



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



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



- 1 Introduction
 - ■動機

- 2 双対平坦構造
 - MaxEnt
 - 展望とまとめ

双対平坦座標系



$$\begin{aligned} \left\{ \boldsymbol{\theta}^{i} \right\} & \Leftarrow \text{ dual } \Rightarrow \left\{ \eta_{i} \right\} \\ \boldsymbol{\Psi}^{\star}(\boldsymbol{\eta}) &= \sum_{i=1}^{N} \boldsymbol{\theta}^{i} \eta_{i} - \boldsymbol{\Psi}(\boldsymbol{\theta}), \\ \boldsymbol{\theta}^{i} &= \frac{\partial}{\partial \eta_{i}} \boldsymbol{\Psi}^{\star}(\boldsymbol{\eta}), \quad \eta_{i} &= \frac{\partial}{\partial \boldsymbol{\theta}^{i}} \boldsymbol{\Psi}(\boldsymbol{\theta}). \end{aligned}$$

双対平坦座標系



$$\{\theta^{i}\} \iff \text{dual} \implies \{\eta_{i}\}$$

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

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

- 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



指数型分布族:
$$S_{exp} \equiv \left\{ p_{\theta}(x) \mid p_{\theta}(x) = \exp \left[\sum_{i=1}^{N} \theta^{i} F_{i}(x) - \Psi(\theta) \right] \right\}.$$

θ - and η -coordinates

$$heta^i:$$
 分布を特徴付けるパラメータ $\eta_i=\int dx\ p_ heta(x)\,F_i(x):$ 期待値パラメータ $\Psi^*(oldsymbol{\eta})=\int dx\ p_ heta(x)\ln p_ heta(x)$

 $\mathcal{S}_{exp}:$ 合 $\{\eta_i\}$ 一定の下で、 $S=-\Psi^*(oldsymbol{\eta})$ 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:
$$S_{exp} \equiv \left\{ p_{\theta}(x) \mid p_{\theta}(x) = \exp \left[\sum_{i=1}^{N} \theta^{i} F_{i}(x) - \Psi(\theta) \right] \right\}.$$

θ - and η -coordinates

$$\theta^i$$
: 分布を特徴付けるパラメータ

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

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

$$\mathcal{S}_{ extit{exp}}$$
 : $eta \left\{ \eta_i
ight\}$ 一定の下で、
エントロピ $S = -\Psi^\star(oldsymbol{\eta})$ 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:
$$S_{exp} \equiv \left\{ p_{\theta}(x) \mid p_{\theta}(x) = \exp \left[\sum_{i=1}^{N} \theta^{i} F_{i}(x) - \Psi(\theta) \right] \right\}.$$

θ - and η -coordinates

$$\theta^i$$
: 分布を特徴付けるパラメータ

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

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

$$_{\mathcal{S}_{---}}$$
 各 $\{\eta_i\}$ 一定の下で、

エントロピ $S = -\Psi^*(\eta)$ を最大とする確率分布!

離散確率分布の双対平坦座標系



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

θ - and η -coordinates

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

 $\eta_i = p_i, \quad i = 1, \dots, N.$

$$\Psi(heta) = -\ln
ho_0, \quad \Psi^\star(\eta) = \sum_{n=0}^N
ho_\mu \ln
ho_\mu.$$

離散確率分布の双対平坦座標系



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

θ - and η -coordinates

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

 $\eta_i = p_i, \quad i = 1, \dots, N.$

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

離散確率分布の双対平坦座標系



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

θ - and η -coordinates

$$\theta^{i} = \ln p_{i} - \ln p_{0},$$

 $\eta_{i} = p_{i}, \quad i = 1, \dots, N.$

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



$$\theta^i = \ln p_i - \ln p_0$$
 なので、

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



$$\theta^i = \ln p_i - \ln p_0$$
 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、



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

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



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

$$p_i = \exp(\theta^i - \Psi(\theta)), \quad i = 1, \dots, N.$$

⇒ 最大エントロピ法との関連は?



- 1 Introduction
 - 重加機
 - 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

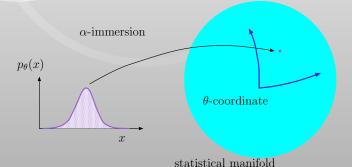


$$\alpha$$
-表現 (or α -immersion)

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

$$lpha$$
-immersion) $p\mapsto \ell^lpha(p):=rac{2}{1-lpha}\,p^{rac{1-lpha}{2}}, \ p\mapsto \ell^{-lpha}(p):=rac{2}{1+lpha}\,p^{rac{1+lpha}{2}}, \quad -1$

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





conjugate reps.

$$au$$
-rep. $p \mapsto au(p) = rac{df(
ho)}{d
ho}(p),$ ho -rep. $p \mapsto
ho(p) = rac{df^{\star}(au)}{d au}(p),$

Legendre 双対な凸関数:

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

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

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



conjugate reps.

$$au ext{-rep.} \quad eta\mapsto au(oldsymbol{
ho})=rac{df(
ho)}{d
ho}(oldsymbol{
ho}), \
ho ext{-rep.} \quad oldsymbol{
ho}\mapsto
ho(oldsymbol{
ho})=rac{df^\star(au)}{d au}(oldsymbol{
ho}),$$

Legendre 双対な凸関数:

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

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

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





α-表現は、<u>共役表現の一例:</u>

$$ho(p) = \ell^{lpha}(p), \qquad au(p) = \ell^{-lpha}(p), \ f_{lpha}(
ho) = rac{2}{1+lpha} \left(rac{1-lpha}{2}
ho
ight)^{rac{2}{1-lpha}}, \quad f_{lpha}^{\star}(au) = rac{2}{1-lpha} \left(rac{1+lpha}{2} au
ight)^{rac{2}{1+lpha}}.$$





指数型分布族の場合:

$$au(p) = p,$$
 $f^*(\tau) = p \ln p,$ $ho = \frac{d}{d\tau} f^*(\tau) = \ln p + 1,$ $f(\rho) = \exp(\rho - 1) = p.$





指数型分布族の場合:

$$au(p) = p,$$
 $f^*(\tau) = p \ln p,$ $ho = \frac{d}{d\tau} f^*(\tau) = \ln p + 1,$ $f(\rho) = \exp(\rho - 1) = p.$

$$-\int dx \ p(x) \ln p(x) = -\int dx \ f^*(\tau(p(x)))$$

エントロピ汎関数 $S = -\Psi^{\star}(\eta)$

12



- 1 Introduction
 - 世沿美相

- 2 双対平坦構造
 - MaxEnt
 - 展望とまとめ



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

ここで、 θ^i および γ は、Lagrange 未定乗数。

14.



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

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

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



$$df^*(\tau)/d\tau = \rho$$
 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、

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



 $\tau'(p(x)) \neq 0$ ならば、

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

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

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

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:

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

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:

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

とする。

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

$$\rho(p_0) = -\gamma,$$

≹K 離散確率分布



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

$$\rho(p_0) = -\gamma,$$

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

⋧ 離散確率分布



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

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

最大エントロピ法との関連は?



$$rac{\partial}{\partial p_i} \Big[- \sum_{\mu=0}^N f^{\star} ig(au(p_{\mu}) ig) + \sum_{k=1}^N heta^k au(p_k) - \gamma \sum_{\mu=0}^N au(p_{\mu}) \Big] = 0,$$

-1

最大エントロピ法との関連は?



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

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

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

最大エントロピ法との関連は?



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

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

各 pk が一定の下で、エントロピを最大化

スコア関数とau-期待値



$$ho(p(x)) = \sum_{i=1}^{N} \theta^{i} F_{i}(x) - \gamma(\theta)$$
 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 au-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au-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au-期待値



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

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 -期待值

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

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

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



$$\mathbb{E}_{\tau}\left[\partial_{i}\rho(\boldsymbol{p})\right] = \mathbb{E}_{\tau}\left[\boldsymbol{F}_{i}\right] - \partial_{i}\gamma,$$

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

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





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

1

变形指数型分布族



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

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

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

$$\Psi_{\kappa}^{\star} = \sum f_{\kappa}^{\star}(\tau(p_i)) = \sum p_i \ln_{f_{\kappa}} p_i = -S_{\kappa}. \quad \kappa - \mathtt{I} \mathsf{V} \mathsf{F} \mathsf{D} \mathsf{E}$$

变形指数型分布族



κ -指数型分布族

$$\mathcal{S}_{\kappa} \equiv \left\{ p_{ heta}(x) \middle| p_{ heta}(x) = lpha \exp_{\{\kappa\}} \left[\frac{1}{\lambda} \left(\sum_{i=1}^{N} \theta^{i} F_{i}(x) - \gamma(oldsymbol{ heta})
ight)
ight]
ight\}.$$

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

κ -表現

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

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

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



- ■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。