GNU R の紹介と時系列解析への適用

総合情報処理センター
丸田　英徳
hmaruta@net.nagasaki-u.ac.jp

1　はじめに

本稿では，統計計算用の言語および実行環境である GNU R [1][2] の紹介と，その簡単な利用例として，1 次元時系列解析への適用について報告します．なお，筆者は統計学・時系列解析プロパーではないため，不明瞭な点や語句の不備等あるかと思いますが，その点ご容赦ください．

2　GNU R について

2.1　GNU R の特徴と利点

GNU R（以下 R）は統計計算用の言語およびその実行環境であり，次のような特徴を持ちます．

- GNU GPL [3] のもとで配布されるフリーソフトであること．
- 最近の研究成果に基づいた実装が（ライブラリなどの形で）比較的早く提供されること†1．
- GNU GPL に基づいたソースコードの他に，R の最新バージョンにこだわらなければ，
  - Windows(95 and later)
  - Mac(OS X, 8.6 to 9.1)
  - Linux（with X Window System)(Debian, Mandrake, Redhat, Vine)

  のバイナリが配布されています†2．

なお本稿執筆時点での R のリリースバージョンは"R version 1.8.1" で，以下で R というときはこのバージョンを指すことにします．

R はかつてベル研究所で開発された S 言語に似ており，仕様・実装に違いはあるものの，高い互換性を持っています．R で提供される統計手法は多岐にわたり，日々新たな実装が追加されていくため，ここですべてを紹介することは不可能です．一部をあげると,「線形・非線形モデル，古典的統計検定，時系列解析，判別分析，クラスタリング」などがあり，またこれらの結果をグラフィカルに表示する機能も優れています†3．また，開発プロジェクトでは，R に関する基本的な使い方から関数拡張の方法，さらには最近発表された論文の R での実装についての報告等，非常に充実したドキュメントも配布されており，GNU ならではの活発な活動が展開されています．同様の機能をもつソフトウェアに，商用で MATLAB ® や S-PLUS ® があり，総合情報処理センターの研究用計算機でご利用されてる方もいらっしゃることかと思いますが，手元のパソコンで手軽にあるいは予備的な実験解析を行う場合などにおいて，R の利用価値は高いのではないかと思います．

---

†1 例えば乱数生成には近年開発され，現存するアルゴリズムの中でも比較的優れているといわれる Mersenne Twister 法がデフォルトとして利用されます．

†2 筆者は，Windows バイナリ（OS:WindowsXP) および Linux バイナリ（OS:GNU/Linux Debian testing + Custom kernel）で動作させています．

†3 その他にもさまざまな統計ライブラリや GUI ライブラリが提供され，完全ではないですが日本語化も有志の方の手で進んでいるようです

2.2 Rの向かない分野

手軽に統計解析のできる環境であるRにも，向かない分野は当然あります．例えば高い精度の要求される数値解析を必要とする分野や，非常に大きい容量のデータを扱うような分野などがそうです．

本報告でも精度にはこだわらずに，Rを利用します．一部で（小数）精度等の違いがありますがRで得られる結果をそのまま利用することにしますのでご注意ください．

3 時系列解析について（概略）

以下では，本報告で用いる時系列解析の手法等について，概略を述べます．詳細については，統計学・時系列解析を取り扱った文献[4-8]等をご覧ください．

3.1 準備

ここでいう時系列解析とは，得られたデータをもとに時系列モデルを構成し，モデルに基づく予測を行うこととします．典型的な時系列解析では，

- 時間領域での分析：異なる時点間における観測値の関連性
- 周波数領域での分析：時系列データの周期的変動

のどちらか，あるいは両方の手法を用いて，時系列データを解析し，その予測モデルを構成します．ここでは，時間領域に注目した時系列解析手法，その中でも古典的な手法であるBox-Jenkins流のアプローチとそのモデル同定における限界，およびモデル同定のための情報量規準[†4]について概観します．なお，取り扱う時系列モデルは主に線形モデルのみとしエルゴード性をもつものとします．[†5].

3.1.1 定常性

まず，定常性という概念を導入します（厳密には弱定常性）．

定義 1 時系列 $\{X_t\}$ が定常であるとは次の *3* つの性質を持つことをいう．

1. 平均が一定：$E(X_t) = \mu = const.$

2. 分散が一定：$Var(X_t) = const., < \infty$

3. 自己共分散が時点間の差（*lag*）のみに依存：$Cov(X_t, X_{t-k}) = \gamma(k),\ for\ \forall k \in Z$

”おおらかに”定常性を換言すれば,「どの時点から観測を開始しても，”適度”な区間のデータが得られれば，時系列モデルを得ることができる」条件と言えばよいでしょうか．[†6]

3.1.2 自己相関関数と偏自己相関関数

定常時系列[†7]の時間領域での分析に使われる最も重要な統計量に自己相関関数と偏自己相関関数があり，それらは次で定義されます．以下では平均 μ の推定量を標本平均 $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} X_t$ とします．

定義 2 自己相関関数 $\rho(k)$ は自己共分散関数を基準化した

$$\rho(k) \stackrel{\mathrm{def}}{=} \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} = \frac{\gamma(k)}{\gamma(0)}$$

[†4]ここでは”規準”という漢字を用いましたが，分野によって（書き手によって？）”基準”を用いることもあるようです（ちなみに英語では”criterion”）

[†5]筆者の力量では広大な非線形モデルの世界を紹介することはできません．

[†6]筆者の素人的見解であり，当然異論のある方もいらっしゃることかと思います．

[†7]以下では特に断らない限り時系列は定常であると仮定します

で与えられ，またその推定量は自己共分散関数の推定量

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k}\{X_t - \hat{\mu}\}\{X_{t+k} - \hat{\mu}\}$$

を用いて，

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}$$

で与えられる．

定義 3 偏自己相関関数 $\Phi(k)$ は，次の方程式（*Yule-Walker* 方程式）：

$$\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(k-1) \\ \rho(1) & \rho(0) & \cdots & \rho(k-2) \\ \ddots & \ddots & \ddots & \ddots \\ \rho(k-1) & \rho(k-2) & \ddots & \rho(0) \end{bmatrix} \begin{bmatrix} \xi_{k1} \\ \xi_{k2} \\ \ddots \\ \xi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \ddots \\ \rho(k) \end{bmatrix}$$

で与えられる ξ_{kk} により，

$$\Phi(k) \stackrel{\mathrm{def}}{=} \xi_{kk}$$

で定義される．

偏自己相関関数は定義だけではわかりにくいのですが，$\Phi(k)$ は，X_t と X_{t+k} との（$X_{t+1},\cdots,X_{t+k-1}$ を" 調節" した上での）直接的影響を表す関数になります．

3.2 線形時系列モデル

ここで，今回取り扱う最も簡単な線形時系列モデルについて紹介したいと思います．ただし，各モデルについて詳細（定常性の条件，反転可能性など）に取り扱うことはせずに，各モデルが時間領域で持つと特徴ならびに予測に関する特徴を，まとめて紹介します．また，モデルを決定するために必要なパラメータ数の決定（モデルの同定）に用いられる情報量規準についても簡単に紹介します．これらの内容をもとに次章において，実際のデータについて R を用いて" 簡単な" 時系列解析を行いたいと思います．

3.2.1 AR（自己回帰，Auto Regressive）モデル

以下で定義される時系列モデルは自己回帰モデル（AR(p) モデル）と呼ばれます．

$$X_t = \sum_{j=1}^{p}\phi_j X_{t-j} + \epsilon_t$$

ここで，ϵ_t は誤差やランダムに変動するノイズを表しているものと考えます．ϵ_t にはホワイトノイズを最低限仮定しますが，一般的には平均 0，分散 σ^2 の独立同一正規過程 $\epsilon_t \sim i.i.d\ N(0,\sigma^2)$ とすることが多いようです．以下，本報告でもこの独立同一正規性を ϵ_t に仮定します．このモデルを端的に言えば，時刻 t における観測値は，p 時点前までの観測値 $\{X_{t-1},\cdots,X_{t-p}\}$ とランダムな変動との線形和で表される，ということになります．

3.2.2 MA（移動平均，Moving Average）モデル

以下で定義される時系列モデルは移動平均（MA(q)）モデルと呼ばれます．

$$X_t = \sum_{j=1}^{q} \pi_j \epsilon_{t-j} + \epsilon_t$$

このモデルを端的に言えば，時刻 t における観測値は，q 時点前までの観測値（ランダム変動，あるいはショックとも呼ばれます）$\{\epsilon_{t-1}, \cdots, \epsilon_{t-q}\}$ と t 時点でのランダムな変動との線形和で表される，ということになります．また通常 q は有限であるため，常に定常性を満たします．本報告ではこれ以上 MA モデルについては述べません．

3.2.3 ARMA（自己回帰移動平均）モデル

以下で定義される時系列モデルは自己回帰移動平均（ARMA(p,q)）モデルと呼ばれます．

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \pi_j \epsilon_{t-j} + \epsilon_t$$

ARMA(p,q) モデルは AR,MA 両モデルの特徴を併せ持ったモデルで，定常性については AR 部分（定義式の第一項）で決定されます．モデルの包含関係を書けば，AR，MA $\subset$ ARMA ということになります．

3.2.4 各モデルの特徴

以下に各モデルの時間領域での特徴，および予測に関する特徴をまとめます．詳細は，時系列関連の書籍をご覧ください．(なお，すべて数式の展開による結果ですが，フォローするのは面倒です)

自己相関・偏自己相関による特徴付け

	AR(p)	MA(q)	ARMA(p,q)
自己相関関数	緩やかに減衰	$q+1$ 以降は 0（切断）	緩やかに減衰
偏自己相関関数	$p+1$ 以降は 0（切断）	緩やかに減衰	緩やかに減衰

予測値の特徴

	AR(p)	MA(q)	ARMA(p,q)
予測値	h とともに徐々に平均に近づく	$q+1$ 期先以降は平均に一致	$q+1$ 期先以降は AR(p) に一致
予測誤差	h とともに徐々に増加	$q+1$ 期先以降は一定	$q+1$ 期先以降は AR(p) に一致

3.2.5 ARIMA（自己回帰和分移動平均，Auto Regressive Integrated Moving Average）モデルについて

階差オペレータ ∇ を，$\nabla X_t = X_t - X_{t-1}$ で定義します．このとき，以下で定義される時系列は自己回帰和分移動平均（ARIMA(p,d,q)）モデルと呼ばれます．

$$\begin{aligned} Y_t &= \nabla^d X_t \\ Y_t &= \sum_{j=1}^{p} \phi_j Y_{t-j} + \sum_{j=1}^{q} \pi_j \epsilon_{t-j} + \epsilon_t \end{aligned}$$

このモデルは適当な階差をとった時系列に対して，ARMA モデルを当てはめたモデルになっています．定義からわかるように ARIMA モデルは AR モデル・MA モデル・ARMA モデルを特別なモデルとして含みます．なお，R によって実際に時系列解析を行う際には，ARIMA モデル用の R library（関数）を利用します．

3.2.6 Box-Jenkins 流アプローチとその限界

ここでいう Box-Jenkins 流アプローチとは以下の流れを指すこととします.

1. 時系列データの定常化（階差をとるなど）

2. モデル同定（ARMA(p,q) モデルでいうなら (p,q) の決定）

3. 係数パラメータの決定（最尤法，最小 2 乗法など）

4. 得られたモデルの診断（残差の独立性などによる検定）

1. の定常化については，詳しく述べませんが，時系列データの実際の面においては定常性の仮定の上に作られた理論に基づくことが少なくありません. よって，この定常化の作業が前処理として重要になります（この際に周波数解析などを用いて季節変動などを解析します）. 手法としては，階差をとる，対数をとるなど様々ですが，その弊害も考慮しなければなりません. 定常化されたデータについては 2, 3, 4 を順に処理していくわけですが，重要なことはモデルの同定の際に先にあげた各モデルの特徴だけからでは判断できないことがあるということです. 特に ARMA モデルについては，(p,q) の選択になんら役に立たないことも考えられます.

3.3 赤池情報量規準

赤池情報量規準 [9–11] はある統計的モデル族からデータに基づいてモデルを選択する際にモデルの悪さを評価する情報量規準で，以下で定義されます.

定義 4 AIC = -2(最大対数尤度) + 2（モデルのパラメータ数）

AIC は将来の予測誤差を少なくするように，Kullback 情報量から（漸近的に）導出された規準で，AIC を最小にするモデルが選択されます. 第 1 項がモデルの当てはまりの良さ，第 2 項がいわゆる "オッカムの剃刀" を具体化した項になり，同程度の当てはまりの良さなら，パラメータ数の少ないより単純なモデルを選ぶような仕組みになっています. AIC については，その提案以来，様々な議論が統計学・情報工学などの周辺分野を巻き込みながら展開されています. それぞれの観点からたくさんの情報量規準が提案され，現在でもそれは続いていますが，ここではこれ以上の説明は省略します.

ここで ARMA(p,q) モデルの AIC について結果だけ述べておきます.（導出は結構面倒です. 付録に AR(1) モデルの AIC の導出について簡単に説明をつけますので，そちらを参考にしてください.）ARMA(p,q) モデルの AIC(p,q) は以下で与えられます.

$$AIC(p,q) = \log \hat{\sigma^2} + 2\frac{p+q}{T}$$

（ただし，$\hat{\sigma^2}$ は分散 σ^2 の最尤推定値.）

4 R による時系列解析

以下では，R を使って実際の時系列データについて簡単な解析を行ってみたいと思います. 主に Windows 版の R を用いて，解析・結果表示は行います. 他の OS についても数値結果は同じですがグラフィカルな部分では多少違いが生じるかもしれません.

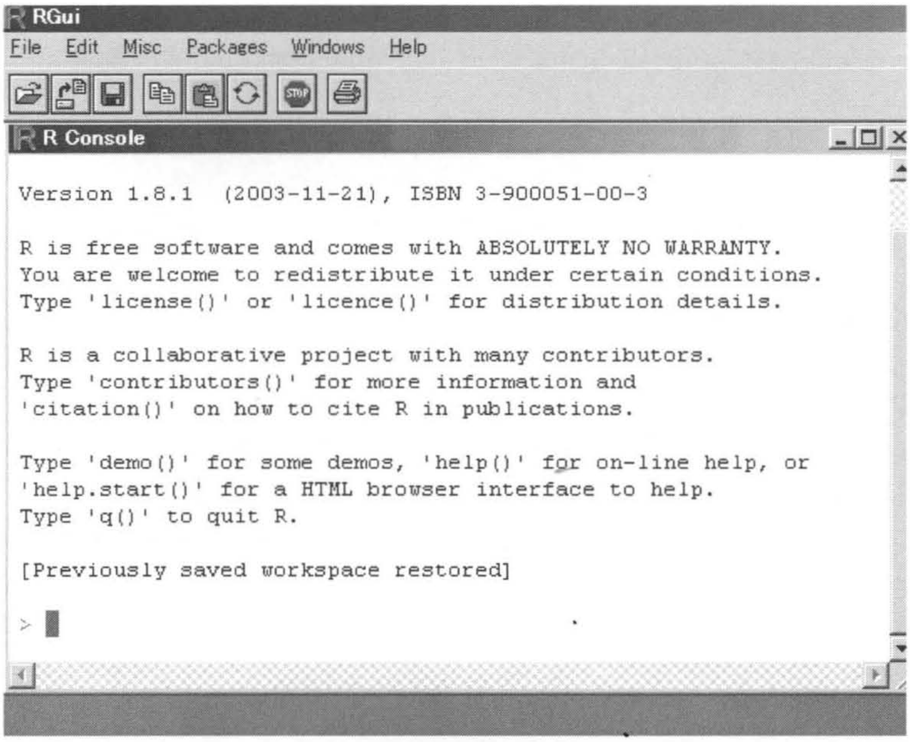

図 1: R の起動時の画面

4.1 R の起動からデータの読み込みまで

図 1 が R の起動時の画面になります (Windows だと，インストール状況にもよりますが，デスクトップ上のアイコンをダブルクリックで，Linux だと kterm などから `R` と入力すると起動します)．データ入力・関数の呼び出し・結果出力は主に R Console とよばれるコンソール上から入力します．例えば `q()` で終了，`help.search(`keyword`)` でキーワード検索などです．また，具体的な関数の利用法などを知りたいときは `help(`具体的な語`)` と入力すると help が表示されます．今回は次のようなデータを用います[†8].

{ 1.0000000 1.0333333 1.0333333 1.0333333 1.0000000 0.9666667 1.0333333 1.0333333 1.0333333 1.0666667 1.0666667 1.0333333 1.0333333 1.0666667 1.1666667 1.1666667 1.1333333 1.1000000 1.0666667 1.0666667 1.1000000 1.1000000 1.1000000 1.1333333 1.1333333 1.1000000 1.1000000 1.1333333 1.1333333 1.1666667 1.1666667 1.1000000 1.1000000 1.0666667 1.0666667 1.0666667 1.0666667 1.0666667 1.0666667 1.0666667 1.0666667 1.1000000 1.1000000 1.1000000 1.1333333 1.1333333 1.1000000 1.0666667 1.1000000 1.1333333 1.1666667 1.1666667 1.1666667 1.1333333 1.1000000}

の 55 点からなる観測データが tab 区切りのテキストファイルとして保存されています (なお，他にもカンマ区切りファイルなども扱えます) [†9]．また，欠損データは含まないとします[†10]．このデータのうち最初の 50 点を解析用のデータとして用い，残りの 5 点をモデルの評価（予測）に用いることとします．

ではまじめに，今回時系列解析に用いる関数群（ライブラリ）のなかで，R が標準で持っていないものを読み込みます．これらを用いる際には事前に CRAN [2] からパッケージとしてダウンロード，

[†8]このような時系列データはいたるところで手にいれることができます

[†9]詳しくは述べませんが，R ではデータフレームとしてデータを取り扱います．データフレームは身長・体重などの属性とその数値データを関連付けて扱うことを可能にします．

[†10]R の関数の中には欠損データの存在も含めた上で解析を行うものもあります．それらは欠損があることをオプション指定することが多いようです

インストールしておく必要があります. (インストールについては `help(install.packages)`, もしくはRのサイトで配布されている「R Installation and Administration」でご確認ください[†11].) 次のコマンドをR Consoleに入力し, ts, tseriesパッケージをロードします.

```
> library(ts)
> library(tseries)
```

次に, 使用する時系列データを`scan`関数でロードします[†12]. データのファイル名を`data55.txt`とすると,

```
> data<-scan(''data55.txt'')
```

となります. (`> data`と入力することで中身を見ることができます.) このデータを時系列として, Rに渡しておきます. (50点の時系列データを`data.ts`とします.)

```
> data.ts<-ts(data,frequency=1,start=c(1,1),end=c(1,50))
```

この時系列をグラフに表示するには`plot`関数を使います.

```
> plot(data.ts,type=''o'',main=''Time Series Data (50 samples)'',ylim=c(0.7,1.3),
  xlab=''time'')
```

図2が表示結果です. ここで本格的な時系列解析を行うためには, 定常性についての議論が必要と

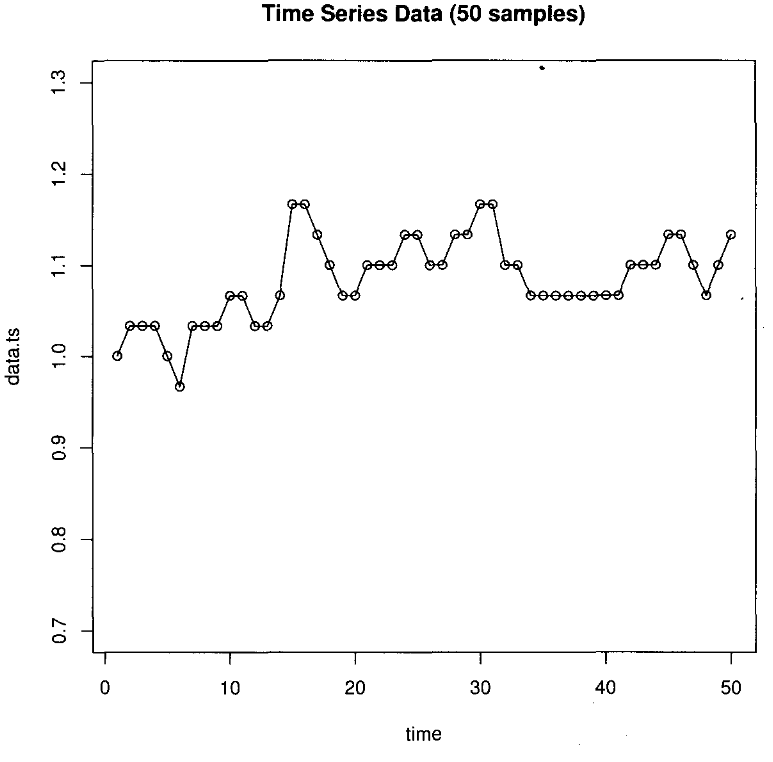

図 2: 時系列データ

なるところですが, 今回はこのままのデータに直接時系列モデルを当てはめることにします.

[†11] Windowsの場合はGUIで簡単にできます.

[†12] 扱うファイルはworking directoryに置く必要があります. また, より高機能なファイル読みこみの関数として`read.table`が用意されています.

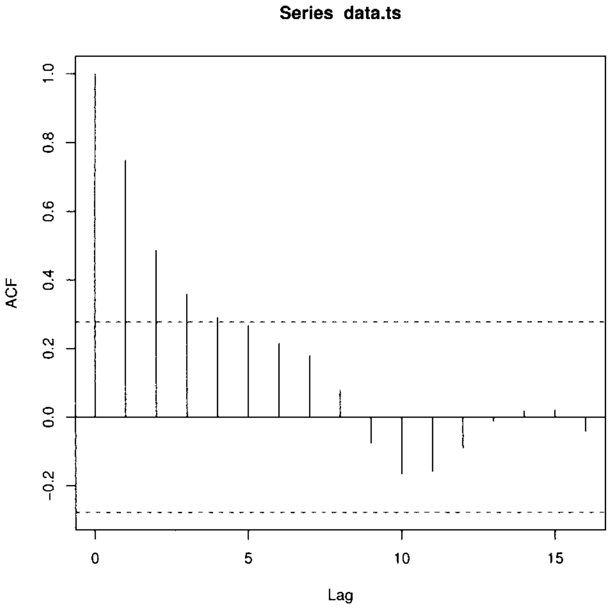

図 3: 標本自己相関関数

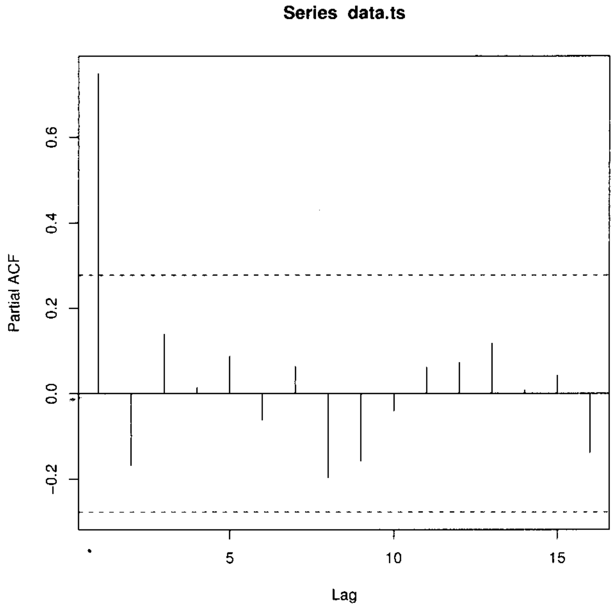

図 4: 標本偏自己相関関数

4.2 自己相関関数と偏自己相関関数の推定

では次にこのデータの（標本）自己相関関数と（標本）偏自己相関関数を求めてみます．それぞれ以下のようなコマンドを入力すると，図 3，図 4 の結果が表示されます．[†13]

```
> acf(data.ts)
> pacf(data.ts)
```

この結果からは，AR 成分・MA 成分両方から構成されているようにみえます．そこで，ARMA モデルの当てはめを検討してみたいと思います．

4.3 ARMA モデルの当てはめ

ARMA モデルの当てはめですが，ここではあまり複雑なモデル（p, q が大きい）モデルは考えずに，*ad hoc* に $1 \leq p \leq 6$，$0 \leq q \leq 6$ までについて AIC に基づいて，モデルの選択を行います．実際の R の関数は `arima` を用います．この関数は ARIMA モデルの当てはめをするもので (p, d, q) に応じて，AIC(p, d, q) や係数 $\{\phi_j\}$，$\{\pi_j\}$ などを求めることができます．また同時にそれぞれのパラメータを選択した際のモデルの AIC が計算されます．ここでは ARMA(p, q)=ARIMA$(p, 0, q)$ として利用します．以下のコマンドを実行することで，パラメータ p, q およびそのときの AIC の値が標準出力に表示されます[†14]．

```
> for (q in 0:6){
> for (p in 1:6){
> x<-arima(data.ts,c(p,0,q),method=c(''ML''),optim.control=list(100000))
> print(c(p,q,x$aic))
```

[†13]図の点線は，相関がないと判断できる信頼限界ということらしいですが，これ自身が信頼できないので，取り扱いには注意が必要です．単なる目安程度と思ってください．

[†14]Warning messages: 1: possible convergence problem: … のようなメッセージが出ることがありますが，これは反復計算による収束判定の条件をみたさないことが原因で，取り扱いには注意が必要です．`optim.control` オプションが `arima` 内で呼ばれる反復計算用の関数 `optim` に渡される引数です．(今回はある程度大きく（∼100000）してみましたが，Warning は消えませんでしたので無視しました．）詳細は `help(optim)` をご覧ください．

```
> }
> }
```

（AICの値は多いので書きませんが）AICを最小にするモデルとしてARMA(4,5)が選ばれました．以下のコマンドで各係数の値と切片が分かります．

```
> arima(data.ts,c(4,0,5),method=c(''ML''),optim.control=list(100000))
```

実際に式に書くと以下のような結果になります．

$$
\begin{aligned}
X_t &= 0.5452X_{t-1} - 0.5234X_{t-2} + 0.6516X_{t-3} + 0.0948X_{t-4} \\
&\quad +0.5469\epsilon_{t-1} + 0.6567\epsilon_{t-2} + 0.0350\epsilon_{t-3} - 0.4560\epsilon_{t-4} - 0.1507\epsilon_{t-5} + \epsilon_t + 1.0794
\end{aligned}
$$

この結果は他の一般的な時系列データの解析結果と比べて，パラメータ数が大きい（i.e. モデルが複雑）ので，ARMAモデルの当てはめそのものに問題があるかもしれません．また，サンプル数によって，AICそのものにバイアスがかかることが知られており，検定等により結果が棄却される可能性もあります．[†15]ここでは検定等の評価は行わずこのモデルを用いて実際に予測してみます．

4.4 予測による検証

では，選択したモデルにより予測を行い実際のデータと比較してみます．予測には，predict関数を用います．以下でデータの名前をわかりやすいものに変更し，5期先まで予測します．

```
> data.past<-data.ts
> data.past.arma<-arima(data.past, c(4,0,5))
> data.pred<-predict(data.past.arma, n.ahead=5)
```

5期先までの予測は，

```
> data.pred$pred
```

で参照することができます．予測結果と実データを比較すると表のようになります．

予測値と実現値の比較

	1期先	2期先	3期先	4期先	5期先
予測値	1.131506	1.116293	1.096350	1.094888	1.105312
実現値	1.1666667	1.1666667	1.1666667	1.1333333	1.1000000

また，以下のコマンドで予測値と実現値がグラフ表示（図5）されます．

```
> ts.plot(data[51:55],data.pred$pred,lty=c(1,2),type="o",ylim=c(0.7,1.3),
   main="Prediction Result")
```

あまり良いとは言えない結果のようですが，このような解析が非常に簡単にRを使うことで実現できるということの紹介にはなったのではないでしょうか？

5 終わりに

以上，GNU Rの紹介と時系列解析への適用の簡単な紹介でしたが，その他にも数多くの機能がRには備わっており，また日々現在の機能の改良や新しい機能の追加が行われています．また，GNU Octave[12]などのGPLライセンスのソフトウェアとの連帯もはかられているようです．このような拡張性の高さもGNUならではのことではないでしょうか？最後にGNU Rをはじめとする，さまざまな方々の努力の恩恵を受けられることを感謝しつつ，つたない文章を終わりたいと思います．

[†15]時系列解析の結果の検定よく使われる統計量にLjung-Box統計量やMcLeod-Li統計量があり，前者による検定はRに実装されています．詳しくはRのBox.test関数のhelpをご覧ください．

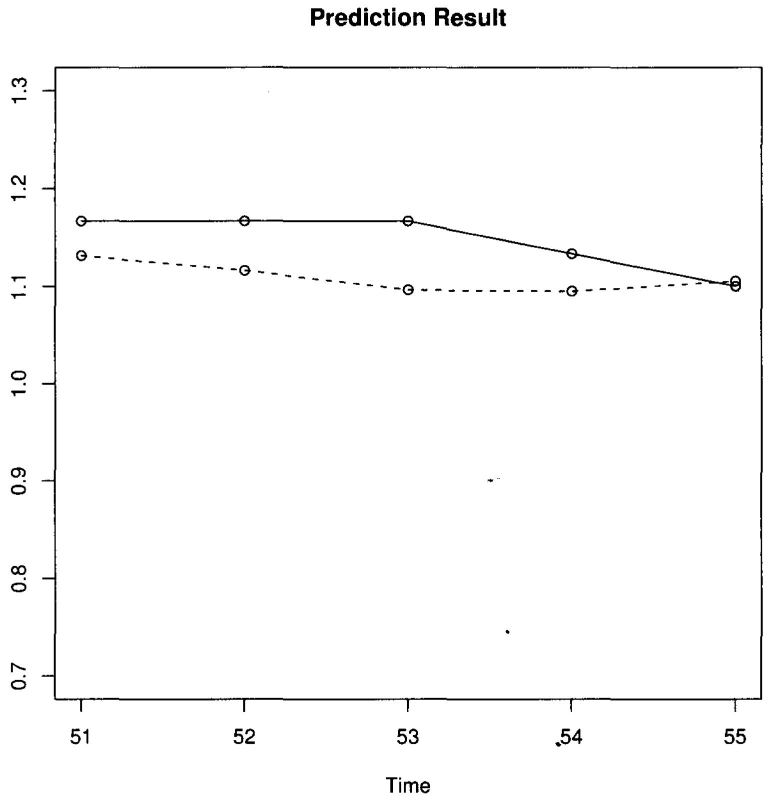

図 5: 予測値と実現値の比較（点線が予測値，実線が実現値）

参考文献

[1] The R project for Statistical Computing, http://www.r-project.org/

[2] The Comprehensive R Archive Network, http://cran.r-project.org/

[3] GNU General Public License (GNU GPL), http://www.gnu.org/copyleft/gpl.html

[4] A. C. ハーベイ，"時系列モデル入門"，東京大学出版会，1985

[5] P. J. Brockwell, R. A. Davis, “Time Series: Theory and Methods ”, Springer Verlag, 1991

[6] 宮野尚哉，"時系列解析入門-線形システムから非線形システムへ-"，サイエンス社，2002

[7] 松葉育雄，"非線形時系列解析"，朝倉書店，2000

[8] 竹村彰通，"現代数理統計学"，創文社，1991

[9] 赤池　弘次，"情報量規準 AIC とは何か"，数理科学，No. 153，p5–p11，サイエンス社，1976

[10] Hirotugu Akaike, "A New Look at the Statistical Model Identification", IEEE Trans. on Automatic Control, Vol. AC-19, No. 6, p716-723, 1974

[11] 竹内　啓，"AIC 基準による統計的モデルの選択をめぐって"，計測と制御，Vol. 22, p29-37, 1983

[12] GNU Octave, http://www.octave.org/

A 付録：AR(1) モデルの AIC の導出

以下では簡単に AR(1) モデルの AIC の導出を行う．

$$X_t = \phi_1 X_{t-1} + \epsilon_t,\ (t = 1, \cdots, T)$$

AR(1) モデルの定常性条件は $|\phi_1| < 1$ であることに注意する [4–8]．ϵ_t が独立に $N(0, \sigma^2)$ に従うとすると

$$f(\epsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right)$$

ここで

$$f(x_1, \cdots, x_T|\phi_1, \sigma^2) = f(x_2, \cdots, x_T|x_1; \phi_1, \sigma^2) f(x_1|\phi_1, \sigma^2)$$

と分解する．まず，第 1 項については ϵ_t の独立性から，

$$\begin{aligned} f(x_2, \cdots, x_T|x_1; \phi_1, \sigma^2) &= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{T-1} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=2}^{T} (x_t - \phi_1 x_{t-1})\right\} \end{aligned}$$

となる．第 2 項については，少し準備がいる．$X_t = \phi_1 X_{t-1} + \epsilon_t$ を書き換えて，

$$X_t = \sum_{j=0}^{J-1} \phi_1^j \epsilon_{t-j} + \phi_1^J X_{t-J}$$

とする．観測値 x_{t-J} が与えられたとき，

$$E(X_t) = E(\sum_{j=0}^{J-1} \phi_1^j \epsilon_{t-j}) + E(\phi_1^J x_{t-J}) = \phi_1^J x_{t-J}$$

となる．ここで，J が十分大きいとき（$J \to \infty$），第 2 項は消滅する．すなわち，時系列が十分遠い過去のある時点から出発したとすれば，

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$$

と書くことができよう．よって，$E(X_t) = 0$ であり，

$$\gamma(0) = E(X_t^2) = E\left(\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}\right)^2 = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j} = \frac{\sigma^2}{1 - \phi_1^2}$$

以上から，定常性より，$X_1 \sim N(0, \frac{\sigma^2}{1-\phi^2})$ となり，

$$f(x_1|\phi_1, \sigma^2) = \frac{\sqrt{1 - \phi_1^2}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(1 - \phi_1^2) x_1^2}{2\sigma^2}\right\}$$

となる．以上から，

$$f(x_1, \cdots, x_T|\phi_1, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T \exp\left\{-\frac{1}{2\sigma^2}\left[(1 - \phi_1^2) x_1^2 + \sum_{t=2}^{T} (x_t - \phi_1 x_{t-1})^2\right]\right\}$$

対数尤度は,

$$l(\phi_1,\sigma^2) = -\frac{T}{2}\log 2\pi\sigma^2 + \frac{1}{2}\log(1-\phi_1^2) - \frac{1}{2\sigma^2}\{\Theta_0(\phi_1)+\Theta(\phi_1)\}$$

ただし,

$$\Theta_0(\phi_1) = (1-\phi_1^2)X_1^2,\ \Theta(\phi_1) = \sum_{t=2}^{T}(X_t - \phi_1 X_{t-1})^2$$

となる．また，X_1 を定数と見たときの（X_1 が与えられたときの）条件付対数尤度 (対数尤度から定数項を除く）を

$$l^*(\phi_1,\sigma^2) = -\frac{T}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\Theta(\phi_1)$$

とする．l を最大にする (ϕ_1,σ^2) を厳密最尤推定量，l^* を最大にする (ϕ_1,σ^2) を条件付最尤推定量という．式からわかるように，データ数が多いときは，これらは同値な推定量となる．よって条件付対数尤度を用いることにすると，l^* を σ^2 で微分して，$\hat{\sigma^2}(\phi_1) = \frac{\Theta}{T-1}$ となる．次に ϕ_1 に関する最大化だが，l^* に $\hat{\sigma^2}(\phi_1)$ を代入すると結局 $\hat{\sigma^2}(\phi_1)$ の最小化に帰着する．(またこのときの $\hat{\phi}_1$ は条件付最小2乗推定量となる.）これから最大対数尤度 l^*_{max} は

$$l^*_{max} = -\frac{1}{2}\log\hat{\sigma^2}(\phi_1)$$

となる．(定数項を除く）AR(1) モデルの AIC を書けば（次数の選択に関係ない定数項を除けば）

$$AIC = \log\hat{\sigma^2}(\phi_1) + \frac{2}{T-1}$$

T が十分大きければ，$T-1 \to T$ と考えてよく[†16],

$$AIC = \log\hat{\sigma^2}(\phi_1) + \frac{2}{T}$$

となる．

[†16] あるいは，X_1 を定数とするとサンプル数は $T-1$ になるので改めて $T-1 \to T$ とする．