線形回帰に基づく予測の誤差分散と回帰パラメータの共分散との関係について

情報理工学科　実世界情報コース　島田伸敬

2021 年 12 月 6 日

運動する台車の位置 x を時刻 t ごとに計測し、データ $\{(t_i, x_i)\}(i = 0 \cdots N)$ を得たとする。等速度運動を仮定すると、

$$x = x_0 + v_0 t \tag{1}$$

ただし、x_0 と v_0 は $t = 0$ における台車の位置と速度（初期速度で一定）である。ここで二乗残差の総和

$$S = \sum (x_i - x_0 - v_0 t_i)^2 \tag{2}$$

を最小とする最小二乗法で x_0 と v_0 を推定すれば、x_0 と v_0 は以下の正規方程式

$$N x_0 + \left(\sum t_i\right) v_0 = \left(\sum x_i\right) \tag{3}$$

$$\left(\sum t_i\right) x_0 + \left(\sum t_i^2\right) v_0 = \left(\sum t_i x_i\right) \tag{4}$$

の解である (センシング工学レジュメ後半を参照)。

このとき x_0, v_0 の推定分散 $\sigma_{x_0}^2$, $\sigma_{v_0}^2$、ならびに x_0 と v_0 の共分散 σ_{x_0,v_0} [1]はそれぞれ、

$$\sigma_{x_0}^2 = \frac{\sum t^2}{\Delta} \sigma_x^2 \tag{5}$$

$$\sigma_{v_0}^2 = \frac{N}{\Delta} \sigma_x^2 \tag{6}$$

$$\sigma_{x_0,v_0} = -\frac{\sum t}{\Delta} \sigma_x^2 \tag{7}$$

$$\Delta = N\left(\sum t^2\right) - \left(\sum t\right)^2. \tag{8}$$

ただし、σ_x^2 は x の一回あたりの観測誤差の分散であり、不偏分散を用いて以下の様に表される (同レジュメ参照)。

$$\sigma_x^2 = \frac{1}{N-2} \sum (x_i - x_0 - v_0 t_i)^2 \tag{9}$$

ここで、任意の時刻 τ における台車の位置 $x(\tau)$ を式 1 を用いて予測することを考える。

$$\hat{x}(\tau) = x_0 + v_0 \tau \tag{10}$$

$\hat{x}(\tau)$ の分散 $\sigma_{\hat{x}(\tau)}^2$ は $\sigma_{x_0}^2$,$\sigma_{v_0}^2$,σ_{x_0,v_0} を用いて表すと、独立でない一般の二変数の誤算伝播則 (二乗和バージョン) を使って以下の様になる (授業スライドレジュメ参照)。

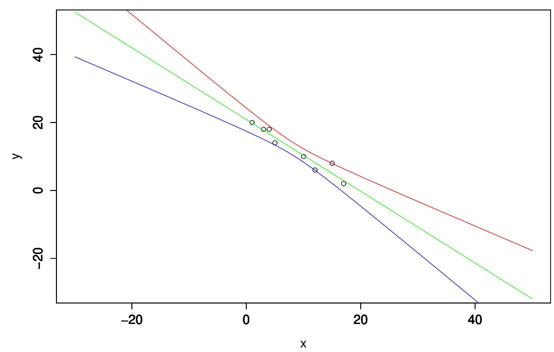

図 1: 線形回帰と誤差推定の例：真の直線が 99%(=3σ) の確率で入る区間を示している。

$$\sigma_{\hat{x}(\tau)}^2 = \left(\frac{\partial x(\tau)}{\partial x_0}\right)^2 \sigma_{x_0}^2 + \left(\frac{\partial x(\tau)}{\partial v_0}\right)^2 \sigma_{v_0}^2 + 2\left(\frac{\partial x(\tau)}{\partial x_0}\right)\left(\frac{\partial x(\tau)}{\partial v_0}\right) \sigma_{x_0,v_0} \tag{11}$$

$\frac{\partial x(\tau)}{\partial x_0} = 1, \frac{\partial x(\tau)}{\partial v_0} = \tau$ より、

$$\sigma_{\hat{x}(\tau)}^2 = \sigma_{x_0}^2 + \tau^2 \sigma_{v_0}^2 + 2\tau \sigma_{x_0,v_0} \tag{12}$$

となる。$\sigma_{\hat{x}(\tau)}^2$ は τ の二次関数であり、二次項の係数 $\sigma_{v_0}^2 > 0$ であるからグラフは下に凸、すなわち最小値を与える τ が一つある。$\sigma_{\hat{x}(\tau)}^2$ が最小、すなわち予測が最も確からしくなる時刻 τ^* を求めると、

$$\frac{\partial \sigma_{\hat{x}(\tau)}^2}{\partial \tau} = 2\tau \sigma_{v_0}^2 + 2\sigma_{x_0,v_0} = 0 \tag{13}$$

$$\begin{aligned} \tau^* &= -\frac{\sigma_{x_0,v_0}}{\sigma_{v_0}^2} \\ &= \left(\frac{\sum t}{\Delta}\sigma_x^2\right) \Big/ \left(\frac{N}{\Delta}\sigma_x^2\right) \\ &= \frac{\sum t}{N} = \bar{t} \end{aligned} \tag{14}$$

[1] 導出は付録参照。

となる。このとき $\sigma^2_{\hat{x}(\tau^*)}$ は

$$
\begin{aligned}
\sigma^2_{\hat{x}(\tau^*)} &= \frac{\sum t^2}{\Delta}\sigma_x^2 + \left(\frac{\sum t}{N}\right)^2 \frac{N}{\Delta}\sigma_x^2 - 2\frac{\sum t}{N}\frac{\sum t}{\Delta}\sigma_x^2 \\
&= \frac{\sigma_x^2}{N\Delta}\left(N\left(\sum t^2\right) - \left(\sum t\right)^2\right) = \frac{\sigma_x^2}{N} \quad (15)
\end{aligned}
$$

である。つまり、予測 $\hat{x}(\tau)$ が最も正確 (推定分散が最小) になるのは時刻データの平均値 $\bar{t}$ のときで、そのときの $\hat{x}(\tau)$ の予測分散 $\sigma^2_{\hat{x}(\bar{t})}$ の値は位置データの平均 $\bar{x}$ の推定分散である $\frac{\sigma_x^2}{N}$ に一致することがわかる (このことはモデル式の傾き（＝速度）v_0=0 と固定したときを考えると自明である)。さらに時刻が $\bar{t}$ から離れるほど予測の推定分散は大きくなることもわかる (図 1)。$\sigma^2_{\hat{x}(\tau)}$ は下に凸な二次関数 (τ^2 の係数が正だから) で τ^* で最小値を取るのだから、τ が τ^* から離れるほど大きくなる。このことは一般に内挿よりも外挿のほうが信頼性に劣るという直観と一致する。

最尤推定で当てはめた運動モデルのパラメタ x_0, v_0 の推定分散を求める式 5,6,7 からわかるように、推定パラメタ x_0, v_0 の推定誤差の間には相関がある。もちろん時刻 τ の予測位置の最良推定値 $\hat{x}(\tau)$ とモデルパラメタ x_0, v_0 の間にも相関がある。これらを 3 次元のベクトル $(x(\tau), x_0, v_0)^T$(T は転置) とみれば、その分散は

$$
\begin{pmatrix}
\sigma_{x_0}^2 + 2\tau\sigma_{x_0,v_0} + \tau^2\sigma_{v_0}^2 & \sigma_{x_0}^2 + \tau\sigma_{x_0,v_0} & \sigma_{x_0,v_0} + \tau\sigma_{v_0}^2 \\
\sigma_{x_0}^2 + \tau\sigma_{x_0,v_0} & \sigma_{x_0}^2 & \sigma_{x_0,v_0} \\
\sigma_{x_0,v_0} + \tau\sigma_{v_0}^2 & \sigma_{x_0,v_0} & \sigma_{v_0}^2
\end{pmatrix} \quad (16)
$$

と行列で表現される（共分散行列。$x(\tau)$ と x_0,v_0 との共分散は各自導出してみよ。よい練習である）。計測データ $\{(t, x)\}$ が時々刻々と得られるとき、推定パラメタ（状態ベクトルという）とその共分散行列を更新しながら最尤推定（あるいは最小二乗推定）していくことができる。これを**状態フィルタリング**と呼び、第 14 回で講義する**線形カルマンフィルタ**がその代表的なものであるが、これは実は第 12、13 回で講義する時系列データの取扱いの内容の発展になっているのである。

付録

線形回帰の推定パラメタ間の共分散は次のようにして求めるとよい。

モデルを $x = x_0 + v_0 t$ とする。このとき推定パラメタ $\hat{x}_0$、$\hat{v}_0$ はそれぞれ正規方程式から次のように求められる。

$$
\hat{x}_0 = \frac{1}{\Delta}\left\{\left(\sum t^2\right)\left(\sum x\right) - \left(\sum t\right)\left(\sum tx\right)\right\}
$$

$$
\hat{v}_0 = \frac{1}{\Delta}\left\{N\left(\sum tx\right) - \left(\sum t\right)\left(\sum x\right)\right\}
$$

これらが最尤推定値であるからこれを中心に真のパラメタは誤差 $\delta x_0 = x_0 - \hat{x}_0$、$\delta v_0 = v_0 - \hat{v}_0$ を持つ。したがって、共分散 σ_{x_0,v_0} は次のように定義される

$$
\sigma_{x_0,v_0} = E[\delta x_0 \delta v_0] \quad (17)
$$

$E[\cdot]$ は $\delta t_1, \cdots, \delta t_N$, $\delta x_1, \cdots, \delta x_N$ に関する期待値である。ところで δx_0 や δv_0 は $\hat{x}_0$,$\hat{v}_0$ の計算に用いた計測値 $\{(t_i, x_i)\}(i = 0 \sim N)$ の計測誤差 $(\delta t_i, \delta x_i)$ からの誤差伝播で計算することができる。ここで t_i の計測誤差 δt_i は 0 とすると x_i の誤差 δx_i だけを考えればよく、$E[\cdot]$ の期待値も $\delta x_1, \cdots, \delta x_N$ についてのみ取れば良い。

$$
\delta x_0 = \sum_i \left(\frac{\partial \hat{x}_0}{\partial x_i}\right)\delta x_i \quad (18)
$$

$$
\delta v_0 = \sum_i \left(\frac{\partial \hat{v}_0}{\partial x_i}\right)\delta x_i \quad (19)
$$

ただし、

$$
\left(\frac{\partial \hat{x}_0}{\partial x_i}\right) = \frac{1}{\Delta}\left[\left(\sum t^2\right) - \left(\sum t\right)t_i\right] \quad (20)
$$

$$
\left(\frac{\partial \hat{v}_0}{\partial x_i}\right) = \frac{1}{\Delta}\left[N t_i - \left(\sum t\right)\right] \quad (21)
$$

である。これを代入して展開していくと、

$$
\begin{aligned}
E[\delta x_0 \delta v_0] &= \frac{1}{\Delta^2}E\left[\left\{\sum_i \left(\left(\sum t^2\right) - \left(\sum t\right)t_i\right)\delta x_i\right\}\cdot\right. \\
&\qquad \left.\left\{\sum_j \left(N t_j - \left(\sum t\right)\right)\delta x_j\right\}\right]. \quad (22)
\end{aligned}
$$

右辺を展開すると $\delta x_i \cdot \delta x_j$ の積の項の和になるが、係数 $\{t_i\}$ は期待値 $E[\cdot]$ に関して定数である (先に $\delta t_i = 0$ と仮定したため期待値は $\delta x_1 \cdots \delta x_N$ に対してのみ取れば良い) から、$E[\cdot]$ の外にくくり出せる。さらにここで x の計測誤差 δx_i と $\delta x_j (i \neq j)$ はそれぞれ独立と仮定すると、$E[\delta x_i \cdot \delta x_j] = 0 (i \neq j)$ であり、結果として $E[\delta x_i^2] = \sigma_x^2$ の項だけが残る。

$$
\begin{aligned}
E[\delta x_0 \delta v_0] &= \frac{1}{\Delta^2}\sum_i \left[\left\{\left(\sum t^2\right) - \left(\sum t\right)t_i\right\}\cdot\right. \\
&\qquad \left.\left\{N t_i - \left(\sum t\right)\right\}E[\delta x_i^2]\right] \\
&= -\frac{\left(\sum t\right)}{\Delta}\sigma_x^2. \quad (23)
\end{aligned}
$$

すなわち、t_i の平均 (つまり重心) が正であれば $\hat{x}_0$ と $\hat{v}_0$ は負の相関を持ち、逆に t_i の平均が負であれば正の相関を持つ。t_i の平均にちょうど x 軸を合わせておけば無相関となることを含意する。