

『AI・データサイエンスのための数学入門』学習補助資料

過去を引き継ぎ、新しい誤差が加わる

AR(1)・平均回帰・多期先予測・ウォルド分解

1 観測値と予測値を区別する

AR(1) モデルは $y_t = c + \phi y_{t-1} + \varepsilon_t$ です。予測時に分かる部分と、まだ分からない部分に分けると、

$$y_t = \underbrace{c + \phi y_{t-1}}_{1 \text{ 期先の予測値}} + \underbrace{\varepsilon_t}_{\text{新しい誤差}} .$$

以下では、誤差が平均 0・分散 σ_ε^2 を持ち、過去の観測とも独立な通常の因果的 AR モデルを考えます。係数は既知として計算します。

平均 0 だけで条件付き期待値を決めない

過去の情報を $\mathcal{F}_{t-1}$ と書くと、予測に必要なのは $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ です。この条件の下で $E[y_t | \mathcal{F}_{t-1}] = c + \phi y_{t-1}$ となります。単に $E[\varepsilon_t] = 0$ だけでは不十分です。

本文の来客数の例を整理する

$c = 20, \phi = 0.8, y_0 = 100$ とします。

t	直前の値を使った予測	与えた誤差	モデルの実現値
1	$20 + 0.8(100) = 100$	5	105
2	$20 + 0.8(105) = 104$	-10	94
3	$20 + 0.8(94) = 95.2$	0	95.2

105 と 94 は、誤差を加えて得た値です。将来の誤差は予測時には未知なので、誤差をあらかじめ知っているように予測式に入れません。

人数なのに 95.2 でもよいか

期待値は整数でなくても構いません。ただし、通常の AR モデルは連続量のモデルであり、実現値の非負整数性を保証しません。人数を直接モデル化するときには、適用範囲に注意します。

定数項 20 は、長期平均 20 ではない

c は回帰式の切片です。この例の長期平均は $20/(1 - 0.8) = 100$ です。「毎日のベース 20 人」という説明だけでは、系列が落ち着く水準を読み違えることがあります。

2 平均からのずれで書き直す

定常平均を m とすれば、期待値を取って

$$m = c + \phi m, \quad m = \frac{c}{1 - \phi}.$$

$c = (1 - \phi)m$ を元の式へ代入すると、

$$y_t - m = \phi(y_{t-1} - m) + \varepsilon_t.$$

過去の値そのものより、平均からのずれを何倍して引き継ぐかと考えると分かりやすくなります。

正の係数と負の係数

$0 < \phi < 1$ なら、予測値は直前の値と平均の間にあります。 $-1 < \phi < 0$ なら、平均の反対側に移り、ずれの絶対値が小さくなります。

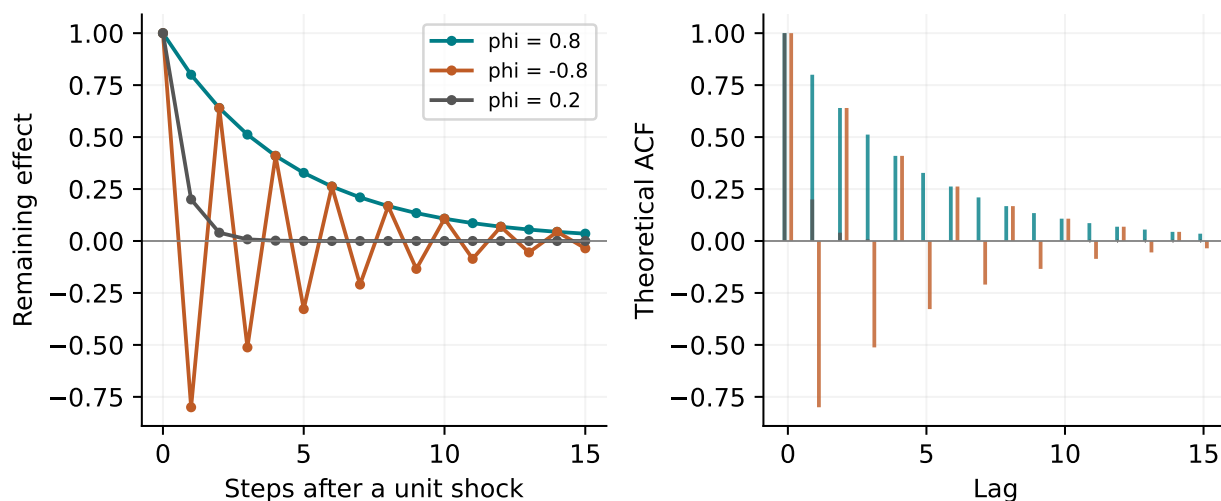


図 1：左は 1 単位のショックが h 時点後に残す効果 ϕ^h 、右は定常 AR(1) の理論 ACF。正の係数では同じ向きに減衰し、負の係数では符号が交互に変わります。

「平均へ戻る」は、毎回の実現値の保証ではない

y_{t-1} が平均より低ければ、正の係数のモデルでは予測値は上に動きます。しかし、大きな負の誤差が加われば実現値はさらに低くなります。平均回帰は予測値や条件付き平均についての性質です。

$|\phi| < 1$ と、初期値の扱い

$|\phi| < 1$ なら、過去の誤差から構成される定常な因果的解が存在します。ただし y_0 を固定して始めた系列が、初日から定常になるとは限りません。初期値の影響は ϕ^t に比例して弱まります。

$\phi = 1$ では $y_t = y_{t-1} + c + \varepsilon_t$ となり、通常はランダムウォーク ($c \neq 0$ ならドリフト付き) です。 ϕ が 1 に近い場合と、1 に等しい場合は区別します。

3 分散と自己相関を導く

$x_t = y_t - m$ とすると、 $x_t = \phi x_{t-1} + \varepsilon_t$ です。ここでは定常な因果的解、 $|\phi| < 1$ 、 $\sigma_\varepsilon^2 > 0$ を仮定します。

系列の分散と、誤差の分散の関係

誤差は過去と独立なので、 $v = V[x_t] = V[x_{t-1}]$ と置くと、

$$v = \phi^2 v + \sigma_\varepsilon^2, \quad \boxed{V[y_t] = v = \frac{\sigma_\varepsilon^2}{1 - \phi^2}}.$$

例えば $\phi = 0.8$, $\sigma_\varepsilon = 6$ なら $v = 36/0.36 = 100$ 、系列の標準偏差は 10 です。新しい誤差の標準偏差 6 と、系列全体の標準偏差 10 は別の量です。

自己相関は係数のべきになる

$\gamma_k = \text{Cov}(x_t, x_{t-k})$ とします。 $k \geq 1$ では、誤差と過去の共分散が 0 なので、

$$\gamma_k = \phi \text{Cov}(x_{t-1}, x_{t-k}) = \phi \gamma_{k-1}.$$

繰り返すと $\gamma_k = \phi^k \gamma_0$ 。よって、

$$\boxed{\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k}.$$

$\phi = 0.8$ なら $\rho_1 = 0.8$, $\rho_2 = 0.64$, $\rho_3 = 0.512$ です。有限データから計算する標本 ACF は、必ずこの値に一致するわけではありません。

1 回のショックが残す影響

同じ将来の誤差を与える 2 つの系列を比べ、ある時点で片方にだけ D を加えると、次の時点の差は ϕD 、さらに次は $\phi^2 D$ です。 h 時点後の差は $\phi^h D$ となります。

負の係数は、規則的な季節性そのものではない

$\phi < 0$ の ACF は符号を交互に変えながら減衰します。これは固定された大きさの季節変動が永久に続くという意味ではありません。系列には毎時点、新しい誤差も加わります。

4 1 期先予測と、多期先予測

時点 t までを観測したときの、 h 期先の予測を $\hat{y}_{t+h|t}$ と書きます。1 期先では

$$\hat{y}_{t+1|t} = c + \phi y_t = m + \phi(y_t - m).$$

2 期先以降は、まだ見ていない値の代わりに予測値を代入します。

$$\hat{y}_{t+h|t} = m + \phi^h(y_t - m).$$

例： $c = 20, \phi = 0.8, y_t = 120$

平均は 100 なので、1・2・3 期先の予測値は

$$116, \quad 112.8, \quad 110.24.$$

ただし翌日の実測が 110 だったなら、その時点からの新しい 1 期先予測は $20 + 0.8(110) = 108$ です。以前に作った 2 期先予測 112.8 とは違います。

先へ行くほど、新しい誤差が積み重なる

反復代入すると、予測誤差は

$$y_{t+h} - \hat{y}_{t+h|t} = \sum_{j=0}^{h-1} \phi^j \varepsilon_{t+h-j}.$$

したがって条件付き分散は

$$\sigma_\varepsilon^2 \sum_{j=0}^{h-1} \phi^{2j} = \sigma_\varepsilon^2 \frac{1 - \phi^{2h}}{1 - \phi^2}.$$

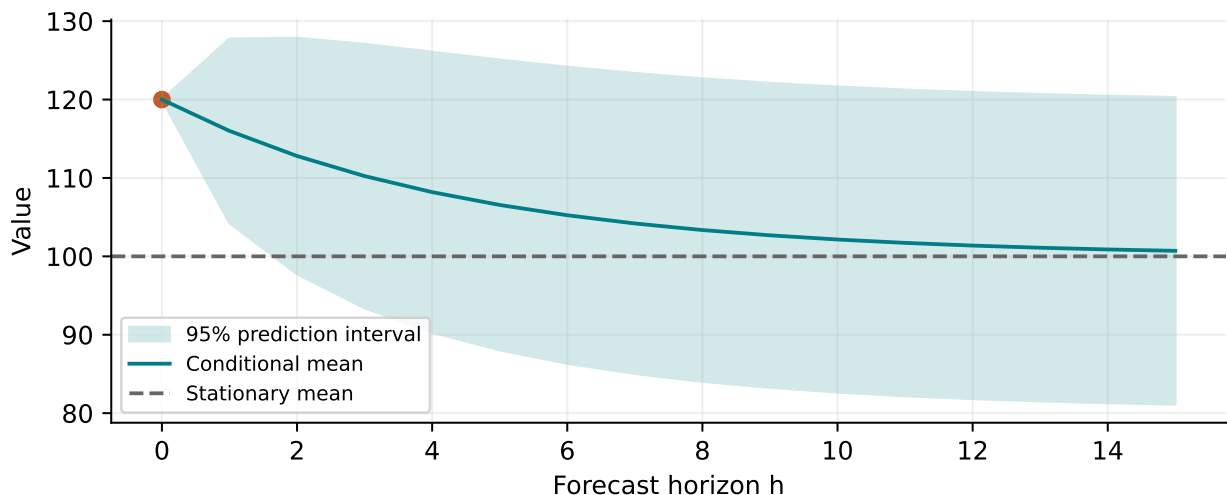


図 2：上の例で $\sigma_\varepsilon = 6$ 、独立な正規誤差を仮定。帯は予測値 $\pm 1.96 \times$ 予測誤差の標準偏差。係数の推定誤差は含めていません。

係数既知・正規誤差の条件で、この帯は各予測時点の約 95% 予測区間です。正規性を仮定しない場合、同じ分散公式は使えても 95% という確率は保証されません。

5 係数をどう学び、予測をどう評価するか

観測値 $y_0, y_1, \dots, y_n$ から、説明変数 $x_t = y_{t-1}$ と目的変数 y_t の n 組を作ります。

時点	説明変数	目的変数
1	y_0	y_1
2	y_1	y_2
$\vdots$	$\vdots$	$\vdots$
n	y_{n-1}	y_n

最小二乗法で当てはめる

$$\sum_{t=1}^n (y_t - c - \phi x_t)^2$$

を小さくする c, ϕ を求めます。 $\bar{x} = n^{-1} \sum x_t$ 、 $\bar{y}_+ = n^{-1} \sum_{t=1}^n y_t$ と置くと、分母が正なら

$$\hat{\phi} = \frac{\sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y}_+)}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad \hat{c} = \bar{y}_+ - \hat{\phi} \bar{x}.$$

$\bar{x}$ と $\bar{y}_+$ は対象とする時点が違うため、同じ値とは限りません。また、制約なしで推定した $\hat{\phi}$ が必ず $(-1, 1)$ に入るわけではありません。

未来を隠して予測を評価する

例えば前半の期間で係数を求め、後半を検証に使います。時点を進めながら、**その時点までに得た情報だけで**予測を作ります。未来の観測から計算した平均や係数を混ぜると、実際より良く予測できたように見えてしまいます。

誤差の大きさを測る

$$MAE = \frac{1}{q} \sum_{t=1}^q |e_t|, \quad RMSE = \sqrt{\frac{1}{q} \sum_{t=1}^q e_t^2}.$$

両者とも元のデータと同じ単位です。RMSE は大きな誤差をより強く反映します。二乗平均と算術平均の関係から $RMSE \geq MAE$ です。

小さい誤差だけでなく、残差の並びも確認

残差に周期性や自己相関が残っていれば、AR(1) だけでは説明しきれていない可能性があります。過去の値を複数使う AR(p) や、季節要因・外部の説明変数を考えるきっかけになります。

6 ウォルド分解を AR(1) から理解する【発展】

中心化した AR(1) を繰り返し展開すると、

$$\begin{aligned} x_t &= \varepsilon_t + \phi x_{t-1} \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 x_{t-2} \\ &= \sum_{j=0}^{q-1} \phi^j \varepsilon_{t-j} + \phi^q x_{t-q}. \end{aligned}$$

定常で $|\phi| < 1$ なら $E[(\phi^q x_{t-q})^2] = \phi^{2q} V[x_t] \rightarrow 0$ 。よって平均二乗の意味で

$$x_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \phi^{2j} = \frac{1}{1-\phi^2} < \infty.$$

AR(1) は、過去の誤差の影響を幾何級数の重みで積み重ねたものとも見られます。

一般の定理で重要な「線形」という限定

ウォルド分解では、有限の 2 次モーメントを持つ弱定常過程を、線形に決定論的成分と、イノベーションの無限和で表せる成分に分けます。本文の「予測」は、基本的には過去の値の線形結合と、その平均二乗極限による予測を意味します。

イノベーションは、現在の値から、過去全体による最良線形予測を引いた差です。過去と直交しますが、一般に独立である必要はなく、条件付き期待値が 0 とも限りません。AR(1) で独立な誤差を仮定した場合は、より強い条件を置いています。

純粋に非決定論的でも、全部が予測不能ではない

AR(1) の x_t では、 $\phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots$ は過去から予測でき、新しい ε_t が予測誤差になります。「非決定論的成分」全体と「その時点の新しい誤差」は別です。

移動平均による平滑化と、MA 型モデルは別

ここでの無限移動平均表示は誤差項の重み付き和です。観測値 y_t の窓内平均を取る SMA とは、対象と目的が異なります。また、ウォルド分解の決定論的成分は、単純に直線トレンドを意味しません。定理の対象は弱定常過程です。

参考：MIT OpenCourseWare, 18.S096 (2013), Lecture 8, “Time Series Analysis”, Wold Representation Theorem.

講義資料 (MIT OpenCourseWare)

7 本文の練習問題を解く

問 1：平均回帰

過去に対する誤差の条件付き期待値が 0 というモデルの仮定により、 $y_{t-1} = 20$ のとき、

$$E[y_t \mid y_{t-1} = 20] = 5 + 0.8(20) = \boxed{21}.$$

定常平均は $m = 5 + 0.8m$ から

$$0.2m = 5, \quad \boxed{m = 25}.$$

20 は平均 25 より低く、予測値 21 は平均へ近づく方向です。平均からの差は $20 - 25 = -5$ から $21 - 25 = -4$ へ、0.8 倍になっています。

問 2：各時点で観測を更新する 1 期先予測

前時点の**実測値**を使います。2 期先・3 期先を時点 0 からまとめて予測する計算とは異なります。

t	y_{t-1}	$\hat{y}_t = 5 + 0.8y_{t-1}$	y_t	$e_t = y_t - \hat{y}_t$
1	20	21	22	1
2	22	22.6	21	-1.6
3	21	21.8	24	2.2

$$MAE = \frac{1 + 1.6 + 2.2}{3} = \boxed{1.6},$$

$$RMSE = \sqrt{\frac{1^2 + (-1.6)^2 + 2.2^2}{3}} = \sqrt{\frac{8.4}{3}} = \boxed{\sqrt{2.8} \approx 1.6733}.$$

比較：時点 0 からの多期先予測なら

$y_0 = 20$ から観測を更新せずに予測すると、

$$\hat{y}_{1|0} = 21, \quad \hat{y}_{2|0} = 21.8, \quad \hat{y}_{3|0} = 22.44.$$

これは上の表の予測とは別です。予測の評価では「いつの情報で、何期先を予測したのか」を明記します。

定数項と平均、誤差分散と系列分散を区別

このモデルでは $c = 5$ ですが平均は 25 です。さらに誤差分散を σ_ε^2 とすれば、定常系列の分散は $\sigma_\varepsilon^2/0.36$ となります。式に現れる似た量を同一視しないようにします。

8 追加問題と解答

問 A：平均を中心に書く

$y_t = 6 + 0.7y_{t-1} + \varepsilon_t$ の定常平均を求め、中心化した式を書け。

解答： $m = 6/0.3 = \boxed{20}$ 。 $y_t - 20 = 0.7(y_{t-1} - 20) + \varepsilon_t$ 。

問 B：負の係数の予測

$m = 10, \phi = -0.5, y_t = 14$ のとき、 $1 \cdot 2 \cdot 3$ 期先の予測を求めよ。

解答： $10 + (-0.5)^h(14 - 10)$ より、 $\boxed{8}$ 、 $\boxed{11}$ 、 $\boxed{9.5}$ 。平均の両側を交互に移りながら近づきます。

問 C：系列の分散と ACF

$\phi = 0.6, \sigma_\varepsilon^2 = 4$ の定常 AR(1) について、系列の分散、標準偏差、 ρ_1, ρ_2 を求めよ。

解答：分散 $4/(1 - 0.36) = \boxed{6.25}$ 、標準偏差 $\boxed{2.5}$ 。 $\rho_1 = \boxed{0.6}$ 、 $\rho_2 = \boxed{0.36}$ 。

問 D：2 期先の予測誤差

問 C のモデルで、2 期先予測誤差の分散を求めよ。1 期先と比べよ。

解答： $4(1 + 0.6^2) = \boxed{5.44}$ 。1 期先の 4 より大きい値です。2 つの将来の誤差が予測誤差に入ります。

問 E：ショックの持続

$\phi = 0.8$ のモデルで、ある時点にだけ 10 の追加ショックを与える。同じ将来の誤差を持つ系列と比べて、2 時点後と 5 時点後の差を求めよ。

解答： $10(0.8)^2 = \boxed{6.4}$ 、 $10(0.8)^5 = \boxed{3.2768}$ 。

問 F：定常性と初期値

$y_0 = 0$ を固定し、 $y_t = 0.8y_{t-1} + \varepsilon_t$ 、独立な誤差の分散を 1 とする。 $V[y_1], V[y_2]$ を求め、初日から定常か答えよ。

解答： $V[y_1] = \boxed{1}$ 、 $V[y_2] = \boxed{1.64}$ 。分散が変化するので初日からは定常ではありません。長期的には $1/(1 - 0.64) = 25/9$ に近づきます。

問 G：ウォルド分解の読み方

「純粋に非決定論的な成分は、過去からまったく予測できない」という説明を訂正せよ。

解答：過去のイノベーションの寄与には予測できる部分があります。1 期先予測で新しく残る誤差が、現在のイノベーションです。一般の定理では線形予測を考えます。