離散と連続確率分布における双対平坦構造

和田 達明 [a]、松添 博 [b]

a) 茨城大、b) 名工大

sul 13 settembre 2016 alla Hokudai

Outline

Outline

双対平坦座標系

$$\{\theta^i\} \quad \Leftarrow \text{ dual } \Rightarrow \quad \{\eta_i\}$$

$$\Psi^\star(\boldsymbol{\eta}) = \sum_{i=1}^{N} \theta^i \eta_i - \Psi(\boldsymbol{\theta}),$$

$$\theta^i = \frac{\partial}{\partial \eta_i} \Psi^\star(\boldsymbol{\eta}), \quad \eta_i = \frac{\partial}{\partial \theta^i} \Psi(\boldsymbol{\theta}).$$

双対平坦座標系

$$\{\theta^i\} \quad \Leftarrow \text{ dual } \Rightarrow \quad \{\eta_i\}$$

$$\Psi^\star(\boldsymbol{\eta}) = \sum_{i=1}^{N} \theta^i \eta_i - \Psi(\boldsymbol{\theta}),$$

$$\theta^i = \frac{\partial}{\partial \eta_i}\Psi^\star(\boldsymbol{\eta}), \quad \eta_i = \frac{\partial}{\partial \theta^i}\Psi(\boldsymbol{\theta}).$$

- 離散確率分布と 連続確率分布とでは双対平坦座標系の取り方が異なっている。
- しかし、これらを 区別する用語はないようである?

連続確率分布の双対平坦座標系

$$\text{指数型分布族：}\ \mathcal{S}_{exp} \equiv \left\{ p_\theta(x) \;\middle|\; p_\theta(x) = \exp\left[\sum_{i=1}^{N} \theta^i F_i(x) - \Psi(\theta)\right] \right\}.$$

θ- and η-coordinates

$$\theta^i:\quad \text{分布を特徴付けるパラメータ}$$

$$\eta_i = \int dx\, p_\theta(x)\, F_i(x):\quad \text{期待値パラメータ}$$

$$\Psi^\star(\eta) = \int dx\, p_\theta(x) \ln p_\theta(x)$$

$\mathcal{S}_{exp}$： **各 $\{\eta_i\}$ 一定の下で、**
エントロピ $S = -\Psi^\star(\eta)$ を最大とする確率分布!

連続確率分布の双対平坦座標系

指数型分布族： $\mathcal{S}_{exp} \equiv \left\{ p_\theta(x) \;\middle|\; p_\theta(x) = \exp\left[\sum_{i=1}^{N} \theta^i F_i(x) - \Psi(\theta)\right] \right\}.$

θ- and η-coordinates

$$\theta^i : \quad \text{分布を特徴付けるパラメータ}$$

$$\eta_i = \int dx\, p_\theta(x)\, F_i(x) : \quad \text{期待値パラメータ}$$

$$\Psi^\star(\eta) = \int dx\, p_\theta(x) \ln p_\theta(x)$$

$\mathcal{S}_{exp}$: **各 $\{\eta_i\}$ 一定の下で、エントロピ $S = -\Psi^\star(\eta)$ を最大とする確率分布!**

連続確率分布の双対平坦座標系

指数型分布族： $\mathcal{S}_{exp} \equiv \left\{ p_\theta(x) \;\middle|\; p_\theta(x) = \exp\left[\sum_{i=1}^{N} \theta^i F_i(x) - \Psi(\theta) \right] \right\}.$

θ- and η-coordinates

$$\theta^i : \quad \text{分布を特徴付けるパラメータ}$$

$$\eta_i = \int dx \; p_\theta(x)\, F_i(x) : \quad \text{期待値パラメータ}$$

$$\Psi^\star(\eta) = \int dx \; p_\theta(x) \ln p_\theta(x)$$

$\mathcal{S}_{exp}$: **各 $\{\eta_i\}$ 一定の下で、**
エントロピ $S = -\Psi^\star(\eta)$ を最大とする確率分布!

離散確率分布の双対平坦座標系

$$\mathcal{S}^N \equiv \left\{ \mathbf{p} = (p_\mu) \mid p_\mu > 0, \sum_{\mu=0}^{N} p_\mu = 1 \right\}, \text{ with } p_0 := 1 - \sum_{k=1}^{N} p_k .$$

θ- and η-coordinates

$$\theta^i = \ln p_i - \ln p_0 ,$$
$$\eta_i = p_i , \quad i = 1, \cdots, N.$$

$$\Psi(\theta) = -\ln p_0 , \quad \Psi^\star(\eta) = \sum_{\mu=0}^{N} p_\mu \ln p_\mu .$$

離散確率分布の双対平坦座標系

$$\mathcal{S}^N \equiv \left\{ \mathbf{p} = (p_\mu) \mid p_\mu > 0, \sum_{\mu=0}^{N} p_\mu = 1 \right\}, \text{ with } p_0 := 1 - \sum_{k=1}^{N} p_k.$$

θ- and η-coordinates

$$\theta^i = \ln p_i - \ln p_0,$$
$$\eta_i = p_i, \quad i = 1, \cdots, N.$$

$$\Psi(\theta) = -\ln p_0, \quad \Psi^\star(\eta) = \sum_{\mu=0}^{N} p_\mu \ln p_\mu.$$

離散確率分布の双対平坦座標系

$$\mathcal{S}^N \equiv \left\{ \mathbf{p} = (p_\mu) \mid p_\mu > 0, \sum_{\mu=0}^{N} p_\mu = 1 \right\}, \text{ with } p_0 := 1 - \sum_{k=1}^{N} p_k.$$

θ- and η-coordinates

$$\theta^i = \ln p_i - \ln p_0,$$
$$\eta_i = p_i, \quad i = 1, \cdots, N.$$

$$\Psi(\boldsymbol{\theta}) = -\ln p_0, \quad \Psi^\star(\boldsymbol{\eta}) = \sum_{\mu=0}^{N} p_\mu \ln p_\mu.$$

離散確率分布は指数型分布族

$\theta^i = \ln p_i - \ln p_0$ なので、

$$\ln p_i = \ln p_i - \ln p_0 + \ln p_0 = \theta^i + \ln p_0,$$

離散確率分布は指数型分布族

$\theta^i = \ln p_i - \ln p_0$ なので、

$$\ln p_i = \ln p_i - \ln p_0 + \ln p_0 = \theta^i + \ln p_0,$$

$\ln p_0 = -\Psi(\theta)$ なので、

離散確率分布は指数型分布族

$\theta^i = \ln p_i - \ln p_0$ なので、

$$\ln p_i = \ln p_i - \ln p_0 + \ln p_0 = \theta^i + \ln p_0,$$

$\ln p_0 = -\Psi(\theta)$ なので、

指数型分布族

$$p_i = \exp\left(\theta^i - \Psi(\boldsymbol{\theta})\right), \quad i = 1, \ldots, N.$$

離散確率分布は指数型分布族

$\theta^i = \ln p_i - \ln p_0$ なので、

$$\ln p_i = \ln p_i - \ln p_0 + \ln p_0 = \theta^i + \ln p_0,$$

$\ln p_0 = -\Psi(\theta)$ なので、

指数型分布族

$$p_i = \exp\left(\theta^i - \Psi(\boldsymbol{\theta})\right), \quad i = 1, \ldots, N.$$

$\Rightarrow$ **最大エントロピ法との関連は？**

Outline

1. Introduction
   - 動機
   - 共役表現
2. 双対平坦構造
   - MaxEnt
   - 展望とまとめ

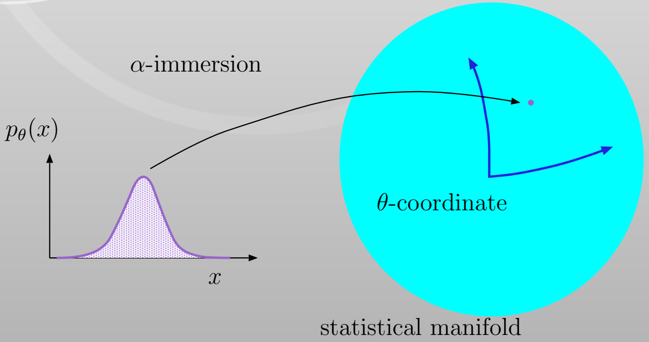
情報幾何学： α 表現

α-表現 (or α-immersion)

$$p \mapsto \ell^{\alpha}(p) := \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}},$$

$$p \mapsto \ell^{-\alpha}(p) := \frac{2}{1+\alpha} p^{\frac{1+\alpha}{2}}, \quad -1 < \alpha < 1.$$

Amari, Nagaoka: *Methods of Information Geometry*, (AMS 2001)



共役表現

conjugate reps.

$$\tau\text{-rep.}\quad p \mapsto \tau(p) = \frac{df(\rho)}{d\rho}(p),$$

$$\rho\text{-rep.}\quad p \mapsto \rho(p) = \frac{df^\star(\tau)}{d\tau}(p),$$

Legendre 双対な凸関数：

$$f(\rho) = \rho\,\tau(\rho) - f^*(\tau(\rho)),$$

$$f^\star(\tau) = \rho(\tau)\,\tau - f(\rho(\tau)).$$

J. Zhang, Neural Comp. **16**, 159, (2004).

共役表現

conjugate reps.

$$\tau\text{-rep.}\quad p \mapsto \tau(p) = \frac{df(\rho)}{d\rho}(p),$$

$$\rho\text{-rep.}\quad p \mapsto \rho(p) = \frac{df^{\star}(\tau)}{d\tau}(p),$$

Legendre **双対な凸関数：**

$$f(\rho) = \rho\,\tau(\rho) - f^{*}\big(\tau(\rho)\big),$$

$$f^{\star}(\tau) = \rho(\tau)\,\tau - f\big(\rho(\tau)\big).$$

J. Zhang, Neural Comp. **16**, 159, (2004).

α-表現は、共役表現の一例：

$$\rho(p) = \ell^{\alpha}(p), \qquad \tau(p) = \ell^{-\alpha}(p),$$

$$f_{\alpha}(\rho) = \frac{2}{1+\alpha}\left(\frac{1-\alpha}{2}\rho\right)^{\frac{2}{1-\alpha}}, \quad f^{\star}_{\alpha}(\tau) = \frac{2}{1-\alpha}\left(\frac{1+\alpha}{2}\tau\right)^{\frac{2}{1+\alpha}}.$$

指数型分布族の場合：

$$\tau(p) = p, \qquad f^{\star}(\tau) = p \ln p,$$

$$\rho = \frac{d}{d\tau} f^{\star}(\tau) = \ln p + 1, \qquad f(\rho) = \exp\left(\rho - 1\right) = p.$$

指数型分布族の場合：

$$\tau(p) = p, \qquad f^\star(\tau) = p \ln p,$$

$$\rho = \frac{d}{d\tau} f^\star(\tau) = \ln p + 1, \qquad f(\rho) = \exp\left(\rho - 1\right) = p.$$

$$-\int dx\, p(x) \ln p(x) = -\int dx\, f^\star\left(\tau(p(x))\right)$$

エントロピ汎関数 $S = -\Psi^\star(\eta)$

Outline

1. Introduction
   - 動機
   - 共役表現
2. 双対平坦構造
   - MaxEnt
   - 展望とまとめ

MaxEnt (最大エントロピ法)

$$\frac{\delta}{\delta p(x)}\left[-\int dx\, f^{\star}(\tau(p(x))) + \sum_{i=1}^{N}\theta^{i}\int dx\, \tau(p(x))\, F_i(x) \right.$$
$$\left. -\gamma\int dx\, \tau(p(x))\right] = 0,$$

ここで、θ^i および γ は、Lagrange 未定乗数。

MaxEnt (最大エントロピ法)

$$\frac{\delta}{\delta p(x)}\left[-\int dx\ f^{\star}(\tau(p(x))) + \sum_{i=1}^{N}\theta^{i}\int dx\ \tau(p(x))\ F_i(x)\right.$$
$$\left. -\gamma\int dx\ \tau(p(x))\right] = 0,$$

ここで、θ^i および γ は、Lagrange 未定乗数。
$df^{\star}(\tau)/d\tau = \rho$ を利用して、

$$\rho(p(x))\ \tau'(p(x)) = \sum_{i=1}^{N}\theta^{i}F_i(x)\ \tau'(p(x)) - \gamma\ \tau'(p(x)).$$

MaxEnt (最大エントロピー法)

$df^\star(\tau)/d\tau = \rho$ を利用して、

$$\rho(p(x))\,\tau'(p(x)) = \sum_{i=1}^{N} \theta^i F_i(x)\,\tau'(p(x)) - \gamma\,\tau'(p(x)).$$

$\tau'(p(x)) \neq 0$ ならば、

$$\rho\big(p(x)\big) = \sum_{i=1}^{N} \theta^i\, F_i(x) - \gamma.$$

MaxEnt (最大エントロピ法)

$\tau'(p(x)) \neq 0$ ならば、

$$\rho(p(x)) = \sum_{i=1}^{N} \theta^i F_i(x) - \gamma.$$

$$\rho(p(x)) = \ln p(x) + 1,$$

$$\Rightarrow \quad \ln p(x) = \sum_{i=1}^{N} \theta^i F_i(x) - \Psi(\theta), \quad \text{with} \quad \Psi(\theta) = \gamma + 1.$$

離散確率分布

$$\rho(p(x)) = \sum_{i=1}^{N} \theta^i F_i(x) - \gamma,$$

の離散版を得るために：

$$p(x_k) \to p_k, \quad F_i(x_k) \to \delta_{i,k}, \quad i, k = 1, \ldots, N.$$

とする。

離散確率分布

$$\rho(p(x)) = \sum_{i=1}^{N} \theta^i F_i(x) - \gamma,$$

の離散版を得るために：

$$p(x_k) \to p_k, \quad F_i(x_k) \to \delta_{i,k}, \quad i, k = 1, \ldots, N.$$

とする。

$$\rho(p_k) = \theta^k - \gamma, \quad k = 1, \ldots, N.$$
$$\rho(p_0) = -\gamma,$$

離散確率分布

$$\rho(p_k) = \theta^k - \gamma, \quad k = 1, \ldots, N.$$
$$\rho(p_0) = -\gamma,$$

$$\theta^k = \rho(p_k) - \rho(p_0), \quad k = 1, \ldots, N.$$

離散確率分布

$$\theta^k = \rho(p_k) - \rho(p_0), \quad k = 1, \ldots, N.$$

$$\rho(p) = \ln p + 1, \quad \Rightarrow \quad \theta^i = \ln p_k - \ln p_0$$

最大エントロピ法との関連は？

$$\frac{\partial}{\partial p_i}\left[-\sum_{\mu=0}^{N} f^{\star}(\tau(p_\mu)) + \sum_{k=1}^{N} \theta^k \tau(p_k) - \gamma \sum_{\mu=0}^{N} \tau(p_\mu)\right] = 0,$$

最大エントロピ法との関連は？

$$\frac{\partial}{\partial p_i}\left[-\sum_{\mu=0}^{N} f^\star(\tau(p_\mu)) + \sum_{k=1}^{N}\theta^k\tau(p_k) - \gamma\sum_{\mu=0}^{N}\tau(p_\mu)\right] = 0,$$

$$\tau(p) = p, \quad f^\star(\tau) = p\ln p, \quad \Rightarrow$$

$$\frac{\partial}{\partial p_i}\left[-\sum_{\mu=0}^{N} p_\mu \ln p_\mu + \sum_{k=1}^{N}\theta^k p_k - \gamma\sum_{\mu=0}^{N} p_\mu\right] = 0,$$

最大エントロピ法との関連は？

$$\tau(p) = p, \quad f^\star(\tau) = p \ln p, \quad \Rightarrow$$

$$\frac{\partial}{\partial p_i}\left[-\sum_{\mu=0}^{N} p_\mu \ln p_\mu + \sum_{k=1}^{N} \theta^k p_k - \gamma \sum_{\mu=0}^{N} p_\mu \right] = 0,$$

各 p_k が一定の下で、エントロピを最大化

スコア関数と τ-期待値

$$\rho(p(x)) = \sum_{i=1}^{N} \theta^i F_i(x) - \gamma(\theta) \quad なので、$$

スコア関数： $s(x) \equiv \partial_i \rho(p(x)) = F_i(x) - \partial_i \gamma,$

スコア関数と τ-期待値

スコア関数： $s(x) \equiv \partial_i \rho(p(x)) = F_i(x) - \partial_i \gamma,$

Then

$$\int dx\ \tau(p(x))\ s(x) = \int dx\ \tau(p(x)) F_i(x) - \partial_i \gamma \int dx\ \tau(p(x)).$$

スコア関数と τ-期待値

スコア関数： $s(x) \equiv \partial_i \rho(p(x)) = F_i(x) - \partial_i \gamma,$

Then

$$\int dx \, \tau(p(x)) \, s(x) = \int dx \, \tau(p(x)) F_i(x) - \partial_i \gamma \int dx \, \tau(p(x)).$$

τ-期待値

$$\mathbb{E}_\tau \left[F_i \right] \equiv \frac{\int dx \, \tau(p(x)) F_i(x)}{\int dx \, \tau(p(x))}.$$

スコア関数と τ-期待値

$$\text{スコア関数：}\quad s(x) \equiv \partial_i \rho(p(x)) = F_i(x) - \partial_i \gamma,$$

Then

$$\int dx\ \tau(p(x))\ s(x) = \int dx\ \tau(p(x)) F_i(x) - \partial_i \gamma \int dx\ \tau(p(x)).$$

τ-期待値

$$\mathbb{E}_\tau\left[F_i\right] \equiv \frac{\int dx\ \tau(p(x)) F_i(x)}{\int dx\ \tau(p(x))}.$$

$$\mathbb{E}_\tau\left[\partial_i \rho(p)\right] = \mathbb{E}_\tau\left[F_i\right] - \partial_i \gamma,$$

where we used $\mathbb{E}_\tau\left[1\right] = 1$.

η-座標系

$$\mathbb{E}_\tau\left[\partial_i\rho(p)\right] = \mathbb{E}_\tau\left[F_i\right] - \partial_i\gamma,$$

$\tau(p) = p, \quad \rho(p) = \ln p + 1$ の場合、$\Psi(\theta) = \gamma + 1$ だったので：

$$\begin{aligned} &\mathrm{E_p}\left[\mathrm{F_i}\right] - \partial_\mathrm{i}\gamma = \mathrm{E_p}\left[\partial_\mathrm{i}\ln\mathrm{p}\right] = \partial_\mathrm{i}\mathrm{E_p}\left[1\right] = 0, \\ \Rightarrow \quad &\eta_i = \partial_i\Psi(\theta) = \partial_i(\gamma+1) = \mathrm{E_p}\left[\mathrm{F_i}\right]. \end{aligned}$$

Outline

変形指数型分布族

κ-指数型分布族

$$\mathcal{S}_\kappa \equiv \left\{ p_\theta(x) \middle| p_\theta(x) = \alpha \exp_{\{\kappa\}} \left[\frac{1}{\lambda} \left(\sum_{i=1}^{N} \theta^i F_i(x) - \gamma(\boldsymbol{\theta}) \right) \right] \right\}.$$

$$\exp_{\{\kappa\}}(x) := \left(\kappa x + \sqrt{1 + \kappa^2 x^2} \right)^{\frac{1}{\kappa}}.$$

κ-表現

$$\tau(p) = p, \qquad f_\kappa^\star(\tau) = \tau \ln_{\{\kappa\}} \tau.$$

$$\rho(p) = \lambda \ln_{\{\kappa\}} \left(\frac{p}{\alpha} \right), \quad f_\kappa(\rho) = \alpha \exp_{\{\kappa\}} \left(\frac{\rho}{\lambda} \right) u_\kappa \left[\alpha \exp_{\{\kappa\}} \left(\frac{\rho}{\lambda} \right) \right].$$

$$\Psi_\kappa^\star = \sum_i f_\kappa^\star(\tau(p_i)) = \sum_i p_i \ln_{\{\kappa\}} p_i = -S_\kappa. \quad \kappa - エントロピ$$

変形指数型分布族

κ-指数型分布族

$$\mathcal{S}_\kappa \equiv \left\{ p_\theta(x) \middle| p_\theta(x) = \alpha \exp_{\{\kappa\}} \left[\frac{1}{\lambda} \left(\sum_{i=1}^{N} \theta^i F_i(x) - \gamma(\boldsymbol{\theta}) \right) \right] \right\}.$$

$$\exp_{\{\kappa\}}(x) := \left(\kappa x + \sqrt{1 + \kappa^2 x^2} \right)^{\frac{1}{\kappa}}.$$

κ-表現

$$\tau(p) = p, \qquad f_\kappa^\star(\tau) = \tau \ln_{\{\kappa\}} \tau.$$

$$\rho(p) = \lambda \ln_{\{\kappa\}} \left(\frac{p}{\alpha} \right), \quad f_\kappa(\rho) = \alpha \exp_{\{\kappa\}} \left(\frac{\rho}{\lambda} \right) u_\kappa \left[\alpha \exp_{\{\kappa\}} \left(\frac{\rho}{\lambda} \right) \right].$$

$$\Psi_\kappa^\star = \sum_i f_\kappa^\star(\tau(p_i)) = \sum_i p_i \ln_{\{\kappa\}} p_i = -S_\kappa. \quad \kappa - エントロピ$$

Summary

- 離散と連続確率分布における双対平坦座標系に注目。
- 共役表現と最大エントロピ法を利用して統一化。