

$$\overrightarrow{W_n} = (x_{n+\tau 1}, x_{n+\tau 2}, \dots, x_{n+\tau d})$$



すなわち初期値の設定によって均等分布するか否かが決まることになる。証明は、伏見[2]およびFushimi[21]を参照されたい。この定理を前述の例を使って説明する。まず2次元で均等分布した2ビットの例である「4.4.5.5.1 M系列乱数の例 1」では(表 4.4.4 参照)、 $W_0 = 0.11$ 、 $W_1 = 0.11 \cdots W_4 = 0.10$ などであるので $\overrightarrow{W_n}$ は、

$$\overrightarrow{W_0} = (1,1), \overrightarrow{W_1} = (1,1), \overrightarrow{W_2} = (1,1), \overrightarrow{W_3} = (1,0), \overrightarrow{W_4} = (1,0)$$

などのベクトルである。 $\{W_n\}$ は5次特性多項式から構成された2ビットのM系列乱数列である。これが1次均等分布するための必要十分条件は、 (5×2) 行列

$$\begin{bmatrix} \overrightarrow{W_0} \\ \overrightarrow{W_1} \\ \vdots \\ \overrightarrow{W_4} \end{bmatrix}$$

すなわち

$$\begin{bmatrix} 11 \\ 11 \\ 11 \\ 10 \\ 10 \end{bmatrix}$$

の2つの5次列ベクトル

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

が独立であること、2次均等分布するための必要十分条件は、(5×4)行列

$$\begin{bmatrix} \overline{W_0} & \overline{W_1} \\ \overline{W_1} & \overline{W_2} \\ \vdots & \vdots \\ \overline{W_4} & \overline{W_5} \end{bmatrix}$$

すなわち

$$\begin{bmatrix} 1111 \\ 1111 \\ 1110 \\ 1010 \\ 1000 \end{bmatrix}$$

の4つの5次列ベクトル

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

が独立であること、である。いずれも満たされるのでこの例では1次および2次均等分布している。

1次元で均等分布した5ビットの例である「4.4.5.5.2 M系列乱数の例2」の場合(表4.4.3参照)は、(5×5)行列

$$\begin{bmatrix} 11111 \\ 11110 \\ 11100 \\ 11000 \\ 10001 \end{bmatrix}$$

の5つの5次列ベクトルが独立であるので1次均等分布しているが、1次元でも均等分布しなかつ

た 5 ビットの例である「4.4.5.5.3 M 系列乱数の例 3」の場合 (表 4.4.5 参照) は、 (5×5) 行列

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

の 5 つの 5 次元ベクトルが独立ではないので、1 次元均等分布しない。第 5 列 = 第 3 列 - 第 4 列の関係にある。

$(p \times dk)$ 行列にあつて $p < dk$ であると dk 個の p 次元ベクトルは独立ではありえないので、次の補題が成立する。

4.4.5.6.2 補題

p 次元の原始多項式によって生成される M 系列から構成される d ビットの M 系列乱数の均等分布の最大次数は

$$\lfloor p/d \rfloor$$

である。ここに $\lfloor x \rfloor$ は x を超えない最大整数である。◆

4.4.5.6.3 例

2 次元均等分布する 5 ビットの M 系列乱数を作成する。上記の補題から P は 10 以上必要であり、表 4.4.2 から 10 次元原始多項式として $f(x) = 1 + x^3 + x^{10}$ を選択する。 τd 等のずれを $(0, 2, 4, 6, 8)$ とした時の 2 連数のプロットが図 4.4.9、ずれを $(0, 1, 2, 3, 4)$ とした時の 2 連数のプロットが図 4.4.10 である。前者の初期値は、表 4.4.6 であり、「伏見・手塚の定理」における (10×10) 行列は、表 4.4.7 である。この行列の各列は独立である (この場合は正方行列であるので行列式がゼロでないことでも判断できる)。よつて 2 次元均等分布する。図 4.4.9 からも判断できる。

一方、後者の初期値および行列は、それぞれ表 4.4.8、表 4.4.9 となり、各列は独立ではない (例えば第 2 列と第 6 列が同じである)。よつて 2 次元均等分布しない。これは図 4.4.10 からも判断できる。

表 4.4.6

ずれ(0,2,4,6,8)

$\vec{W}_0 =$	0	0	0	0	0
$\vec{W}_1 =$	0	0	0	0	1
$\vec{W}_2 =$	0	0	0	0	0
$\vec{W}_3 =$	0	0	0	1	0
$\vec{W}_4 =$	0	0	0	0	1
$\vec{W}_5 =$	0	0	1	0	0
$\vec{W}_6 =$	0	0	0	1	0
$\vec{W}_7 =$	0	1	0	0	1
$\vec{W}_8 =$	0	0	1	0	0
$\vec{W}_9 =$	1	0	0	1	0

表 4.4.7

ずれ(0,2,4,6,8)

0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	1	0	0
0	0	1	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	1	0	0	1
0	1	0	0	1	0	0	0	1	0	0
0	0	1	0	0	0	1	0	0	1	0
1	0	0	1	0	0	0	1	0	0	1

表 4.4.8

ずれ(0,1,2,3,4)

$\vec{W}_0 =$	0	0	0	0	0
$\vec{W}_1 =$	0	0	0	0	0
$\vec{W}_2 =$	0	0	0	0	0
$\vec{W}_3 =$	0	0	0	0	0
$\vec{W}_4 =$	0	0	0	0	0
$\vec{W}_5 =$	0	0	0	0	1
$\vec{W}_6 =$	0	0	0	1	0
$\vec{W}_7 =$	0	0	1	0	0
$\vec{W}_8 =$	0	1	0	0	1
$\vec{W}_9 =$	1	0	0	1	0

表 4.4.9

ずれ(0,1,2,3,4)

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	1	0
0	0	0	1	0	0	0	0	1	0	0
0	0	1	0	0	0	0	1	0	0	1
0	1	0	0	1	0	0	0	1	0	0
1	0	0	1	0	0	0	0	1	0	0

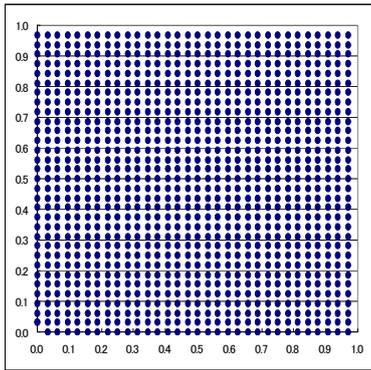


図 4.4.9

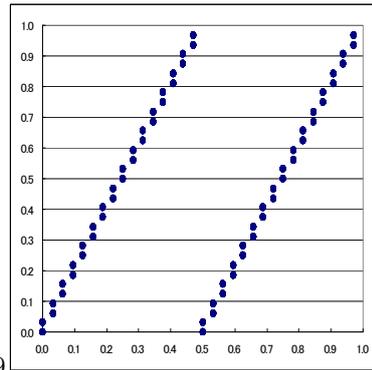


図 4.4.10

4.4.5.7 M系列乱数の性質

1次元均等分布するM系列乱数の性質は以下のとおり。その他の性質を含め、詳細は伏見[2]およびGolomb[23]を参照されたい。

1. $d = p$ の場合 (特性多項式の次数と同じビット(桁)を取る場合)、数列 $\{2^p \cdot W_n\}$ 中の数値 (整数になっている) の出現頻度を数える。一周期中、1 から $2^p - 1$ までの数値が 1 回ずつ出現しゼロは出現しない。
2. $d < p$ の場合 (特性多項式の次数未満のビット(桁)を取る場合)、同じく、数列 $\{2^d \cdot W_n\}$ 中の数値 (整数になっている) の出現頻度を数える。一周期中、1 から $2^d - 1$ までの数値が 2^{p-d} 回出現し、ゼロは $2^{p-d} - 1$ 回出現する (一回少ない)。

これは、 $d = p$ の場合には 1 から $2^p - 1$ までの数値が一回ずつ出現するが、はじめの (左の) d 桁のみを取ると、1 から $2^d - 1$ までの数値が 2^{p-d} 回出現するからである。残りはゼロであるが、 $(2^p - 1) - (2^d - 1)(2^{p-d}) = 2^{p-d} - 1$ 回出現する。

3. 数列 $\{W_n\}$ の一周期の平均値は、一様分布の平均値である 0.5 にほぼ等しい。一周期の平均を \bar{W} 、 $T = 2^p - 1$ を周期とすると、

$$\bar{W} = \frac{1}{T} \sum_{i=0}^{T-1} W_i = \frac{1}{T2^d} \sum_{i=0}^{T-1} (2^d \cdot W_i)$$

である。ところが、上記 1. および 2. の性質から、

$$\begin{aligned}
\bar{W} &= \frac{1}{(2^p - 1)2^d} \left\{ 2^{p-d} \sum_{i=1}^{2^d-1} i + (2^{p-d} - 1) \cdot 0 \right\} \\
&= \frac{2^{p-d}}{(2^p - 1)2^d} \frac{1}{2} (2^d - 1)(2^d) \\
&\approx \frac{1}{2}
\end{aligned}$$

となる。 p および d が十分大きいとき $2^p - 1 \approx 2^p$ 、 $2^d - 1 \approx 2^d$ であることを使った。

4. 数列 $\{W_n\}$ の一周期の分散も、おおよそ一様分布の分散の $1/12$ である。実際に、分散を σ^2 とすると、

$$\begin{aligned}
\sigma^2 &= \frac{1}{T} \sum_{i=0}^{T-1} W_i^2 - \bar{W}^2 \\
&= \frac{1}{T(2^d)^2} \sum_{i=0}^{T-1} (2^d \cdot W_i)^2 - \bar{W}^2 \\
&= \frac{1}{T(2^d)^2} (2^{p-d}) \sum_{i=1}^{2^d-1} (i)^2 - \bar{W}^2
\end{aligned}$$

であり、

$$\begin{aligned}
&= \frac{2^{p-d}}{(2^p - 1)(2^d)^2} \frac{1}{6} (2^d - 1)(2^d) \{2(2^d - 1) + 1\} - \bar{W}^2 \\
&\approx \frac{2^{p-d}}{(2^p)(2^d)^2} \frac{1}{6} (2^d)(2^d) \{2(2^d)\} - \left(\frac{1}{2}\right)^2 \\
&= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}
\end{aligned}$$

である。ここでも、 p および d が十分大きいとき $2^p - 1 \approx 2^p$ 、 $2^d - 1 \approx 2^d$ であることを使った。



4.4.6 メルセンヌ・ツイスター

4.4.6.1 概要

Matsumoto-Nishimura[22]で提案された MT19937 と呼ばれる GF(2)上の M 系列乱数について概要を述べる。詳細は原文を参照していただきたい。MT19937 は $2^{19937} - 1$ という非常に長い周期を持つ M 系列乱数である。この数が素数(メルセンヌ素数)であるところからメルセンヌ・ツイスター(Mersenne Twister)と呼ばれている。

MT19937 は 32 ビットの乱数として使えば $\lfloor 19937/32 \rfloor = 623$ 次均等分布が理論的に保証される。また、19937 次 137 項の原始多項式によって生成される。項数が非常に多いにもかかわらず、実装では 3 項漸化式とほぼ同様の計算手順で構築できる。よって生成速度も他の 3 項漸化式による M 系列乱数と大きく異ならない。

MT19937 のポイントは $19937 = 623 \times 32 + 1$ と分解でき、19937 がメルセンヌ指数となっているところである。

まず、1 ワードを $d = 32$ ビット(桁)とし $\vec{W}_k (k = 0, 1, \dots)$ を各ビットに 0 か 1 を要素にもつ 32 次元(横)ベクトルとする。これを 623 個横につなぎ、右端に 1 ビットの「端数ビット」を更につないで、 $19937 = 623 \times 32 + 1$ ビットの(横)ベクトル $\vec{X}_k (k = 0, 1, \dots)$ を次のように構築する。

$$\begin{aligned}\vec{X}_0 &= (\vec{W}_{623}, \vec{W}_{622}, \dots, \vec{W}_1, \vec{W}_0^u) \\ \vec{X}_1 &= (\vec{W}_{624}, \vec{W}_{623}, \dots, \vec{W}_2, \vec{W}_1^u) \\ &\vdots \\ \vec{X}_k &= (\vec{W}_{k+623}, \vec{W}_{k+622}, \dots, \vec{W}_{k+1}, \vec{W}_k^u)\end{aligned}$$

ここに \vec{W}_k^u は \vec{W}_k の左端の 1 ビットのみを取った(1ビット)ベクトルとする。また \vec{X}_0 はゼロベクトルではないとする。このとき $\vec{W}_{k+624}, (k = 0, 1, \dots)$ を

$$\vec{W}_{k+624} = \vec{W}_{k+397} \oplus (\vec{W}_k^u \mid \vec{W}_{k+1}^l)A$$

によって生成する。ここに、 \vec{W}_k^u は \vec{W}_k の左端の 1 ビット、 \vec{W}_{k+1}^l は \vec{W}_{k+1} の右端の 31 ビットを指す。また、 $(\vec{W}_k^u \mid \vec{W}_{k+1}^l)$ は両者をつないで 32 ビットのベクトルとしたものを指す。

A は GF(2)上の (32×32) 行列で、

$$A = \begin{bmatrix} 0 & 0 & & & \\ & 1 & & & \\ & 0 & \ddots & & \\ & & & 1 & \\ a_{31} & a_{30} & \cdots & a_0 & \end{bmatrix}$$

とし、 $a_{31}, a_{30}, \dots, a_0$ (0 か 1 をとる係数)は、それぞれ順に、

1001 1001 0000 1000 1011 0000 1101 1111

とする³⁸。このとき \overline{X}_k の (19937 × 19937) 遷移行列を B 、すなわち、

$$\overline{X}_{k+1} = \overline{X}_k B, \quad (k = 0, 1, \dots)$$

とすると B は正則行列であり、

$$B = \begin{bmatrix} 0 & I_w & 0 & 0 \\ 0 & 0 & I_w & 0 \\ \vdots & & & \\ 0 & & & \\ I_w & & & \\ 0 & & & \\ \vdots & & \ddots & \\ 0 & & 0 & I_w & 0 \\ 0 & & 0 & 0 & I_{w-r} \\ S & & 0 & 0 & 0 \end{bmatrix}, S = \begin{bmatrix} 0 & I_r \\ I_{w-r} & 0 \end{bmatrix} A$$

となる。ここに、 $w = 32, r = 31$ であり I_w, I_r, I_{w-r} はそれぞれ 32、31、1 次元の正方単位行列とする。

このとき、この行列の固有多項式

$$\begin{aligned} f(x) &= |x \cdot I_{19937} - B| \\ &= (x^{624} + x^{397})(x^{623} + x^{396})^{31} \\ &\quad + a_0(x^{624} + x^{397})(x^{623} + x^{396})^{30} \\ &\quad \vdots \\ &\quad + a_{29}(x^{624} + x^{397})(x^{623} + x^{396})^1 \\ &\quad + a_{30}(x^{624} + x^{397}) \\ &\quad + a_{31} \end{aligned}$$

³⁸ 4 ビットずつ区切ってあるのは読みやすくするためだけである。ちなみに 16 進表記では 9908B0DF である。

は GF(2)上の 19937 次(137 項)原始多項式であることが実証できる。すると、任意の $k(=0,1,\dots)$ に
 対して、 (19937×19937) 行列

$$\begin{bmatrix} \overline{X}_k \\ \overline{X}_{k+1} \\ \vdots \\ \overline{X}_{k+19936} \end{bmatrix}$$

は正則となり、「伏見・手塚の定理」から、 \overline{X}_k の一部の 32 ビットの系列、例えば、 $\{\overline{W}_{k+1}\}$
 $(k=0,1,\dots)$ は $\lfloor 19937/32 \rfloor = 623$ 次均等分布する。

MT19937 では、これらをC言語で実装している。また、均等分布の精度を上げるため、 $\{\overline{W}_{k+1}\}$ に
 更にある (32×32) 行列をかけて出力している(Tempering と呼ばれる)。

4.4.6.2 要点

1. 32 ビットの乱数列を用いる場合は理論的に 623 次均等分布が保証される。
2. 部分列に対する均等分布は検定しなければならない。
3. ある程度の検定はパスしているが、目的に応じて各自検定することが必要であることは、線形合同法、M 系列乱数と異なる。
4. Matsumoto-Nishimura[22]の論文中に例示されている算法では 624 個の初期値を、

$$\overline{W}_{n+1} = 69069 \cdot \overline{W}_n \pmod{2^{32}}$$

で生成している。当初の 624 個はこの線形合同法による乱数列であることに注意されたい。
 また、この場合、初期値 \overline{W}_0 としてゼロは取れない。

4.5 参考文献

乱数関連

- [1] Knuth, D. E. (1981): "The Art of Computer Programming. Vol. 2: Seminumerical Algorithms, 2nd Ed.", Addison-Wesley「準数値算法／乱数」渋谷政昭(訳),サイエンス社、または、
Knuth, D. E. (1998): "The Art of Computer Programming. Vol. 2: Seminumerical Algorithms, 3rd Ed.", Addison-Wesley「日本語版」有澤誠/和田英一(監訳),ASCII.
(一様乱数の基本書。非常に読みやすい。線形合同法、検定、特にスペクトル検定が詳しい。)
- [2] 伏見正則(1989):「乱数」, 東京大学出版会
(M 系列乱数、各種検定の概要がわかりやすい。各種の分布に従う乱数の生成も一通り網羅。M 系列乱数の理論背景となる代数学の知識を前提としているので、別途補わなければならない。)
- [3] Tezuka, Syu(1995):"Uniform Random numbers: Theory and Practice", Kluwer Academic publishers.
(M 系列乱数までの一通りの理論を網羅。)

モンテカルロ法一般

- [4] Sheldon M. Ross(2001): "Simulation 3rd Ed", ACADEMIC PRESS
(SOA の教科書。各種乱数の生成、分散減少法、検定を網羅。英文は読みやすい。但し一様乱数の生成方法にはふれていない。)
- [5] Sheldon M. Ross(2002): "Introduction to Probability Models 8th Ed", ACADEMIC PRESS
(SOA の教科書。第 11 章がシミュレーション。各種乱数の生成方法に関しては[4]より事例が豊富であるが、その他の部分は[4]の集約版といった記述である。)
- [6] Brian D. Ripley(1987): "Stochastic simulation", WILEY
(一様乱数の生成、各種乱数の生成、分散減少法などを網羅。各種乱数の計算アルゴリズムが豊富であり、各々の効率性を比較している。専門的。)
- [7] George S. Fishman(1996): "Monte carlo: Concepts, Algorithms, and Application", Springer
(モンテカルロ関連の集大成。700 ページの大書である。アルゴリズムが非常に豊富である。乱数の検証結果が参考になる。専門的。)

シミュレーション一般

- [8] 津田孝夫(1995):「モンテカルロ法とシミュレーション 電子計算機の確率論的応用(三訂版)」, 倍風館
(M 系列乱数の具体的計算方法が参考になる。工学的諸問題、物理現象への応用などが主題。)
- [9] 関根智明ほか著(1976):「シミュレーション」, 日科技連
(シミュレーション関連の全般。オペレーション・リサーチへの応用が主題である。)

金融工学への応用

- [10] 湯前祥二, 鈴木輝好(2000):「モンテカルロ法の金融工学への応用」, 朝倉書店
(金融工学への応用を主題としている。アメリカン・オプションの価格をモンテカルロ法により求める手法を紹介している。)

代数学/整数論

一様乱数の生成では体論(特に有限体論)が基礎になっている。以下はいずれも代数学の入門書である。[11]は読みやすい。[12]はコンパクトにまとまっている。[14]は読み物としても読める古典的名著。[13][15]も古典的名著。代数学では[11][13]、整数論では[14]が入門書として適していると思われる。他はやや専門的である。

- [11] 渡辺敬一, 草場公邦(1994):「代数の世界」, 朝倉書店すうがくぶっくす 13
- [12] 石田信(1978):「代数学入門」, 実教出版株式会社
- [13] 松坂和夫(1976):「代数系入門」, 岩波書店
- [14] 遠山啓(1972):「初等整数論:日評数学選書」, 日本評論社
- [15] 高木貞治(1971):「初等整数論講義 第2版」, 共立出版株式会社

SOAの過去の問題

- [16] SOA cause 3 examination, from 2000 to 2002.

確率・統計関連

- [17] 国沢清典編:「確率統計演習2統計」, 倍風館
- [18] A.M. Mood: “Introduction to the Theory of Statistics(3rd ed)”, McGraw-Hill

数値表

- [19] Abramowitz, M. and Stegun, I. A. (Eds.) (1970): “*Hand book of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*”, Dover Publication., New York.

専門的文献

- [20] G. S. Fishman and L. R. Moore: “An Exhaustive Analysis of Multiplicative Congruential Random Number Generators with Modulus $2^{31}-1$ ”, SIAM Journal on Scientific and Statistical Computing 7(1986),24-45.
- [21] M. Fushimi and S. Tezuka: “The k-distribution of Generalized Feedback Shift Register Pseudorandom Numbers”, Communications of the ACM 26(1983) 516-523.
- [22] M. Matsumoto and T. Nishimura: “Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudorandom Number Generator”, ACM Transactions on Modelling and Computer Simulation 4 (1994), 254-266.

フィードバック・シフト・レジスター

- [23] S. W. Golomb: “*Shift Register Sequence*”, Holden-Day (1967)

第5章 線形計画法

本章では線形計画法の初歩を紹介する。限定された内容であるが初等的な線形計画問題を具体的に解く方法を習得することを企図したものである¹。これが、より高度な知識習得の契機となることを期待している。

5.1 定式化・例と本章の問題の範囲

5.1.1 定式化

「ある制約された範囲内で目的を達成するための適切な方法を求めよ」²という形の問題を「最適化問題」(または「最適問題」)という。

このとき、目的達成が一定の関数で表され、一定の制約条件のもとでその関数の最適化(大体の場合は当該関数の「最大値」または「最小値」をもって最適化とする)によって問題の解決がなされる場合、その関数を「目的関数」という。そして、制約条件と目的関数が線形(一次式)で表されるものを「線形計画問題」という。本章では最適化が目的関数の最大値によって達成できるとする「最大化問題」のみを扱う³。

最大化問題としての線形計画問題の一般的な形は次のようにあらわすことが出来る。

問題の一般的な形

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 \\ \dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m \\ x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 \end{array} \right.$$

の制約条件のもとに、目的関数

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

を最大値にする x_1, x_2, \dots, x_n を求め、そのときの目的関数の値を求めよ。

この場合、制約条件は

$$\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 \\ \dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m \\ x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 \end{array}$$

のような不等式によって与えられ、

目的関数は $c_1x_1 + c_2x_2 + \cdots + c_nx_n$ である。

¹ 本章では線形代数の初歩を既知としている。

² 井上(2001)85 頁または杉山昌平「最適問題」(共立出版, 1967年)「まえがき」。

³ 最小化問題としては例えば古林(1980)。

なお本章ではこのような最大化問題を、

最大化	$z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$
条件	$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1$
	$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2$

	$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m$
	$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$

のように書くことにする。

また、制約条件をみたす変数の値を「実行可能解」といい、制約条件を満たす領域を「実行可能領域」、目的関数を最大にする実行可能解を「最適解」という。

5.1.2 例

次のような問題を考える。

x と y はいずれも非負であるとする。このとき、 $x + 4y \leq 50$ と $4x + y \leq 50$ に囲まれた領域で $x + y$ を最大にしたい、どのようにしたらよいだろうか？

これは、

最大化	$z = x + y$
条件	$x + 4y \leq 50$
	$4x + y \leq 50$
	$x \geq 0, y \geq 0$

と表すことができる。

この場合、目的関数は $z = x + y$ 、

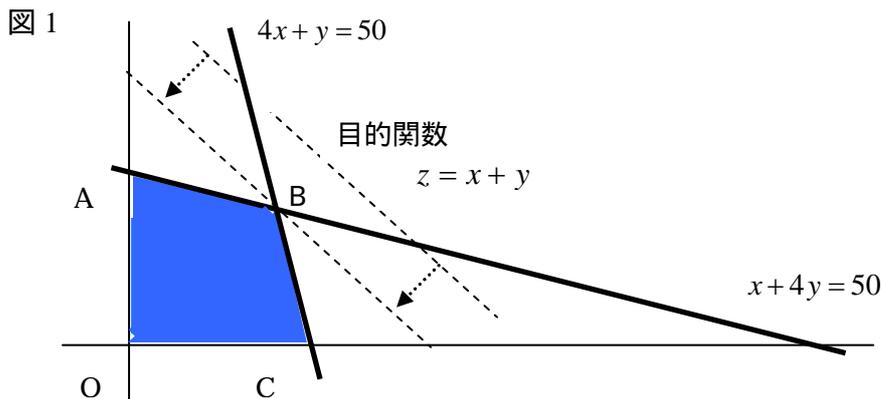
制約条件は

$$\begin{aligned}x + 4y &\leq 50 \\4x + y &\leq 50 \\x \geq 0, y &\geq 0\end{aligned}$$

である。

この問題を図解によって解いてみよう。

図1の領域 OABC は、 $x + 4y \leq 50$ と $4x + y \leq 50$ と変数の非負条件からなる部分であり「実行可能領域」である。さて、この領域の点(実行可能解)のどれを採用したら $z = x + y$ は最大になるのであろうか？図を見る限り答えは点 B(10,10) のところであらう。

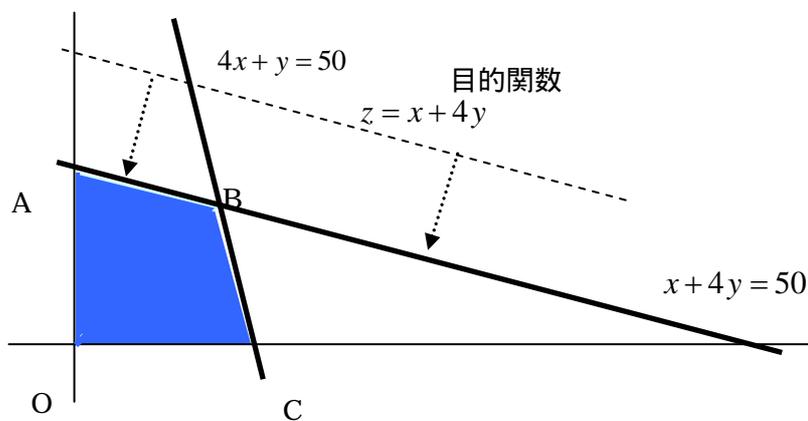


このような形状⁴の実行可能領域にあって、線形関数（図 1 の点線で示す）が最大値を取るといった場合、最大値はこの領域の頂点のどれか（この場合であれば点 B）である。

しかし、目的関数の形状いかんでは、頂点ではなく辺（この場合であれば辺 AB、辺 BC、辺 OA、辺 OC のどれか）で取ることも考えられる。

頂点以外で最大値を取る場合は、図 2 を参照されたい。これは制約条件は例 1 と同じとして目的関数を $z = x + 4y$ とした場合である。この場合は線分 AB 上のどここの部分でも最大値をとることがわかる。

図 2



5.1.3 本章における問題の限定化（1）

さて、このような観測のもと、本章では次のように問題を限定する。

「目的関数を最大化させる最適解は実行可能領域のいずれかの頂点のみである」

したがって、図 2 のように頂点だけでなく辺全体で最大値を取る場合はここでは問題にしないこととする。更に、本章では「変数は常に非負」であることは明示のあるなしに関わらず常に仮定されているものとする。

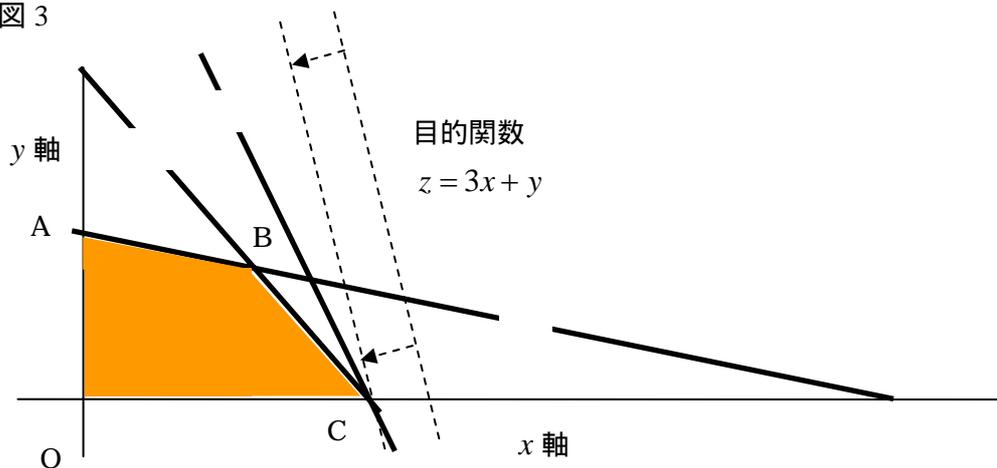
また、退化という現象は起きていないものとする。ここでは次の退化の例を挙げるにと

⁴ 凸集合という。この定義は後に述べる。

どめる。

$$\begin{array}{ll} \text{最大化} & z = 3x + y \\ \text{条件} & x + 4y \leq 60 \quad : \\ & 2x + y \leq 60 \quad : \\ & x + y \leq 30 \quad : \\ & x \geq 0, y \geq 0 \end{array}$$

図 3



交点の座標は $A = (0, 15)$ 、 $B = (20, 10)$ 、 $C = (30, 0)$ である。最大化問題としては座標 C で目的関数は値 90 を取るが、このとき、座標 C をどのように解釈するかといえば、次の三通りが考えられる。

- 1) x 軸と の交点
- 2) x 軸と の交点
- 3) と の交点

しかし、1)以外は実行可能領域からの解釈とは言えない。何となれば、 は実行可能領域には関与しない式だからである。このような状態を退化しているといい、問題によっては最適解に行き着かず巡回してしまうことも皆無とはいえない（その際の解法は別にあるがここでは触れない）。そこで、本章ではこのような状態は最初から排除されているものとする。

5.2 標準形の線形計画問題

不等式 $x + y \leq z$ はある別の変数 $t (\geq 0)$ を介在させることによって、 $x + y + t = z$ と書くことができる。この考え方をういて、線形計画問題を「標準形の線形計画問題」に変換することを考えよう。標準形の線形計画問題とは不等式条件で与えられる線形計画問題を等式による条件に変えることである。

線形計画問題は

$$\begin{aligned}
 \text{最大化} \quad & z = c_1x_1 + c_2x_2 + \cdots + c_nx_n \\
 \text{条件} \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 \\
 & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 \\
 & \cdots \cdots \cdots \\
 & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m \\
 & x_1 \geq 0, x_2 \geq 0, \cdots, x_n \geq 0
 \end{aligned}$$

で表されるが、ここで新たに、 $x_{n+1} \geq 0, x_{n+2} \geq 0, \cdots, x_{n+m} \geq 0$ を次のように加えて、「不等式による制約条件」を「等式による制約条件」に変換する。

最大化	$z = c_1x_1 + c_2x_2 + \cdots + c_nx_n + 0x_{n+1} + 0x_{n+2} + \cdots + 0x_{n+m}$
条件	$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + x_{n+1} = b_1$
	$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + x_{n+2} = b_2$
	$\cdots \cdots \cdots$
	$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + x_{n+m} = b_m$
	$x_1 \geq 0, x_2 \geq 0, \cdots, x_n \geq 0$
	$x_{n+1} \geq 0, x_{n+2} \geq 0, \cdots, x_{n+m} \geq 0$

これを「標準形の線形計画問題」という。これによって不等式による不便さが解消され、線形代数の豊富な結果が利用可能となる。

このようにして付加された変数 $x_{n+1} \geq 0, x_{n+2} \geq 0, \cdots, x_{n+m} \geq 0$ を「スラック変数」(slack) という。このスラック変数に対しても目的関数の係数を 0 として与えると、目的関数はスラック変数に対しても拡大して適用することができる、従って、目的関数を $z = c_1x_1 + c_2x_2 + \cdots + c_nx_n + 0x_{n+1} + 0x_{n+2} + \cdots + 0x_{n+m}$ としている。

このようにスラック変数を入れた場合であっても、等式条件を満たす $m+n$ 次元の実ベクトル空間の領域を改めて「実行可能領域」と呼び、等式条件を満たす点も改めて「実行可能解」(または「可能解」という。

5.2.1.1 例題

次の線形計画問題を標準形の線形計画問題に変形せよ⁵。

$$\begin{aligned}
 \text{最大化} \quad & z = 3x_1 + 2x_2 + 4x_3 \\
 \text{条件} \quad & x_1 + x_2 + 2x_3 \leq 4 \\
 & 2x_1 + 2x_3 \leq 5 \\
 & 2x_1 + x_2 + 3x_3 \leq 7 \\
 & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0
 \end{aligned}$$

⁵ 今野(1987)23頁。なお、この問題は解くことができ、 $x_1 = 2.5, x_2 = 1.5, x_3 = 0$ のとき最大値 10.5 を得る。

解 スラック変数 $x_4 \geq 0$ 、 $x_5 \geq 0$ 、 $x_6 \geq 0$ を導入して次の形にする。

$$\text{最大化 } z = 3x_1 + 2x_2 + 4x_3 + 0x_4 + 0x_5 + 0x_6$$

$$\text{条件 } x_1 + x_2 + 2x_3 + x_4 = 4$$

$$2x_1 + 2x_3 + x_5 = 5$$

$$2x_1 + x_2 + 3x_3 + x_6 = 7$$

$$x_i \geq 0, i = 1, \dots, 6$$

これは標準形の線形計画問題である。

5.2.1.2 例題

目的関数 $z = 4x_1 + 3x_2 + 5x_3$ を最大化する問題を考える。ただし、 $x_1 + x_2 + x_3 \leq 3$ 、 $x_1 + 2x_2 + 3x_3 \leq 4$ 、各変数は正値であるとする。この問題を標準形の線形計画問題として書き下せ。

解⁶

$$\text{最大化 } z = 4x_1 + 3x_2 + 5x_3 + 0x_4 + 0x_5$$

$$\text{条件 } x_1 + x_2 + x_3 + x_4 = 3$$

$$x_1 + 2x_2 + 3x_3 + x_5 = 4$$

$$x_i \geq 0, i = 1, \dots, 5$$

5.2.2 問題 以下を標準形の線形計画問題に変形せよ⁷。

(1) 最大化 $z = 3x_1 + 2x_2$

条件 $6x_1 + 15x_2 \leq 390$

$$2x_1 + x_2 \leq 40$$

$$x_1 \geq 0, x_2 \geq 0$$

(2) 最大化 $z = 5000x_1 + 4500x_2 + 3000x_3 + 500x_4$

条件 $100x_1 + 150x_2 + 80x_3 \leq 1500$

$$50x_1 + 50x_2 + 30x_3 \leq 650$$

$$5x_1 + 10x_2 + 40x_4 \leq 250$$

$$x_i \geq 0, i = 1, \dots, 4$$

5.3 行列表現

線形計画問題と標準形の線形計画問題を行列によって表現することを考えよう。

⁶ この解は $x_1 = 5/2$ 、 $x_2 = 0$ 、 $x_3 = 1/2$ のとき最大値 $25/2$ をとる。

⁷ この問題の解は、(1)は $x_1 = 35/4$ 、 $x_2 = 45/2$ のとき、最大値 $285/4$ を取る。また、(2)は $x_1 = 13$ 、 $x_2 = 0$ 、 $x_3 = 0$ 、 $x_4 = 37/8$ のとき、最大値 67312.5 を得る。

次の線形計画問題

$$\begin{aligned}
 \text{最大化} \quad & z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \\
 \text{条件} \quad & a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \leq b_1 \\
 & a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n \leq b_2 \\
 & \dots\dots\dots \\
 & a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n \leq b_m \\
 & x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0
 \end{aligned}$$

について、つぎのような行列とベクトルを定める。

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \cdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad (\text{本章ではこれを制約条件行列という})$$

$$\vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

すると、線形計画問題は、

$$\begin{aligned}
 \text{最大化} \quad & z = \vec{c}^T \cdot \vec{x} \\
 \text{条件} \quad & A\vec{x} \leq \vec{b} \\
 & \vec{x} \geq \vec{0}
 \end{aligned}$$

と表現される。

ただし、ここで \vec{c}^T はベクトルの転置をあらわし、ベクトルの不等式は各成分においてその不等号が成り立っているものとする。例えば、 $\begin{pmatrix} x \\ y \end{pmatrix} \geq \begin{pmatrix} a \\ b \end{pmatrix}$ であるとは、 $x \geq a$ 、 $y \geq b$ が成り立っていることをいう。

5.3.1 標準形の線形計画問題

さらに、次に下記のように拡大したベクトルと行列を定義する。

$$\vec{\hat{c}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \vec{\hat{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n+m} \end{pmatrix}$$

$$\begin{aligned}x_1 + 3x_2 + x_3 + x_5 &= 4 \\x_1 + x_2 + 3x_3 + x_6 &= 4 \\x_i &\geq 0, \quad i=1, \dots, 5\end{aligned}$$

であるので、

$$\hat{A} = \begin{pmatrix} 3 & 1 & 1 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 & 1 & 0 \\ 1 & 1 & 3 & 0 & 0 & 1 \end{pmatrix} \text{ とし、 } \vec{b} = \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix} \text{ とすると、}$$

制約条件は、 $\hat{A} \cdot \vec{x} = \vec{b}$ と書くことができ、また変数の非負条件は $\vec{x} \geq 0$ と書ける。

従って、この最大化問題は

$$\begin{aligned}\text{最大化 } z &= {}^t \vec{c} \cdot \vec{x} \\ \text{条件 } \hat{A} \cdot \vec{x} &= \vec{b} \\ \vec{x} &\geq 0\end{aligned}$$

と書ける。

記号に関する注意：以下、 A と \hat{A} 、 \vec{c} と \vec{c} など、文脈から明確な場合がほとんどであるので、特に注意を要する場合を除くほか、これらを区別せずに用いるものとする。

5.4 頂点と拡大した制約条件行列との関係⁹

以下では実行可能領域の幾何学的な性質を幾つか解明する。本章における線形計画問題の解法の中心となるものは、実行可能領域の頂点がどこにあるのか、という問題であり、ある頂点から出発して、順次頂点を移動することによって目的関数の値を改良し最終的に最適解を目指すことである。

5.4.1 定義 凸集合

n 次元実ベクトル空間を \mathbf{R}^n とするとき、 \mathbf{R}^n の部分集合 M が「凸集合」であるとは、 M 内の任意の点 \vec{x} と \vec{y} について、これらの点を線分で結合したときその線分上の任意の点が M に含まれることをいう。即ち、 $t\vec{x} + (1-t)\vec{y}$, ($0 \leq t \leq 1$) によって表現される任意の点が M に含まれることをいう。

5.4.2 命題

標準形の線形計画問題の実行可能領域は \mathbf{R}^n 内の凸集合である。

証明

実行可能領域の任意の二点を \vec{x} 、 \vec{y} とするとき、その間の点は $t\vec{x} + (1-t)\vec{y}$ である。この点について、拡大した制約条件行列 A を作用させると、

⁹ 藤田(2003)5 頁

$$A(t\bar{x} + (1-t)\bar{y})$$

\bar{x} 、 \bar{y} が実行可能解であるので、

$$A(t\bar{x} + (1-t)\bar{y}) = tA\bar{x} + (1-t)A\bar{y} = t\vec{b} + (1-t)\vec{b} = \vec{b}$$

を得るが、これはとりもなおさず、 $t\bar{x} + (1-t)\bar{y}$ が実行可能領域の元であることを示している。したがって、実行可能領域の集合は凸集合である。

5.4.3 定義 凸集合の頂点

凸集合の元 \bar{x} が凸集合の「頂点」であるとは、 \bar{x} 以外の凸集合の元 \bar{y} と \bar{z} ($\bar{y} \neq \bar{z}$) があり、 $\bar{x} = (\bar{y} + \bar{z})/2$ とは書けないことをいう。

5.4.4 定理¹⁰

n 個の m 次元列ベクトル $\vec{a}_1, \vec{a}_2, \vec{a}_3, \dots, \vec{a}_n$ によって、行列 A が $A = (\vec{a}_1 \ \vec{a}_2 \ \vec{a}_3 \ \dots \ \vec{a}_n)$ と表されるとき、 $A \cdot \bar{x} = \vec{b}$ ($\bar{x} \geq 0$) の解 ${}^T \bar{x} = (x_1 \ x_2 \ \dots \ x_n)$ の要素 x_1, x_2, \dots, x_n のうち 0 でないものが $x_{i_1}, x_{i_2}, \dots, x_{i_r}$ であったとする。(これは、 $\vec{b} = x_{i_1} \vec{a}_{i_1} + x_{i_2} \vec{a}_{i_2} + \dots + x_{i_r} \vec{a}_{i_r}$ と書けることを意味している。)

このとき、 $x_{i_1}, x_{i_2}, \dots, x_{i_r}$ を 0 でない成分とする $A \cdot \bar{x} = \vec{b}$ の解ベクトル \bar{x} について、

$$\vec{a}_{i_1}, \vec{a}_{i_2}, \dots, \vec{a}_{i_r} \text{ が一次独立} \Leftrightarrow \bar{x} \text{ が実行可能領域の頂点}$$

つまり、

$$\vec{a}_{i_1}, \vec{a}_{i_2}, \dots, \vec{a}_{i_r} \text{ が一次従属} \Leftrightarrow \bar{x} \text{ が実行可能領域の頂点ではない}$$

である。

証明

実行可能領域の集合を $M = \{ \bar{x} \in \square^n \mid A\bar{x} = \vec{b}, \bar{x} \geq 0 \}$ とする。

ここでは三つのベクトル $\vec{a}_1, \vec{a}_2, \vec{a}_3$ が一次独立で $\sum_{i=1}^3 \alpha_i \vec{a}_i = \vec{b}$ 、 $\alpha_i > 0 \quad i=1,2,3$ であると

き、 $\bar{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ が M の頂点であることを背理法で示そう。

頂点でなければ、 M の異なる二つの元 \bar{y} と \bar{z} で $\bar{x} = (\bar{y} + \bar{z})/2$ と書けるはずである。

\bar{y} と \bar{z} の成分を、

¹⁰ 証明は久志本(1979)79 頁または藤田(2003)7~8 頁によっている。

$$\bar{y} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \bar{z} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ とすると、 } \bar{x} \text{ が頂点ではないので少なくとも一つの成分については}$$

$\beta_i \neq \gamma_i$ でなければならない。

更に、 $\sum_{i=1}^3 \beta_i \bar{a}_i = \bar{b}$ 、 $\sum_{i=1}^3 \gamma_i \bar{a}_i = \bar{b}$ であるので、辺々を引くと、

$$\sum_{i=1}^3 (\beta_i - \gamma_i) \bar{a}_i = \bar{0} \text{ であるが、少なくとも一つの成分について } \beta_i \neq \gamma_i \text{ なのでこれは } \bar{a}_1, \bar{a}_2,$$

\bar{a}_3 が一次独立であることに反する。従って、 \bar{x} は M の頂点である。

$$\text{逆に、 } \bar{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ が } M \text{ の頂点であるが、 } \bar{a}_1, \bar{a}_2, \bar{a}_3 \text{ が一次独立でない場合を考える。一次}$$

独立ではないので、 $\sum_{i=1}^3 \kappa_i \bar{a}_i = \bar{0}$ と書いたとき、少なくとも一つの i について $\kappa_i \neq 0$ となる

はずである。

ここで、二つの異なるベクトルの中点として \bar{x} が書かれることを示す。

$$\sum_{i=1}^3 \alpha_i \bar{a}_i = \bar{b} \text{ と } \sum_{i=1}^3 \kappa_i \bar{a}_i = \bar{0} \text{ であることから、 } \sum_{i=1}^3 (\alpha_i \pm \kappa_i) \bar{a}_i = \bar{b} \text{ であることに注意したい。}$$

また、任意の $\varepsilon > 0$ によっても、 $\sum_{i=1}^3 \varepsilon \kappa_i \bar{a}_i = \bar{0}$ であるから $\sum_{i=1}^3 (\alpha_i \pm \varepsilon \kappa_i) \bar{a}_i = \bar{b}$ である。しか

し、任意の κ に対して $^T (\alpha_1 \pm \varepsilon \kappa_1, \alpha_2 \pm \varepsilon \kappa_2, \alpha_3 \pm \varepsilon \kappa_3, 0, \dots, 0)$ が実行可能解を与えている

とは限らない。なぜならば、実行可能領域の非負性が要請されていることから ε の大きいかんでは M の元になるとは限らないからである。

そこで、 $\alpha_i > 0 (i=1,2,3)$ であることから、 $\varepsilon > 0$ を微小にとることによって $\alpha_1 \pm \varepsilon \kappa_1 > 0$, $\alpha_2 \pm \varepsilon \kappa_2 > 0$, $\alpha_3 \pm \varepsilon \kappa_3 > 0$ とすることができる。そのような ε によって M に含まれる二点 \bar{y} と \bar{z} を次のように定義する。

$$\vec{y} = \begin{pmatrix} \alpha_1 + \varepsilon\kappa_1 \\ \alpha_2 + \varepsilon\kappa_2 \\ \alpha_3 + \varepsilon\kappa_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \vec{z} = \begin{pmatrix} \alpha_1 - \varepsilon\kappa_1 \\ \alpha_2 - \varepsilon\kappa_2 \\ \alpha_3 - \varepsilon\kappa_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

このようにすると、

$$\frac{\vec{y} + \vec{z}}{2} = \frac{\begin{pmatrix} \alpha_1 + \varepsilon\kappa_1 \\ \alpha_2 + \varepsilon\kappa_2 \\ \alpha_3 + \varepsilon\kappa_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \alpha_1 - \varepsilon\kappa_1 \\ \alpha_2 - \varepsilon\kappa_2 \\ \alpha_3 - \varepsilon\kappa_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}{2} = \frac{\begin{pmatrix} 2\alpha_1 \\ 2\alpha_2 \\ 2\alpha_3 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}{2} = \vec{x}$$

となることから、 \vec{x} が頂点ではなくなることからこれは矛盾。従って、 $\vec{a}_1, \vec{a}_2, \vec{a}_3$ は一次独立である。ここでは、三つのベクトルについて証明したが、一般的にも同様に証明される。

この定理から、実行可能領域の頂点が制約条件行列の一次独立な列ベクトルによる \vec{b} が対応していることがわかる。これは、ベクトル空間の基底の変更が頂点の移動に相当することを示唆している。

5.4.4.1 例題

次の形の実行可能領域の頂点を求めてみよう。

$$M = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \left| \begin{array}{l} x_1 + 5x_2 + x_3 = 180 \\ x_1 + x_2 + x_4 = 100 \\ x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0 \end{array} \right. \right\}$$

これらの係数を行列で表現すると、

$$\vec{a}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{a}_2 = \begin{pmatrix} 5 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{a}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{a}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

である。ここで、

$$x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3 + x_4\vec{a}_4 = \begin{pmatrix} 180 \\ 100 \end{pmatrix}$$

を考えることになる。

まず、 $\vec{a}_1, \dots, \vec{a}_4, \begin{pmatrix} 180 \\ 100 \end{pmatrix}$ は二つずつペアにするとそれらは線形独立であり、 $\begin{pmatrix} 180 \\ 100 \end{pmatrix}$ は二次元

ベクトルであるので、 $\vec{a}_1, \dots, \vec{a}_4$ の一つでは $\begin{pmatrix} 180 \\ 100 \end{pmatrix}$ は表現できず、また、三つ以上の $\vec{a}_1, \dots, \vec{a}_4$ では線形独立ではありえない。従って、 $\vec{a}_1, \dots, \vec{a}_4$ のうち二つずつのベクトルによる線形結合による $\begin{pmatrix} 180 \\ 100 \end{pmatrix}$ の表現が頂点を与える。そうすると、可能性があるのは ${}_4C_2 = 6$ 個の組合せとなり、次のような表を作成することができる。

	基底のペア	x_1	x_2	x_3	x_4	非負条件
	$\langle \vec{a}_1, \vec{a}_2 \rangle$	80	20	0	0	
	$\langle \vec{a}_1, \vec{a}_3 \rangle$	100	0	80	0	
	$\langle \vec{a}_1, \vec{a}_4 \rangle$	180	0	0	-80	×
	$\langle \vec{a}_2, \vec{a}_3 \rangle$	0	100	-320	0	×
	$\langle \vec{a}_2, \vec{a}_4 \rangle$	0	36	0	64	
	$\langle \vec{a}_3, \vec{a}_4 \rangle$	0	0	180	100	

変数の非負条件を満たすものは、 $\langle \vec{a}_1, \vec{a}_2 \rangle$ 、 $\langle \vec{a}_1, \vec{a}_3 \rangle$ 、 $\langle \vec{a}_2, \vec{a}_4 \rangle$ 、 $\langle \vec{a}_3, \vec{a}_4 \rangle$ の4つである。

従って、頂点の座標は、

$$P_1 \begin{pmatrix} 80 \\ 20 \\ 0 \\ 0 \end{pmatrix} \quad P_2 \begin{pmatrix} 100 \\ 0 \\ 80 \\ 0 \end{pmatrix} \quad P_3 \begin{pmatrix} 0 \\ 36 \\ 0 \\ 64 \end{pmatrix} \quad P_4 \begin{pmatrix} 0 \\ 0 \\ 180 \\ 100 \end{pmatrix}$$

の4点である。

5.5 シンプレックス基準

5.5.1 例題

シンプレックス基準の一般形は次の節に記述するが、その前に、次の例題を考えてみよう。

$$\begin{aligned} \text{最大化} \quad & z = 6x_1 + 12x_2 \\ \text{条件} \quad & x_1 + 5x_2 \leq 180 \\ & x_1 + x_2 \leq 100 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

これを標準形に直すと次のような形になる。

$$\begin{aligned} \text{最大化} \quad & z = 6x_1 + 12x_2 + 0x_3 + 0x_4 \\ \text{条件} \quad & x_1 + 5x_2 + x_3 = 180 \\ & x_1 + x_2 + x_4 = 100 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0 \end{aligned}$$

この線形計画問題の頂点の座標は実は前節の最後で求めた頂点の座標である。

$$P_1 \begin{pmatrix} 80 \\ 20 \\ 0 \\ 0 \end{pmatrix} \quad P_2 \begin{pmatrix} 100 \\ 0 \\ 80 \\ 0 \end{pmatrix} \quad P_3 \begin{pmatrix} 0 \\ 36 \\ 0 \\ 64 \end{pmatrix} \quad P_4 \begin{pmatrix} 0 \\ 0 \\ 180 \\ 100 \end{pmatrix}$$

さて、これに目的関数を当てはめると、次のようになる。

座 標	$6x_1 + 12x_2 + 0x_3 + 0x_4$
${}^T(80, 20, 0, 0)$	720
${}^T(100, 0, 80, 0)$	600
${}^T(0, 36, 0, 64)$	432
${}^T(0, 0, 180, 100)$	0

従って、 $x_1 = 80$, $x_2 = 20$ のとき 720 が得られ、これが線形計画問題の解答となる。これは立派な解法であり、この方法で解答しても全く問題はない。

しかし、実際的にはこの方法は変数が少なくまた制約条件の式数が少ない場合に限定される。例えば式が 10 で変数が 20 程度になると ${}_{20}C_{10} = 184,756$ の組合せを考えなければならぬし、式が 5 で変数が 10 程度でも ${}_{10}C_5 = 252$ であり人がやろうとすると相当の努力を必要とする。従って、効果的に解を求める方法が必要になる。

5.5.2 シンプレックス基準¹¹

ここでは、 n 変数の線形計画問題で変数の非負条件を除いて m 個の条件式からなるものを考える。この線形計画問題をスラック変数を入れて標準化する。すると、拡大した制約条件行列 A を使って、

$$\begin{array}{ll} \text{最大化} & z = {}^T\vec{c} \cdot \vec{x} \\ \text{条 件} & A\vec{x} = \vec{b} \\ & \vec{x} \geq 0 \end{array}$$

と書くことが出来る。

行列 A を改めて詳細に書けば、

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & 0 & 0 & \cdots & 1 \end{pmatrix}$$

$\leftarrow m \rightarrow$ (top row) ↑ m (right side)

である。このとき、 ${}^T\vec{a}_1 = (a_{11} \ a_{21} \ \cdots \ a_{m1})$ 、 ${}^T\vec{a}_2 = (a_{12} \ a_{22} \ \cdots \ a_{m2})$ などと書くと、 A は列ベクトル $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_{n+m}$ によって、 $A = (\vec{a}_1 \ \vec{a}_2 \ \cdots \ \vec{a}_{n+m})$ と書ける。

¹¹ 井上(2001)96 頁

すると、 $A\vec{x} = \vec{b}$ は、 $\sum_{j=1}^{n+m} \vec{a}_j x_j = \vec{b}$ と書ける。

さて、 $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_{n+m}$ の中から、線形独立な m 個のベクトルをとって、その選ばれたベクトルに列番号 $1, 2, \dots, m$ を改めて振りなおす。このようにしてできた、正方行列を改めて $A_0 = (\vec{a}_1 \ \vec{a}_2 \ \dots \ \vec{a}_m)$ とおく。このような一次独立な基底を選択することはこの議論の中で常に可能である。なんとなれば、 A の $n+1$ 列目以降の m 個の列ベクトルは m 行 m 列の単位行列をなすからである。

すると、この A_0 を使うと、 $\vec{b} = A_0 \vec{x}^{(0)}$ なる表現を得ることができる。

$\vec{b} = A_0 \vec{x}^{(0)}$ (より具体的に表示すると、 $\vec{b} = x_1^{(0)} \vec{a}_1 + x_2^{(0)} \vec{a}_2 + \dots + x_m^{(0)} \vec{a}_m$) を満たす $\vec{x}^{(0)} (\geq \vec{0})$ は一つの実行可能解である。

ここで ${}^T \vec{c}_0 = (c_1, c_2, \dots, c_m)$ とする。ただし、 c_1, c_2, \dots, c_m は A_0 に採用された基底 $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ に対応した変数に対する目的関数の係数の順番に並んでいるものとする。すると、このときの目的関数の値は、 ${}^T \vec{c}_0 \cdot \vec{x}^{(0)}$ である。

次に、基底 $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ のうちの一つのベクトルを非基底 $\vec{a}_j \in \{\vec{a}_{m+1}, \vec{a}_{m+2}, \dots, \vec{a}_{n+m}\}$ で置き換えて目的関数の値が上昇するかどうかを考えてみよう。

この \vec{a}_j は $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ の線形結合で書くことができるので、ある $\vec{u}^{(j)}$ が存在して、 $\vec{a}_j = A_0 \vec{u}^{(j)}$ となる (より具体的に表示すると、 $\vec{a}_j = u_1^{(j)} \vec{a}_1 + u_2^{(j)} \vec{a}_2 + \dots + u_m^{(j)} \vec{a}_m$)。 \vec{a}_j に正のパラメータ θ を乗じると、 $\theta \cdot \vec{a}_j = \theta \cdot A_0 \cdot \vec{u}^{(j)}$ である。これと $\vec{b} = A_0 \vec{x}^{(0)}$ を組み合わせると $\vec{b} - \theta \cdot \vec{a}_j = A_0 \vec{x}^{(0)} - \theta \cdot A_0 \vec{u}^{(j)} = A_0 (\vec{x}^{(0)} - \theta \cdot \vec{u}^{(j)})$ である。

従って、 $\vec{b} = A_0 (\vec{x}^{(0)} - \theta \cdot \vec{u}^{(j)}) + \theta \cdot \vec{a}_j$ であり θ をうまく取れば実行可能解を与える。すなわち、 $\vec{b} = \sum_{k=1}^m (x_k^{(0)} - \theta \cdot u_k^{(j)}) \vec{a}_k + \theta \cdot \vec{a}_j$ であるから、各 \vec{a}_k の係数 $(x_k^{(0)} - \theta \cdot u_k^{(j)})_{(k=1,2,\dots,m)}$ と \vec{a}_j の係数 θ が非負であればこれは実行可能解を与える。

しかし、その実行可能解が最適解であるならば、少なくとも凸集合の頂点でなければならず、これは 5.4.4 の定理よりベクトル $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ と \vec{a}_j が線形独立でなければならないが、ここには $m+1$ 個のベクトルがあることからこれはあり得ない。すなわち、 \vec{a}_j を基底に参加させるのであれば、 $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ のうちどれかを退場させなければならないのである。

さて、目的関数の値を具体的に成分を用いて表現してみよう。 \vec{a}_j に対応する目的関数の係数を c_j とすると、

$$\begin{aligned} & c_1 (x_1^{(0)} - \theta \cdot u_1^{(j)}) + c_2 (x_2^{(0)} - \theta \cdot u_2^{(j)}) + \dots + c_m (x_m^{(0)} - \theta \cdot u_m^{(j)}) + c_j \theta \\ & = c_1 x_1^{(0)} + c_2 x_2^{(0)} + \dots + c_m x_m^{(0)} - \theta (c_1 \cdot u_1^{(j)} + \dots + c_m \cdot u_m^{(j)} - c_j) \end{aligned}$$

ここで、 $z_0 = c_1 x_1^{(0)} + c_2 x_2^{(0)} + \dots + c_m x_m^{(0)}$, $z_j = c_1 \cdot u_1^{(j)} + \dots + c_m \cdot u_m^{(j)}$ とおくと、

目的関数の値は $z_0 - \theta(z_j - c_j)$ と書ける。これをまとめると、

$z_j < c_j$ の場合は、 $-(z_j - c_j) > 0$ 、 $z_0 \leq z_0 - \theta(z_j - c_j)$ であるから、目的関数の値は上昇する。

しかし、 $z_j \geq c_j$ の場合は、 $-(z_j - c_j) \leq 0$ 、 $z_0 \geq z_0 - \theta(z_j - c_j)$ であるから、目的関数の値は上昇しない。

従って、この $z_j - c_j$ の正負によって、新たに選んだベクトル \bar{a}_j を導入して値が上昇するかが判断できるということになる。これを「シンプレックス基準」という。

それでは、ここで、 $-(z_j - c_j) > 0$ であったとしよう。すると、新たなベクトル \bar{a}_j を導入した場合の目的関数の値は $z_0 - \theta(z_j - c_j) \geq z_0$ ということになる。ここで問題となるのが、 θ の範囲である。

実行可能解は $x_1^{(0)} - \theta \cdot u_1^{(j)}$, $x_2^{(0)} - \theta \cdot u_2^{(j)}$, ..., $x_m^{(0)} - \theta \cdot u_m^{(j)}$, θ であり、解が負にはならないことから、

$$x_1^{(0)} - \theta \cdot u_1^{(j)} \geq 0, x_2^{(0)} - \theta \cdot u_2^{(j)} \geq 0, \dots, x_m^{(0)} - \theta \cdot u_m^{(j)} \geq 0, \theta \geq 0$$

という連立した条件式によって、 θ の取りえる範囲が定まる。

すなわち、 $0 \leq \theta \leq \frac{x_k^{(0)}}{u_k^{(j)}} (k=1, \dots, m)$ という条件式が成立し、非負条件からこの内一番条件がきついものが θ の範囲を限定する。

また、基底変更後の目的関数の値は $z_0 - \theta(z_j - c_j)$ であるから、最大化問題の解として

はそのような θ のなかでも最も大きいもの、すなわち $\frac{x_l^{(0)}}{u_l^{(j)}} = \min_{(k=1, \dots, m)} \left\{ \frac{x_k^{(0)}}{u_k^{(j)}} \right\}$ であるとする、

$0 \leq \theta \leq \frac{x_l^{(0)}}{u_l^{(j)}}$ であるが、そのような取りうる θ の最大値 $\theta = \frac{x_l^{(0)}}{u_l^{(j)}}$ を採用することになる。そ

うすると、その際、 $x_1^{(0)} - \theta \cdot u_1^{(j)}$, $x_2^{(0)} - \theta \cdot u_2^{(j)}$, ..., $x_m^{(0)} - \theta \cdot u_m^{(j)}$ のいずれかは 0 となるはずである。

例えば $k=2$ のときの値 $\frac{x_2^{(0)}}{u_2^{(j)}}$ が最も小さかったとして説明を進めると、

$x_2^{(0)} - \theta \cdot u_2^{(j)} = x_2^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_2^{(j)} = 0$ となることから、 \bar{a}_2 の係数が 0 となる。これは基底とし

て \bar{a}_2 が退場し、代わりに $\theta = \frac{x_2^{(0)}}{u_2^{(j)}}$ を係数に引き連れて \bar{a}_j が登場することを示している。

このとき実行可能解は、 $x_1^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_1^{(j)}$, $x_3^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_3^{(j)}$, ..., $x_m^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_m^{(j)}$, $\frac{x_2^{(0)}}{u_2^{(j)}}$ であ

り、このとき \vec{b} のベクトル表示は、

$$\vec{b} = \left(x_1^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_1^{(j)} \right) \vec{a}_1 + \left(x_3^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_3^{(j)} \right) \vec{a}_3 + \dots + \left(x_m^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_m^{(j)} \right) \vec{a}_m + \left(\frac{x_2^{(0)}}{u_2^{(j)}} \right) \vec{a}_j$$

である。

また、これに対応する目的関数の係数は $c_1, c_3, \dots, c_m, c_j$ であることから、目的関数の値は、実行可能解と目的関数の係数を乗じることによって、

$$c_1 \cdot \left(x_1^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_1^{(j)} \right) + c_3 \cdot \left(x_3^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_3^{(j)} \right) + \dots + c_m \cdot \left(x_m^{(0)} - \frac{x_2^{(0)}}{u_2^{(j)}} \cdot u_m^{(j)} \right) + c_j \cdot \frac{x_2^{(0)}}{u_2^{(j)}}$$

となる。

ここでは、 \vec{a}_2 の係数を 0 にすることによって、値の上昇を実現したが、これは、最初の基底 $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m$ から \vec{a}_2 を外して、 \vec{a}_j を導入したと解釈することができる。

さて、次に改めて別のベクトルを入れて、値が上昇するかを吟味することになるが、どのように新たなベクトルを導入しても、 $-(z_j - c_j) > 0$ とはならない場合はシンプレックス基準からそれ以上の基底の変更を行っても目的関数の値が上昇しないことを示しており、そのとき、最適問題の解が与えられていることとなる。

5.5.2.1 例題

この様子を次の例題で実行してみよう

$$\begin{aligned} \text{最大化} \quad & z = 2x_1 + x_2 \\ \text{条件} \quad & x_1 + x_2 \leq 5 \\ & 3x_1 + x_2 \leq 10 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

スラック変数を入れて標準化すると、

$$\begin{aligned} \text{最大化} \quad & z = 2x_1 + x_2 + 0x_3 + 0x_4 \\ \text{条件} \quad & x_1 + x_2 + x_3 = 5 \\ & 3x_1 + x_2 + x_4 = 10 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, \end{aligned}$$

制約条件の係数行列は

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 3 & 1 & 0 & 1 \end{pmatrix}$$

である。

また後々使用するので、 $c_1 = 2, c_2 = 1, c_3 = 0, c_4 = 0$ は確認しておく。

以下、最初の基底の選び方のいかににより解答の筋道が変わるのでその相違を見ながら二つの方法を考えたいと思う。いずれの方法でもよい。

方法

この問題では、解のありかがグラフなどから制約条件式の交点であることが見て取れるので、最初の基底を

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

としてしまうことが考えられる。

すると、

$$A_0 = \begin{pmatrix} 1 & 1 \\ 3 & 1 \end{pmatrix} \text{を得る。}$$

まず、交点を得よう。これは制約条件式を連立方程式と見たときの解そのものである。

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \frac{5}{2} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \frac{5}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

となる。従って、目的関数の値は、

$$z_0 = (2 \ 1) \begin{pmatrix} \frac{5}{2} \\ \frac{5}{2} \end{pmatrix} = 15/2$$

である。このとき、更に、基底を変換して値が上がるかどうかを見なければならぬ。

まず、 $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ が基底としてふさわしいかどうかを判定しよう。

ここで判定する式の形が、 $z_0 - \theta(z_3 - c_3)$ であることを思い出そう。

ちなみに、 $z_3 = (c_1 \ c_2) A_0^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (2 \ 1) \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}$ であること、 $c_3 = 0$ から、

$$z_0 - \theta(z_3 - c_3) = 15/2 - \theta(1/2 - 0) < 15/2$$

従って、 $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ は新たな基底とは採用し得ない。

同様にして $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ も基底とはなり得ないことがわかるので、 $x_1 = \frac{5}{2}$, $x_2 = \frac{5}{2}$ のとき目的関数の

値 $= \frac{15}{2}$ がこの線形計画問題の解である。

方法

方法 のようにあらかじめ解の所在が明らかな場合はそれでよいが、そのような解がわからないとして解を順次探っていこう。

基底 $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ を取ることにする。

採用したベクトルと目的関数の係数の順番を合わせるために、形式的ではあるが

$\bar{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\bar{a}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\bar{a}_3 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\bar{a}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ とおき、 $c_1 = 0$, $c_2 = 0$, $c_3 = 2$, $c_4 = 1$ とおくこと

にする。

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{とおくと、}$$

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 10 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

であることから、これは一つの解をあたえている。

この係数に相当する c_0 については、 ${}^T c_0 = (0, 0)$ であることから、目的関数の値 $z_0 = 0$ となる。

さて、ここで、基底の変更を行うと値が増えるのか減るのか、また増えたとしたら、どのベクトルを採用したら効率が良いのか判断しなければならない。

$\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ を導入すると、

$$z_3 = (0, 0) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 0$$

であり、 $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ に相当する係数は $c_3 = 2$ である。

すると、目的関数の増加の様子は

$$0 - \theta(0 - 2) = 2\theta \text{ である。}$$

同じことを、 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ について行くと、

$$0 - \theta(0 - 1) = \theta$$

となって、 $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ を新たな基底に入れた方が効率がよい。

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

であることから、

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} - \theta \begin{pmatrix} 1 \\ 3 \end{pmatrix} \text{ を計算すると、}$$

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} = (5 - \theta) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (10 - 3\theta) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \theta \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

である。

各係数が正の値をとる範囲を考察すると、

$$0 \leq \theta \leq 5 \quad 0 \leq \theta \leq 10/3$$

のいずれか範囲の小さいほうとなるので、 $\theta = 10/3$ が上限である。

この上限を取ると目的関数は $2\theta = 20/3$ である。

そして、このとき $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ は基底から退場し、新たに $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ が基底となる。

従って、現在の基底は $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ である。

さて、これに更に基底の変更を行って目的関数の値が上昇するか検討しなければならない。

今度検討しなければならない基底は $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ である。

改めて、 $\begin{pmatrix} 5 \\ 10 \end{pmatrix}$ をこの現在の基底を用いて表現すると、

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \frac{5}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{10}{3} \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

であることと、

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{2}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

から、

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \left(\frac{5}{3} - \frac{2}{3}\theta\right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \left(\frac{10}{3} - \frac{1}{3}\theta\right) \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \theta \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

である。これから θ の最大値としては、 $\theta = 5/2$ である。

これにより $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ の係数が 0 となりこれが基底から外れて、 $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ が基底になることがわかる。

すなわち、 $\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \frac{5}{2} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \frac{5}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ となる。

すると、目的関数の値としては、 $2x_1 + x_2 = 2 \cdot \frac{5}{2} + 1 \cdot \frac{5}{2} = \frac{15}{2}$ である。

さて、更に、基底の変更が必要かどうかを検討する。

ここで、 $A_1 = \begin{pmatrix} 1 & 1 \\ 3 & 1 \end{pmatrix}$ とすると、この逆行列は $A_1^{-1} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$ である。

検討すべき基底は $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ と $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ である。

・ $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ の場合

定義により $c_1 = 0$ である。

$z_1 = (2, 1) \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}$ であることから、

$$z_1 - c_1 = z_1 = 1/2 > 0$$

である。このことから、 $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ を導入しても目的関数の値を減少させるだけなので取りえな

い。

・ $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ を導入した場合

定義により $c_2 = 0$ である。

$$z_2 = (2, 1) \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2} \text{ であることから、}$$

$$z_2 - c_2 = z_2 = 1/2 > 0$$

である。このことから、 $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ を導入しても目的関数の値を減少させるだけなのでこれも取

りえない。

これより、これ以上の基底の変更は必要ないことがわかる。

従って、

$$x_1 = \frac{5}{2}, x_2 = \frac{5}{2} \text{ のとき、 } z = \frac{15}{2} \text{ を得る。}$$

5.6 まとめ

線形画問題を解く手順としては。

標準形の線形計画問題に直す。

本章では実行可能領域の頂点で最大値をとることを前提としている。

頂点を移動させることと、制約条件の係数ベクトルを基底とした基底の変更をすることが同値であることを使って基底の変更を行うことを考える。

シンプレックス基準から基底の変更の候補をさがす。このとき、他のどの基底を使っても目的関数の値の上昇が認められないときは、そこで最大化問題は終了している。シンプレックス基準から新たな基底を導入すると目的関数の値が上昇することが認められたならば、その基底を使って(その代わり何らかの基底が退場するはずである)目的関数を上昇させる。

に戻り、このサイクルを繰り返す。

5.7 練習問題

5.7.1 問題

次の線形計画問題を解け。ただし、いずれの問題においても変数は非負であるとする。

$$(1) \quad \begin{array}{ll} \text{最大化} & w = x + y \\ \text{条件} & 3x + y \leq 12 \\ & x + 3y \leq 12 \end{array}$$

$$(2) \quad \begin{array}{ll} \text{最大化} & w = x + 2y \\ \text{条件} & x + y \leq 5 \\ & x + 3y \leq 12 \\ & 3x + y \leq 10 \end{array}$$

$$(3) \quad \begin{array}{ll} \text{最大化} & w = x + y \\ \text{条件} & 4x + y \leq 8 \\ & x + 2y \leq 10 \\ & 10x + y \leq 15 \end{array}$$

$$(4) \quad \begin{array}{ll} \text{最大化} & w = 3x + 2y \\ \text{条件} & 6x + 15y \leq 390 \\ & 2x + y \leq 40 \end{array}$$

$$(5) \quad \begin{array}{ll} \text{最大化} & w = x + 2y + z \\ \text{条件} & x + y + z \leq 18 \\ & x + 3y + z \leq 21 \\ & 3x + y + z \leq 24 \end{array}$$

$$(6) \quad \begin{array}{ll} \text{最大化} & w = 2x + y + z \\ \text{条件} & x + y + z \leq 18 \\ & x + 3y + z \leq 21 \\ & 4x + y + z \leq 24 \end{array}$$

解答

以下解答において、例えば $\left(x = \frac{5}{2}, y = \frac{5}{2}\right) \Rightarrow w = \frac{15}{2}$ と書いた場合は、 $x = \frac{5}{2}$, $y = \frac{5}{2}$ にお

いて最大値 $w = \frac{15}{2}$ を取ることを意味する。

以下解答のみを与える。

$$(1) \quad (x = 3, y = 3) \Rightarrow w = 6$$

$$(2) \quad \left(x = \frac{3}{2}, y = \frac{7}{2}\right) \Rightarrow w = \frac{17}{2}$$

$$(3) \quad \left(x = \frac{6}{7}, y = \frac{32}{7}\right) \Rightarrow w = \frac{38}{7}$$

$$(4) \quad \left(x = \frac{35}{4}, y = \frac{45}{2}\right) \Rightarrow w = \frac{285}{4}$$

$$(5) \quad \left(x = 3, y = \frac{3}{2}, z = \frac{27}{2}\right) \Rightarrow w = \frac{39}{2} \quad \text{又は、} \quad \left(x = 0, y = \frac{3}{2}, z = \frac{33}{2}\right) \Rightarrow w = \frac{39}{2}$$

$$(6) \quad \left(x = 2, y = \frac{3}{2}, z = \frac{29}{2}\right) \Rightarrow w = 20 \quad \text{又は、} \quad (x = 2, y = 0, z = 16) \Rightarrow w = 20$$

5.7.2 問題

本章の全ての例題は解ける形になっている。これらを解いてもらいたい。解答は脚注として記してある。

5.7.3 問題

商品 X と商品 Y は、それぞれ成分 A と B という希少素材から構成された別個の商品である。X と Y それぞれに含まれる A と B の含有量は次の表に示すとおりである。

この希少素材は一日の配給量が限定されている。その様子も次の表に示している。

	商品 X 一個における含有量(g)	商品 Y 一個における含有量(g)	配給量(g)
A	2	12	6000
B	8	3	3000
一個の販売単価	15 円	20 円	

これで一日の売り上げを最大化したい。X と Y の生産数をどのようにすればよいか（解は端数を許すものとする）。

解答

X の生産数を x 、Y の生産数を y とすると線形計画問題としては、

最大化 $15x + 20y$

条件 $2x + 12y \leq 6000$

$8x + 3y \leq 3000$

である。

解答としては、 $x = 200$, $y = 1400/3$ である。このとき、 $37000/3 = 12,333$ を得る。

5.8 参考文献

- 1 久志本茂：最適化問題の基礎，森北出版（1979）
- 2 古林 隆：線形計画法入門，産業図書（1980）
- 3 今野 浩：線形計画法，日科技連（1987）
- 4 片桐 他：行列と線形計算，東京電機大学出版局(1995)
- 5 井上博夫：線形経済学，白桃書房(2001)
- 6 藤田岳彦：ファイナンスの最適化入門，講談社（2003）