

『AI・データサイエンスのための数学入門』学習補助資料

窓を動かして、変化の傾向を読む

移動平均の計算・平滑化と遅れ・指数平滑化

1 移動平均の値を、どの時点に置くか

本文の移動平均は、現在と過去を使う**後方移動平均**です。時点 t の値は、その時点でそろっている直近 k 個の平均です。

$$MA_t = \frac{1}{k} \sum_{i=0}^{k-1} y_{t-i} \quad (t \geq k).$$

例：幅 3 の窓を 1 つ進める

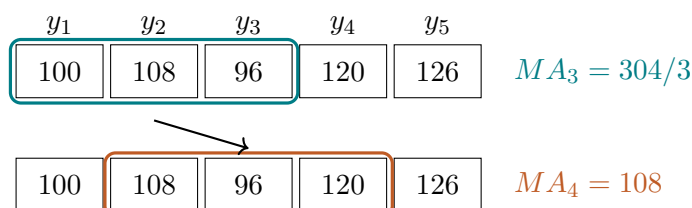


図 1：上段は時点 3 の窓、下段は時点 4 の窓。100 が窓から出て、120 が入ります。108 と 96 は両方の窓に含まれます。

$$MA_3 = \frac{100 + 108 + 96}{3} = \frac{304}{3}, \quad MA_4 = \frac{108 + 96 + 120}{3} = 108.$$

MA_3 は時点 3 に、 MA_4 は時点 4 に対応させます。グラフに描くときも、どの時点の平均なのかをそろえます。

最初の $k-1$ 時点には完全な窓がない

幅 3 なら $t=3$ から計算でき、 n 個の観測から $n-2$ 個の平均が得られます。最初の 2 時点を 0 と置くわけではありません。小さい窓で代用するなら、別の規則であることを明記します。

観測間隔と欠測にも注意

7 個の平均が「7 日平均」になるのは、日次データが欠けずに並んでいる場合です。欠測値を除いて 7 個取ると、7 日より長い期間になることがあります。

2 平滑化と反応の速さは、両立しにくい

次の人工データは、16 時点目に水準が 100 から 140 へ上がり、独立な小さなノイズが加わったものです。

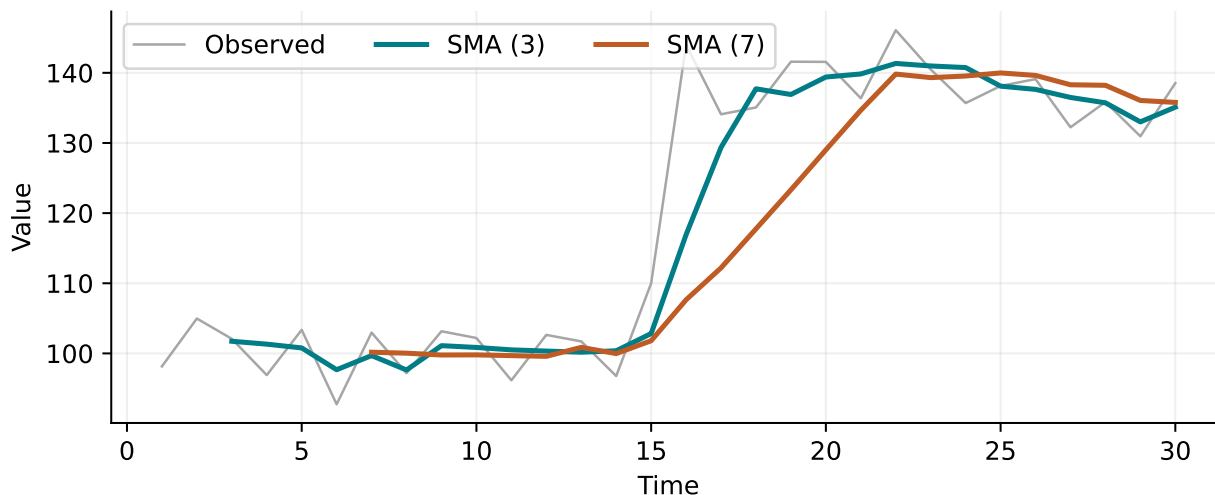


図 2：灰色は観測値、青緑は幅 3、橙は幅 7 の後方移動平均。幅 7 は細かな変動をならしやす一方、水準変化への追従が遅れます。

直線トレンドで、遅れを計算する

$y_t = a + bt$ を代入すると、

$$\begin{aligned} MA_t &= \frac{1}{k} \sum_{i=0}^{k-1} \{a + b(t-i)\} \\ &= a + bt - \frac{b}{k} \frac{k(k-1)}{2} = \boxed{a + b \left(t - \frac{k-1}{2} \right)}. \end{aligned}$$

直線的な変化では $(k-1)/2$ 時点前の水準に一致します。 $b=2, k=5$ なら現在値より 4 低くなります。これは**直線トレンドに対する結果**であり、あらゆる波形の転換点が必ず同じだけ遅れるという意味ではありません。

中心移動平均ならどうなるか

奇数幅 $k = 2h + 1$ の中心移動平均は

$$CMA_t = \frac{1}{2h+1} \sum_{i=-h}^h y_{t+i}.$$

直線トレンドでは $CMA_t = a + bt$ です。ただし、 $h \geq 1$ なら未来の値が必要なので、リアルタイムの判断には使えません。

平滑化と予測を区別する

MA_t は観測後に計算する平滑化値です。これを翌時点の予測値に使うことはできますが、その予測の良さは、学習・調整に使わなかった未来の観測で確かめる必要があります。

3 なぜノイズが弱まるのか

一定水準 m の周りで $y_t = m + \varepsilon_t$ とします。ノイズが独立で、平均 0、分散 σ^2 なら、

$$MA_t = m + \frac{1}{k} \sum_{i=0}^{k-1} \varepsilon_{t-i}.$$

期待値の線形性と独立な和の分散から、

$$E[MA_t] = m, \quad V[MA_t] = \frac{1}{k^2} \sum_{i=0}^{k-1} \sigma^2 = \boxed{\frac{\sigma^2}{k}}.$$

標準偏差は $\sigma/\sqrt{k}$ です。ノイズの標準偏差が 6 なら、幅 9 では 2 になります。相関したノイズでは共分散の項が加わるため、この式をそのまま使えません。

滑らかになっても、値どうしは独立ではない

隣の窓とは $k-1$ 個のノイズが共通です。上の仮定で

$$\text{Cov}(MA_t, MA_{t+1}) = \frac{k-1}{k^2} \sigma^2.$$

$k \geq 2, \sigma^2 > 0$ なら正です。滑らかな連続性の一部は、平均する窓が重なることによって生まれています。

周期と窓幅が一致する場合

季節成分が $6, -2, -4, 6, -2, -4, \dots$ なら、連続 3 個の平均は常に 0 です。一般に、周期 p で 1 周期の和が 0 の成分は、幅 p の平均で消えます。トレンドや他の変動まで消えるわけではありません。

窓幅が大きいほど必ず滑らかになる？

一般的な傾向としてはそうですが、任意のデータに対する厳密な単調性ではありません。例えば $y_t = (-1)^t$ は幅 2 なら平均 0 ですが、幅 3 では $MA_t = (-1)^t/3$ となり、振動が残ります。

外れ値は消えず、複数の平均に広がる

ある 1 つの観測だけが本来の値より D 大きいと、その観測を含む平均は D/k だけ高くなります。系列の端でなければ、連続する k 個の後方移動平均に影響します。

ノイズを抑える効果と、急変・外れ値・周期をどう扱うかを合わせて窓幅を選びます。

4 指数平滑化：過去の情報を徐々に弱める

初期値 $s_1 = y_1$ として、 $t \geq 2$ で

$$s_t = \alpha y_t + (1 - \alpha)s_{t-1} = s_{t-1} + \alpha(y_t - s_{t-1})$$

とします。直前の平滑化値から観測値に向かって、差の α 倍だけ修正する式です。

繰り返し代入すると、重みが見える

$$s_3 = \alpha y_3 + \alpha(1 - \alpha)y_2 + (1 - \alpha)^2 y_1,$$

一般には

$$s_t = \alpha \sum_{j=0}^{t-2} (1 - \alpha)^j y_{t-j} + (1 - \alpha)^{t-1} y_1.$$

重みの合計は 1 です。初期値の項を除けば、重みは等比数列になっています。

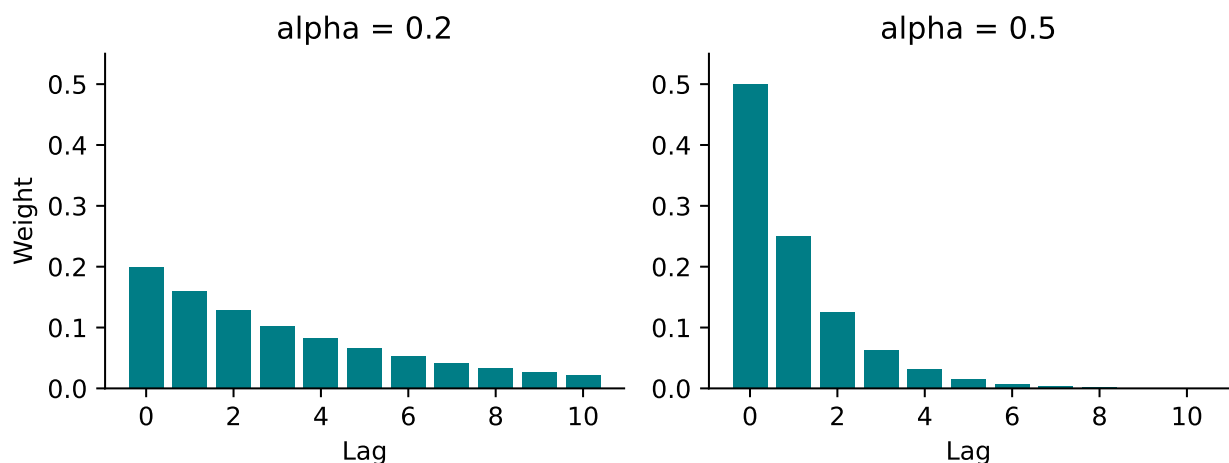


図 3：通常の重み $\alpha(1 - \alpha)^j$ とラグ j の関係。 $\alpha = 0.5$ は現在を強く重視し、0.2 は過去の情報を相対的に長く残します。初期値由来の項は別に扱います。

	単純移動平均	指数平滑化
重み	窓内で一律 $1/k$	時間とともに等比的に減少
過去の扱い	窓外は重み 0	初期値の影響も徐々に減少
調整する量	窓幅 k	係数 α

初期値と、予測誤差に注意

$s_1 = y_1$ は初期化の 1 つの方法です。また、観測後の差 $y_t - s_t$ と、直前の平滑化値を予測に使った誤差 $y_t - s_{t-1}$ は別です。両者は $y_t - s_t = (1 - \alpha)(y_t - s_{t-1})$ の関係にあります。

5 本文の練習問題を解く

問 1：3 期間・5 期間移動平均

データは 100, 108, 96, 120, 126, 114, 132 です。

t	y_t	3 期間 SMA	y_{t-3} 期間 SMA	5 期間 SMA
3	96	$304/3 \approx 101.33$	$-16/3 \approx -5.33$	—
4	120	108	12	—
5	126	114	12	110
6	114	120	-6	112.8
7	132	124	8	117.6

例えば、

$$MA_5^{(3)} = \frac{96 + 120 + 126}{3} = 114, \quad MA_5^{(5)} = \frac{100 + 108 + 96 + 120 + 126}{5} = 110.$$

上付きの (3), (5) は窓幅を区別するための記号です。5 期間は 5 時点目から計算できます。一般に 5 期間の方が短期変動をならしやす一方で、上昇への反応は遅くなります。

計算を簡単にする更新公式

隣り合う窓を引き算すると共通項が消えるので、

$$MA_t - MA_{t-1} = \frac{y_t - y_{t-k}}{k} \quad (t \geq k+1).$$

例えば幅 3 では、 $MA_6 = 114 + (114 - 96)/3 = 120$ と更新できます。

問 2：指数平滑化

$\alpha = 0.5, s_1 = 100$ なので、

$$s_2 = 0.5(108) + 0.5(100) = \boxed{104},$$

$$s_3 = 0.5(96) + 0.5(104) = \boxed{100},$$

$$s_4 = 0.5(120) + 0.5(100) = \boxed{110}.$$

α が大きいほど現在の観測値を強く反映します。SMA は窓内に同じ重みを与え、指数平滑化は通常の過去の重みを等比的に減らします。

観測値との差を、すべてノイズと呼ばない

$y_t - MA_t$ には、ノイズだけでなくトレンドに対する遅れや季節性も残り得ます。2 頁の直線トレンドでは、ノイズがなくても差が生じます。

6 追加問題と解答

問 A：窓の移動

4, 8, 6, 10, 12 の 3 期間移動平均を求めよ。

解答： $MA_3 = \boxed{6}$, $MA_4 = \boxed{8}$, $MA_5 = \boxed{28/3}$ 。各平均は窓の最後の時点に対応します。

問 B：更新公式

幅 4 で $MA_6 = 20$, $y_3 = 12$, $y_7 = 28$ とする。 MA_7 を求めよ。

解答： $MA_7 = 20 + (28 - 12)/4 = \boxed{24}$ 。窓から出るのは y_3 です。

問 C：直線トレンドと遅れ

$y_t = 10 + 3t$ の 5 期間後方移動平均について、 MA_{10} と $y_{10} - MA_{10}$ を求めよ。

解答： $MA_{10} = 10 + 3(10 - 2) = \boxed{34}$ 。 $y_{10} = 40$ なので差は $\boxed{6}$ 。窓の中心は時点 8 です。

問 D：ノイズの分散

独立な平均 0 のノイズの標準偏差が 8 である。一定水準の周りで、幅 16 の移動平均の標準偏差を求めよ。

解答： $8/\sqrt{16} = \boxed{2}$ 。分散は $64/16 = 4$ です。分散と標準偏差を区別します。

問 E：指数平滑化

$s_{t-1} = 80$, $y_t = 100$ とする。 $\alpha = 0.2$ と 0.8 の s_t を求め、違いを説明せよ。

解答：それぞれ $80 + 0.2(20) = \boxed{84}$ 、 $80 + 0.8(20) = \boxed{96}$ 。 0.8 の方が現在の観測値 100 に近くなります。

問 F：外れ値の影響

ある 1 つの観測だけが 50 大きく記録された。幅 5 の移動平均はどう変わるか。系列の端から十分離れているとする。

解答：その観測を含む連続 5 個の移動平均が、それぞれ $50/5 = \boxed{10}$ 高くなります。影響が弱まる一方で複数時点に広がります。

問 G：予測に未来の値を混ぜてよいか

時点 10 で時点 11 を予測するため、時点 10 の中心 3 期間平均 $(y_9 + y_{10} + y_{11})/3$ を使った。この評価の問題点を説明せよ。

解答：予測時には未知の y_{11} を使っています。未来の情報が入り込むため、実際に可能な予測性能を評価できません。時点 10 までに利用できる情報だけを使います。