時系列ゼミのメモ
第 1 回

伊藤真道

5/18

しばらくは，Brockwell·Daves(2016) *Introduction to time series and forecasting 3rd edition*. 外出自粛明けたら，沖本 *(2010)* 経済・ファイナンスデータの計量時系列分析に移行する？.

1 Chap1 の雑まとめ

1.3.3 時系列モデリングの一般的な方法

- 時系列データをプロットし，グラフの大まかな特徴をとらえる．特に以下の四つについて確認する
  1. トレンド (全体的に上がってる？下がってる？)
  2. 季節的な成分 (6 月は降雨量多いけど，11 月は少ない)
  3. グラフの明らかな傾向変化 (世界恐慌の前後で株価の変動の様相は全然違うよね)
  4. 外れ値の有無
- 定常 (**stationaly**) な残差 *(residuals)* を得るために，トレンドや季節的な変動を取り除く．これを達成するために，データの変換 (*e.g.* 対数変換，$Y_t := X_t - X_{t-d}$ 差分をとる *etc.*) を行うケースもある．
- 標本自己相関関数を含む様々な標本に対する統計量を利用して，残差に当てはめるモデルを選択する．
- これまでの残差に対する変換の逆変換を行うことで，元の系列データを復元し，予測 *(forecasting)* を行う．

まとめると，データを可視化し，大まかな特徴をとらえる → 定常な残差を求める → 残差に当てはめるモデルを選び，パラメータの推定を行う → モデルから予測される残差を逆変換し，元の系列の予測値を得る，となる．

以上の手順の他の有効なアプローチとして，フーリエ級数展開によって，時系列データを異なる周期の正弦関数などの和として表現する方法があり，これは信号処理や構造デザインといった工学の分野で利用される．

1.4 Stationary Models and the Autocorrelation Function

ざっくりと定常 *(stationary)* を説明すると，全ての t に対して，X_t と X_t を時間 h だけずらした時の X_{t+h} の統計的性質がそれぞれ等しくなるというのが，それぞれの h の値に対して成り立つこと．h を固定した時のイメージは循環小数的な感じかな．h を固定しない時のイメージは，$1/3 = 0.33333333....$ みたいな．

Def. 1.4.1(平均関数と共分散関数)

$\{X_t\}$ を有限の二乗モーメント $E[X_t^2]$ を持つ系列とする．$\{X_t\}$ の平均関数 *(mean function)* とは，

$$\mu_X(t) := E(X_t)$$

で定義される関数であり，$\{X_t\}$ の共分散関数 *(covariance function)* とは，

$$\gamma_X(r,s) := Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

で定義される関数である．

¡note¿

Def. 1.4.2(weakly stationary)

(i) 平均関数 $\mu_X(t)$ が t に依存しない
(ii) 共分散関数 $\gamma_X(t+h,t)$ がそれぞれのズレ h に対して t に依存しない
(=ズレ h のみの関数になる)
の両方を満たす時，$\{X_t\}$ は弱定常 *(weakly stationary)* であるという．

Remark 1.1. 時系列 $\{X_t, t = 0, \pm 1, \ldots\}$ が強定常 *(strict stationarity)* という性質を持つとは，全ての整数 h，$n > 0$ に対して，$(X_1, X_2, \ldots, X_n)$ と $(X_{1+h}, X_{2+h}, \ldots, X_{n+h})$ が同一の同時分布を持つこととして定義される．明らかに強定常なら弱定常である *(証明のスケッチ考えてみてね)*．以降，定常といったら弱定常を指すものとする．

¡note¿

Remark 1.2. *Def.1.4.2* の *(ii)* から，共分散関数はズレ h のみの関数

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma(t + h, t)$$

のように表すことができる．これは自己共分散関数 *(autocovariance function)* と呼ばれる．

Def. 1.4.3(autocovariance/autocorrelation function)

$\{X_t\}$ を定常な時系列とする．

$$\gamma_X(h) := Cov(X_{t+h}, X_t)$$

を X_t のズレ h の自己共分散関数 *(autocovariance function, ACVF)* と呼び，これを正規化した

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

を X_t のズレ h の自己相関関数 *(autocorrelation function, ACF)* と呼ぶ．

¡notet¿

1.4.1 The sample Autocorrelation Function

実際は，いきなりモデルを構築するわけではない．まずは時系列の依存関係を見つけたり，当てはめるモデルを選択したりするために，標本自己相関関数を計算する．

Def. 1.4.4(sample mean/ACVF/ACF)

$x_1, \ldots, x_n$ を時系列の観測値とする.

$$\bar{x} := \frac{1}{n}\sum_{t=1}^{n} x_t$$

を標本平均 *(sample mean)*,

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}) \ (-n < h < n)$$

を標本自己共分散関数 *(sample autocovariance function)*,

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \ (-n < h < n)$$

を標本自己相関関数 *(sample autocorrelation function)* と呼ぶ.

¡note¿

1.5 Estimation and Elimination of Trend and Seasonal Components

データをプロットしたグラフの精査から，データが以下の発生過程 *(classical decomposition model)*

$$X_t = m_t + s_t + Y_t \tag{1.5.1}$$

の実現値であることがわかる時もある．ここで，m_t はトレンドの効果を表す徐々に変化する関数，s_t は周期 d の季節変動を表す関数，Y_t は定常なランダムノイズを表す項である．時系列データの分析のスタートは，データを変換して定常な系列を得ることであったことから，これらの m_t, s_t を推定し取り除くことが重要である．*¡note¿*

1.5.1 Estimation and Elimination of Trend in the Absence of Seasonality

季節変動がない場合は,

$$X_t = m_t + Y_t \tag{1.5.2}$$

のようになる．このモデルのトレンド m_t の推定法として，移動平均 *(Moving average)* や，スペクトラルスムージング *(spectral smoothing)* が利用されている．ただし，これらはモデルを構築しているわけではないことに注意.

1. 有限の移動平均フィルタで平滑化
   q を非負整数とし，以下の両側移動平均

   $$W_t := \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t-j} \tag{1.5.3}$$

   を考える．これは $q+1 \le t \le n-q$ に対して,

   $$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q} m_{t-j} + \frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t-j} \approx m_t \tag{1.5.4}$$

   となる．ただし，m_t が区間 $[t-q, t+q]$ にて線形であることと，この区間における誤差項の平均がゼロに近いことを仮定している．以上から，トレンドの推定値として,

   $$\hat{m}_t = \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t-j} \text{を用いる．} \tag{1}$$

2. *Exponential smoothing*
   $\alpha \in [0,1]$ に対して，片側移動平均は,

   $$\hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1} \tag{1.5.7}$$

   のように再帰的に定義される．ただし $\hat{m}_1 = X_1$ とする．順繰りに代入していくと，$t \ge 2$ に対して,

   $$\hat{m}_t = \sum_{j=0}^{t-2} \alpha(1-\alpha)^j X_{t-j} + (1-\alpha)^{t-1} X_1$$

   となり，$X_t, X_{t-1}, \ldots$ にだんだん小さくなる重みを掛け合わせていることがわかる.
3. 平滑化によって高頻度の成分を除去する
   高速フーリエ変換等をもちいる.
4. 多項式フィッティング
   $m_t = a_0 + a_1 t + a_2 t^2 + \cdots$ のように，トレンドを時点の多項式で表されると仮定し，データとの誤差二乗和 $\sum_{t=1}^{n}(x_t - m_t)^2$ を最小にするパラメータ $a_0, \ldots$ を求める.

¡note¿

上記の方法の他には，差分をとるという方法がある．以降の表記を簡単にするために，ズレ *1* の差分作用素 Δ を

$$\Delta X_t = X_t - X_{t-1} = (1-B)X_t \tag{1.5.9}$$

と定義する．ただし，B を後方シフト作用素とし，$BX_t = X_{t-1}$ を表すとする．$B^j(X_t) = X_{t-j}$，$\Delta^j(X_t) = \Delta(\Delta^{j-1}(X_t))$，$\Delta^0(X_t) = X_t$ である．Δ は，

$$\begin{aligned}\Delta^2(X_t) &= \Delta(\Delta(X_t)) = (1-B)(1-B)X_t = (1-2B+B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

のようになることに注意．もし，Δ が線形なトレンド関数 $m_t = c_0 + c_1 t$ に適用されたとすると，$\Delta m_t = m_t - m_{t-1} = c_0 + c_1 t - (c_0 + c_1(t-1)) = c_1$ のように定数関数となる．同様にして, 次数 k のトレンド関数も Δ^k を適用することで, 定数関数にできる．例えば, $X_t = m_t + Y_t, m_t = \sum_{j=0}^{k} c_j t^j$ に対して，Δ^k を適用すると，

$$\Delta^k X_t = k! c_k + \Delta^k Y_t$$

のように，平均 $k! c_k$ の定常過程となる．ここから，与えられた系列 $\{x_k\}$ に対して，$\{\Delta^k x_t\}$ が定常過程からの実現値であると判断できるまで，Δ を適用し続ければ良いということがわかる．実際，実データでは，*1* 次か *2* 次の場合が多いらしい．
¡note¿

1.5.2 Estimation and Elimination of Both Trend and Seasonality

Classical Decomposition Model

$$X_t = m_t + s_t + Y_t,\ (t = 1, \dots, n) \tag{1.5.11}$$

$$where\ E[Y_t] = 0,\ s_{t+d} = s_t,\ and\ \sum_{j=1}^{d} s_j = 0$$

今度は，トレンドと季節変動の両方を取り除くケースを考える．まずは，両方推定してデータから取り除く方法である．まず，$d = 2q$(q が奇数の時は $d = 2q + 1$) として，移動平均を

$$\hat{m}_t = (0.5x_{t-q} + x_{t-q-1} + \cdots + x_{t+q-1} + 0.5x_{t+q})/d \tag{1.5.12}$$

として推定し，データからトレンドを取り除いた系列 $w_k = x_{k+jd} - \hat{m}_{k+jd},\ q < k + jd \le n - q, k = 1, \dots, d$ を計算する．つぎに，季節変動の平均が 0 であったことを考慮して，これを

$$\hat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i,\ k = 1, \dots, d \tag{2}$$

で推定する．定常過程からの実現値は，

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t,\ t = 1, \dots, n \tag{3}$$

として得られる．

次は，差分をとる方法である．ズレ d(季節変動の周期) の差分作用素 Δ_d を

$$\Delta_d X_t = X_t - X_{t-d} = (1 - B^d)X_t$$

と定義する．似ているが $\Delta^d = (1 - B)^d$ であることに注意．これを *classical decomposittion model* に適用すると，

$$\Delta_d X_t = (m_t - m_{t-d}) + (Y_t - Y_{t-d})$$

が得られる．これは，トレンドと定常過程に従う確率変数のみからなる系列であり，トレンドは，前述した移動平均等を用いて取り除くことができる．

¡note¿

1.6 Testing the Estimated Noise Sequence

これまでの手法を用いて，Y_t が求められたとしよう．以降これを残差 *(residuals)* と呼ぶ．残差それぞれに依存関係が見られなければ，観測値は独立な確率変数とみなすことができ，平均や分散を求める以外にそれ以上何かモデリングする必要はない．しかし，依存関係が見られる場合は，より複雑なモデルを試して，依存関係を説明するような定常な系列が得られる変換を探索しなければならない．一度この”うまく説明できるモデル”が得られれば，将来の値の予測にかなり役立つ．

iid 系列であるかどうか判断する方法として，

1. 標本自己相関関数を用いる
   n が大きい時，$Y_1, \ldots, Y_n$ の標本自己相関係数は iid で $N(0, 1/n)$ に従うため，もし $y_1, \ldots, y_n$ が iid 系列からの実現値であるならば，*95%* の標本自己相関係数は，区間 $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$*(95%* 信頼区間*)* に入るはずである．実際に自己相関係数をそれぞれのラグに対して計算してみて，*(i)2* つか *3* つの自己相関係数がこれらの区間の外に出る，*(ii)1* つの自己相関係数が大きく外に出る，のいずれかの場合は，iid の仮定を棄却する．
2. *portmanteau* 検定
   上の方法が，ズレの個数 h 個の自己相関係数の全てを見るのに対し，この検定は，
   $$Q = n \sum_{j=1}^{n} \hat{\rho}^2(j) (\sim \chi_h^2)$$
   という一つの統計量にて実行される．$Y_1, \ldots, Y_n$ が iid サンプルならば，Q は自由度 h の χ^2 分布に従う *(*$\because \sqrt{n}\hat{\rho}(j) \sim N(0,1)$*)* ことを利用して検定を行う．これ以外にも，
   $$Q_{LB} = n(n+2) \sum_{j=1}^{h} \hat{\rho}^2(j)/(n-j) (\sim \chi_h^2)$$
   や
   $$Q_{ML} = n(n+2) \sum_{k=1}^{h} \hat{\rho}_{WW}^2(k)/(n-k) (\sim \chi_h^2)$$
   を用いる方法もある．
3. *The turning point test*
   $y_{i-1} < y_i$，$y_i > y_{i+1}$ や，$y_{i-1} > y_i$，$y_i < y_{i+1}$ となる点の個数を T とすると，
   $$\mu_T = E[T] = 2(n-2)/3$$
   $$\sigma_T^2 = V[T] = (16n-29)/90$$
   となるが，n が大きい時，$T - \mu_T$ は漸近的に $N(\mu_T, \sigma_T^2)$ に従うことを用いて検定を行う．

4. The difference-sign test

この検定では，$y_i - y_{i-1}$ が正となる個数を用いる．これを S とすると，

$$\mu_S = E[S] = (n-1)/2$$
$$\sigma_S^2 = V[S] = (n+1)/12$$

となるが，これも n が大きい時，S が漸近的に $N(\mu_S, \sigma_S^2)$ に従うことを用いて検定を行う．これは，周期的な変動があるときには注意が必要である．

5. The rank test

これは，線形なトレンドを検出するのに有効である．$y_j > y_i$, $j > i$ となるような，(i, j) のペアの個数を P とする．

$$\mu_P = E[P] = n(n-1)/4$$
$$\sigma_P^2 = V[P] = n(n-1)(2n+5)/72$$

となるが，n が大きい時，P は漸近的に $N(\mu_P, \sigma_P^2)$ に従うことを用いて検定を行う．

6. 自己回帰モデルの当てはめ

考えられうる次数に対して *AR* モデルを当てはめ，*AICC* 統計量を算出しこれを最小にする次数を求める．このとき，次数が 0 であるならば，データの系列は，前後と無相関である *(*= ホワイトノイズである*)* といえることを利用．

7. 正規性のチェック

qq プロットをみて，残差に正規性が成り立っているかを確認する．$Y_{(1)}, \ldots, Y_{(n)}$ を $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ の順序統計量とする．もし，$X_{(1)}, \ldots, X_{(n)}$ を $N(0,1)$ からの標本の順序統計量とすると，$X_{(j)}$ と $(Y_{(j)} - \mu)/\sigma$ の分布がともに標準正規分布であることを用いると，$Y_{(j)}$ の期待値は

$$E[Y_{(j)}] = \mu + \sigma m_j,\ m_j = E[X_{(j)}]$$

となる．正規性の仮定が妥当ならば，qq プロットは線形に近くなり，結果として $(m_j, Y_{(j)})$ の重相関係数 R^2 も *1* に近くなる．よって，R^2 がものすんごく小さいとき，正規性の仮定を棄却する．もし，m_j を $\Phi^{-1}((i-0.5)/n)$ で近似すると，重相関係数は

$$R^2 = \frac{(\sum_{i=1}^n (Y_{(i)} - \bar{Y})\Phi^{-1}((i-0.5)/n))^2}{\sum_{i=1}^n (Y_{(i)} - \bar{Y})^2 \sum_{i=1}^n (\Phi^{-1}((i-0.5)/n))^2}$$

と計算できる．

などが利用される．これらを複数試してみることも大事．

¡note¿