

正定値行列の一般化平均 – 色度認知問題への応用 –

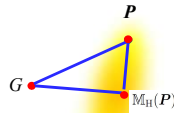
江口 真透 数理・推論研究系 教授

概要: フレッシュ平均とフレッシュ回帰関数

正定値行列の空間上のリーマン幾何と情報幾何

正定値行列の回帰モデルと予測問題

色度空間認知の問題における予測子の提案



1. はじめに

- ◆ (M, ρ) を距離空間とし, X を M 上のランダムオブジェクトとするととき, X の **フレッシュ平均** を

$$\mathbb{M}(X) = \operatorname{argmin}_{y \in M} \mathbb{E}\{\rho^2(X, y)\}$$

と定める. Cf. Fréchet (1948), Kobayashi and Nomizu (1996), Karcher (1997).

- ◆ 特に M がリーマン多様体, ρ をリーマン距離とするととき, \mathbb{M} は **リーマン平均** と呼ぶ. Cf. Cartan mean (1927).

- ◆ 最近, オブジェクトデータの **多様体学習**, ノンパラメトリック統計学の分野が活発な進展を遂げつつある. cf. Patrangenaru & Ellingson (2016).

2. リーマン幾何

- ◆ ここでは, $d \times d$ の正定値行列の空間 $\mathcal{PD} = \{H \in \mathbb{R}^{d \times d} : H = H^\top, H > O\}$ について詳しく考える. 典型的な **リーマン計量** g は

$$g_H(X, Y) = \operatorname{tr}(XH^{-1}YH^{-1})$$

と与えられる. ここで接ベクトル X と Y は $d \times d$ の対称行列である. この計量は合同変換に対して不変であることが分かる: $g_{BBH^\top}(BXB^\top, BYB^\top) = g_H(X, Y)$.

リーマン2乗距離は次で与えられる.

$$D_R(G, H) = \operatorname{tr}\{\{\log(G^{-\frac{1}{2}}HG^{-\frac{1}{2}})\}^2\}$$

- ◆ P を \mathcal{PD} 上の **ランダム行列** とする. このとき, P の **リーマン平均** は次で定められる:

$$\mathbb{M}_R(P) = \operatorname{argmin}_{G \in \mathcal{PD}} \mathbb{E}\{D_R(P, G)\}$$

P が2値で, 等確率 $\mathbb{P}(P = G) = \frac{1}{2}$ and $\mathbb{P}(P = H) = \frac{1}{2}$ ならば,

$$\mathbb{M}_R(P) = G^{\frac{1}{2}}(G^{-\frac{1}{2}}HG^{-\frac{1}{2}})^{\frac{1}{2}}G^{\frac{1}{2}}$$

となる. しかし, 一般の P では行列方程式の解として与えられ, 厳密な解は知れていないので勾配アルゴリズムなどで数値的に求める必要がある. Cf. Moakher (2005).

3. 情報幾何

- ◆ 正定値行列空間のリーマン幾何から, リーマン平均 $\mathbb{M}_R(P)$ が考えられた. しかし, もっと素朴な平均も考えられる. 例えば, 算術平均, 幾何平均, 調和平均などを含む次の **一般化平均** を考えよう.

$$\mathbb{M}_\varphi(P) = \varphi^{-1}(\mathbb{E}\{\varphi(P)\})$$

ここで $\varphi(G)$ は固有分解 $G = BAB^\top$ に対して $B\varphi(\Lambda)B^\top$ とする. 典型例は次である:

- (1) **算術平均** $\mathbb{M}_A(P) = \mathbb{E}(P)$
- (2) **幾何平均** $\mathbb{M}_G(P) = \exp(\mathbb{E}(\log P))$
- (3) **調和平均** $\mathbb{M}_H(P) = (\mathbb{E}(P^{-1}))^{-1}$
- (4) **対数指数平均** $\mathbb{M}_{LE}(P) = \log(\mathbb{E}(\exp P))$

- ◆ \mathcal{PD} 上のウィッシュャート分布の **KL ダイバージェンス** は次のように書ける:

$$D_{KL}(G, H) = \frac{n}{2} \operatorname{tr}[G(H^{-1} - G^{-1}) + \log H - \log G]$$

このとき, \mathcal{PD} 上のランダム行列 P に対して次の **平均ピタゴラスの定理** が成立する.

$$\mathbb{E}D_{KL}(P, G) = \mathbb{E}D_{KL}(P, \mathbb{M}_A(P)) + D_{KL}(\mathbb{M}_A(P), G)$$

$$\mathbb{E}D_{KL}(G, P) = D_{KL}(G, \mathbb{M}_H(P)) + \mathbb{E}D_{KL}(\mathbb{M}_H(P), P).$$

この性質より, 直ちに

$$\mathbb{M}_A(P) = \operatorname{argmin}_{G \in \mathcal{PD}} \mathbb{E}D_{KL}(P, G), \quad \mathbb{M}_H(P) = \operatorname{argmin}_{G \in \mathcal{PD}} \mathbb{E}D_{KL}(G, P)$$

が言える. このように2乗距離に対して定義されたフレッシュ平均がKLダイバージェンスに対して考えると**非対称性**から算術平均と調和平均が導かれることが分かる.

- ◆ 次に**プレグマン・ダイバージェンス**

$$D_\varphi(G, H) = \operatorname{tr}\{\Phi(G) - \Phi(H) - \varphi(G)(H - G)\}$$

を考える. ここで $\Phi = \varphi$. 一般化平均を平均ピタゴラスの定理によって理解できる:

$$\mathbb{E}\{D_\varphi(P, G)\} = \mathbb{E}\{D_\varphi(P, \mathbb{M}_\varphi(P))\} + D_\varphi(\mathbb{M}_\varphi(P), G)$$

$$\mathbb{E}\{D_\varphi(G, P)\} = \mathbb{E}\{D_\varphi(G, \mathbb{M}_\varphi(P))\} + D_\varphi(\mathbb{M}_\varphi(P), G)$$

4. 回帰モデルと予測子の提案

- ◆ 一般の距離空間上のフレッシュ平均は容易に**フレッシュ回帰関数**へ拡張できる. 正定値行列空間のランダム行列 P に対して説明変数ベクトル X が得られるとき,

$$\mathbb{M}_D(P|x) = \operatorname{argmin}_{f: \mathbb{R} \rightarrow \mathcal{PD}} \mathbb{E}[D(P, f(X)) | X = x]$$

を D による P の**予測子**と呼ぶ. Cf. Petersen and Müller (2019)

一般化平均の回帰関数への拡張は条件付き期待値を使って, 自然に

$$\mathbb{M}^{(\varphi)}(P|x) = \varphi^{-1}(\mathbb{E}[\varphi(P) | X = x]).$$

と与えられる. これはプレグマン・ダイバージェンスによる P の予測子となっていることが分かる:

$$\mathbb{M}^{(\varphi)}(P|x) = \operatorname{argmin}_{f: \mathbb{R} \rightarrow \mathcal{PD}} \mathbb{E}[D_\varphi(f(X), P) | X = x]$$

- ◆ このように多様な P の予測子が提案できる. そのため, 予測子の最適性を合同変換に不変なリーマン2乗距離 $D_R(P, Q) = \operatorname{tr}\{\{\log P^{-\frac{1}{2}}Q P^{-\frac{1}{2}}\}^2\}$ を使って導入する.

一般の予測子 $\hat{P}(X)$ に対して**リーマン予測エラー**を以下のように定める:

$$\operatorname{RPE}(\hat{P}) = \mathbb{E}\{D_R(P, \hat{P}(X))\}$$

- ◆ データ $\mathcal{D} = \{(x_i, P_i) : i = 1, \dots, n\}$ が得られたとき, **経験予測子**を

$$\hat{P}_{h,D}^{(\varphi)}(x) = \varphi^{-1}\left(\sum_{i=1}^n \pi_i^D(x, h) \varphi(P_i)\right)$$

と提案する. ここで

$$\pi_i^D(x, h) = \frac{K_h(x, x_i, P_i)}{\sum_{j=1}^n K_h(x, x_j, P_j)}, \quad K_h(x, y, G) = \frac{\exp\{-\frac{1}{2h}(x-y)^\top G^{-1}(x-y)\}}{\det(2\pi h G)^{\frac{1}{2}}}$$

- ◆ 予測子 $\hat{P}_{h,D}^{(\varphi)}(x)$ は, 生成関数 φ が恒等写像ならばナダラヤ・ワトソン型の予測子になる. しかし, 最適な生成関数 φ とバンド幅 h の選択が残されている. 以下の経験リーマン予測エラーを使ってこの選択が可能になる.

データ \mathcal{D} による予測子 $\hat{P}_D(X)$ の**経験リーマン予測エラー**はleave-one-outによって

$$\operatorname{RPE}(\hat{P}_D) = \frac{1}{n} \sum_{i=1}^n D_R(\hat{P}_{D^{(-i)}}(x_i), P_i)$$

と求める. ここで $\mathcal{D}^{(-i)} = \{(x_j, P_j) : j \neq i\}$.

5. 色度認知問題の予測

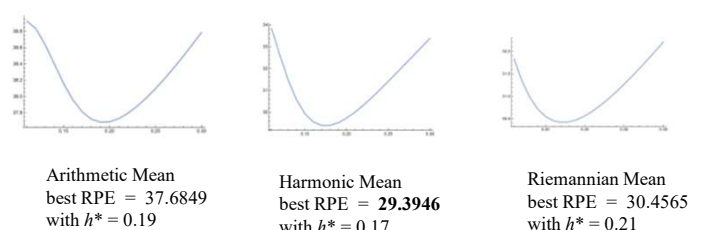
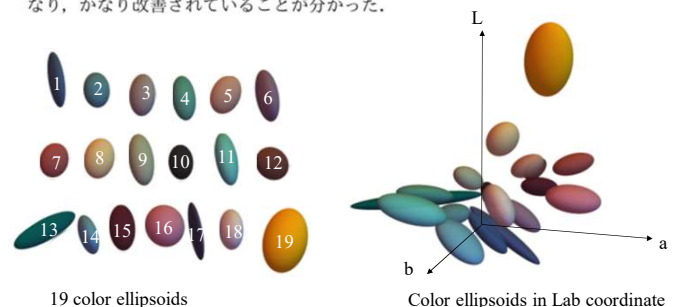
- ◆ 色空間を表す座標としてLabが広く使われている. 座標点 $x = (L, a, b)$ では, L は明るさ, a は緑・赤色度, b は青・黄色度を表す. 色空間の中で人間の**色差領域** (色が区別できない領域)は色によって形状, 大きさともに大きく異なるが**楕円体**で近似できることが報告されている. cf. Melgosa *et al.* (1997).

- ◆ RIT-Dupont 実験データ $\mathcal{D} = \{(x_1, P_1), \dots, (x_{19}, P_{19})\}$ は下図のようになる.

この \mathcal{D} に対して, 予測子 $\hat{P}_{h,D}^{(\varphi)}(X)$ に対して最適バンド幅を以下のように求めた.

$$h_{\text{opt}} = \operatorname{argmin}_{h \in (0, \infty)} \operatorname{RPE}(\hat{P}_{h,D}^{(\varphi)})$$

このようにして提案の予測子の経験リーマン予測エラーを求めたところ, 通常の回帰関数の予測子では**37.685**となるが, 一方で調和平均を使った予測子では**29.395**となり, かなり改善されていることが分かった.



Arithmetic Mean
best RPE = 37.6849
with $h^* = 0.19$

Harmonic Mean
best RPE = **29.3946**
with $h^* = 0.17$

Riemannian Mean
best RPE = 30.4565
with $h^* = 0.21$