擬似スコア関数とプレ・コントラスト関数の幾何学

MATSUZOE Hiroshi
Nagoya Institute of Technology

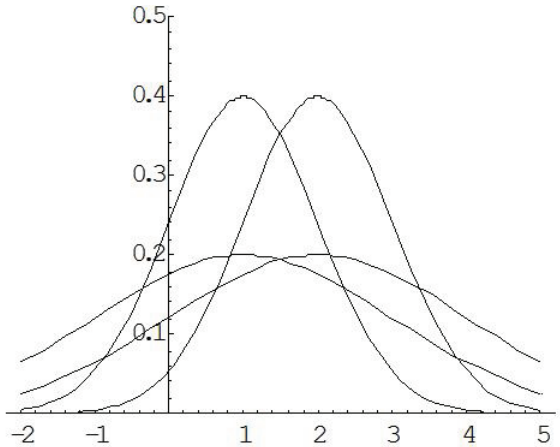
joint works with
HENMI Masayuki
(The Institute of Statistical Mathematics)

1　統計モデルと統計多様体

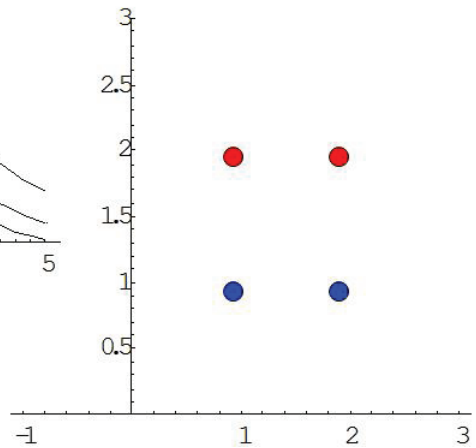
S が Ω 上の 統計モデル （または パラメトリックモデル）
$\overset{\text{def}}{\Longleftrightarrow}$ S が $\xi \in \Xi$ をパラメータとする確率密度関数族で

$$S = \left\{ p(x;\xi) \,\middle|\, \int_\Omega p(x;\xi)dx = 1, p(x;\xi) > 0, \xi \in \Xi \subset R^n \right\}.$$

S を $\{\Xi; \xi^1, \ldots, \xi^n\}$ を局所座標系とする多様体（曲がった空間）とみなす.



多様体



局所座標系

$g^F = (g_{ij})$: S の **Fisher 計量**

$$\overset{\text{def}}{\Longleftrightarrow} \quad g^F_{ij}(\xi) := \int_\Omega \left(\frac{\partial}{\partial \xi^i} \log p(x;\xi)\right)\left(\frac{\partial}{\partial \xi^j} \log p(x;\xi)\right) p(x;\xi)dx$$

$$= \int_\Omega \left(\frac{\partial}{\partial \xi^i} p_\xi\right)\left(\frac{\partial}{\partial \xi^j} \log p_\xi\right) dx \tag{1}$$

$$= \int_\Omega \frac{1}{p(x;\xi)} \left(\frac{\partial}{\partial \xi^i} p_\xi\right)\left(\frac{\partial}{\partial \xi^j} p_\xi\right) dx \tag{2}$$

Proposition 1.1 次の条件は同値

(1) g^F は正値.

(2) $\{\partial_1 p_\xi, \ldots, \partial_n p_\xi\}$ は線形独立.

(3) $\{\partial_1 l_\xi, \ldots, \partial_n l_\xi\}$ は線形独立.

$\partial_i p_\xi \quad \overset{\text{def}}{\Longleftrightarrow}$ m-表現，混合型表現

$\partial_i l_\xi = \left(\dfrac{\partial_i p_\xi}{p_\xi}\right) \quad \overset{\text{def}}{\Longleftrightarrow}$ e-表現，指数型表現.

($p(x;\theta)$ のスコア関数)

1.2 Statistical manifolds

M : a manifold (an open domain in R^n)
h : a (semi-) Riemannian metric on M
∇ : an affine connection on M

Definition 1.2 (Kurose)
We say that the triplet (M, ∇, h) is a statistical manifold

$$\overset{\text{def}}{\Longleftrightarrow} \quad (\nabla_X h)(Y, Z) = (\nabla_Y h)(X, Z).$$

$C(X, Y, Z) := (\nabla_X h)(Y, Z)$, cubic form, Amari-Chentsov tensor field

Definition 1.3
∇^*: the dual connection of ∇ with respect to h

$$\overset{\text{def}}{\Longleftrightarrow} \quad Xh(Y, Z) = h(\nabla^*_X Y, Z) + h(Y, \nabla_X Z).$$

(M, ∇^*, h): the dual statistical manifold of (M, ∇, h).

Remark 1.4 (Original definition by S.L. Lauritzen)
(M, g) : a Riemannian manifold
C : a totally symmetric $(0, 3)$-tensor field
We call the triplet (M, g, C) a statistical manifold.

1.3 Contrast functions

M : a manifold (an open domain in R^n)
$\rho(p,q)$: a function on $M \times M$
$\underline{\rho[X_1 \cdots X_i | Y_1 \cdots Y_j] \text{ : a function on } M}$ defined by

$$\rho[X_1 \cdots X_i | Y_1 \cdots Y_j](r) := (X_1)_{(p)} \cdots (X_i)_{(p)} \, (Y_1)_{(q)} \cdots (Y_j)_{(q)} \rho(p,q)||_{\substack{p=r \\ q=r}}$$

For example,

$$\begin{aligned}
\rho[X|\,](r) &= X_{(p)} \rho(p,q)||_{\substack{p=r \\ q=r}} \\
\rho[X|Y](r) &= X_{(p)} Y_{(q)} \rho(p,q)||_{\substack{p=r \\ q=r}} \\
\rho[XY|Z](r) &= X_{(p)} Y_{(p)} Z_{(q)} \rho(p,q)||_{\substack{p=r \\ q=r}} \\
&\vdots
\end{aligned}$$

Definition 1.5
$\rho : M \times M \to \mathrm{R}$; a contrast function of M

$$\overset{\text{def}}{\Longleftrightarrow} \begin{array}{l} (1)\ \rho(w,w) = 0, \\ (2)\ \rho[X|] = \rho[|X] = 0, \\ (3)\ h(X,Y) := -\rho[X|Y] \quad \text{is a semi-Riemannian metric on } M. \end{array}$$

Example 1.6 When $M = \mathrm{R}^n$, set

$$\rho(x, y) := \frac{1}{2}||x - y||^2, \qquad (x, y) \in \mathrm{R}^n \times \mathrm{R}^n.$$

Then ρ is a contrast function on R^n.

Example 1.7 $S = \{p(x; \theta)\}$: a statistical model

Kullback-Leibler divergence, relative entropy

$$\begin{aligned}\rho_{KL}\left(p(x;\theta), p(x;\theta')\right) &= \int p(x;\theta) \log \frac{p(x;\theta)}{p(x;\theta')} dx \\ &= E_\theta\left[\log p(x;\theta) - \log p(x;\theta')\right]\end{aligned}$$

$$\begin{aligned}\rho_{KL}[\partial_i|\partial_j] &= -\int \partial_i p(\theta) \partial'_j \log p(\theta') dx \Big|_{\theta=\theta'} \qquad \left(\partial_i = \frac{\partial}{\partial\theta^i},\ \partial'_j = \frac{\partial}{\partial\theta'^j}\right) \\ &= -\int \partial_i \log p(\theta) \partial'_j \log p(\theta') p(\theta) dx \Big|_{\theta=\theta'} \\ &= -g^F_{ij} \qquad \text{the Fisher metric} \\ \rho_{KL}[\partial_i\partial_j|\partial_k] &= -\int \left(\partial_i\partial_j l(\theta) \partial'_k l(\theta') + \partial_i l(\theta) \partial_j l(\theta) \partial'_k l(\theta')\right) p(\theta) dx \Big|_{\theta=\theta'} \\ &= -\Gamma^{(m)}_{ij,k} \qquad \text{the mixture connection}\end{aligned}$$

The KL-divergence induces the invariant statistical manifold $(S, \nabla^{(m)}, g^F)$.

Definition 1.7

$\rho : M \times M \to \mathrm{R}$; a contrast function of M

$$\overset{\text{def}}{\Longleftrightarrow} \begin{array}{l} (1)\ \rho(w, w) = 0, \\ (2)\ \rho[X|] = \rho[|X] = 0, \\ (3)\ h(X, Y) := -\rho[X|Y] \quad \text{is a semi-Riemannian metric on } M. \end{array}$$

We can define affine connections ∇ and ∇^* by

$$\begin{aligned} h(\nabla_X Y, Z) &= -\rho[XY|Z], \\ h(Y, \nabla^*_X Z) &= -\rho[Y|XZ]. \end{aligned}$$

$$\Longrightarrow \quad \begin{array}{rcl} \nabla, \nabla^* & : & \text{torsion-free mutually dual with respect to } h. \\ \nabla h, \nabla^* h & : & \text{symmetric } (0,3)\text{-tensor fields.} \end{array}$$

Proposition 1.10

$\rho(p, q)$: *a contrast function on* M

$\Longrightarrow$ *The induced objects* (M, ∇, h), (M, ∇^*, h) *are statistical manifolds.*

Tensor fields B and B^* are defined by

$$\begin{aligned} h(B(X,Y)Z,V) &:= -\rho[XYZ|V] + \rho[\nabla_X \nabla_Y Z|V] \\ h(V,B^*(X,Y)Z) &:= -\rho[V|XYZ] + \rho[V|\nabla^*_X \nabla^*_Y Z] \end{aligned}$$

B : **the Bartlett tensor**
B^* : **the dual Bartlett tensor**

Proposition 1.11 (Eguchi '93)
R, R^* : the curvature tensors of ∇, ∇^*, respectively.
$\Longrightarrow$

$$\begin{aligned} R(X,Y)Z &= B(Y,X)Z - B(X,Y)Z, \\ R^*(X,Y)Z &= B^*(Y,X)Z - B^*(X,Y)Z. \end{aligned}$$

2 Quasi-statistical manifolds

M : a manifold (an open domain in R^n)
h : a non-degenerate $(0,2)$-tensor field on M
∇ : an affine connection on M
$T^\nabla(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$: the torsion tensor of ∇

Definition 2.1
(M,∇,h): a quasi-statistical manifold

$$\overset{\text{def}}{\Longleftrightarrow} \quad (\nabla_X h)(Y,Z) - (\nabla_Y h)(X,Z) = -h(T^\nabla(X,Y),Z)$$

In addition, if h is a semi-Riemannian metric, then we say that (M,∇,h) is a statistical manifold admitting torsion (SMAT).

Definition 2.2
∇^*: (quasi-) dual connection of ∇ with respect to h

$$\overset{\text{def}}{\Longleftrightarrow} \quad Xh(Y,Z) = h(\nabla^*_X Y, Z) + h(Y, \nabla_X Z).$$

Proposition 2.3
The dual connection ∇^* of ∇ is torsion free.

We remark that $(\nabla^*)^* \neq \nabla$ in general.

Proposition 2.4
If h is symmetric $h(X,Y) = h(Y,X)$
or skew-symmetric $h(X,Y) = -h(Y,X)$
$$\Longrightarrow (\nabla^*)^* = \nabla$$

Proposition 2.5
(M, ∇^*, h) : *∇^* is torsion free and dual of ∇,*
h is a non-degenerate $(0,2)$-tensor field,
$\Longrightarrow$ *(M, ∇, h) is a quasi-statistical manifold.*

Suppose that (M, ∇, h) is a statistical manifold admitting torsion.

(1) (M, ∇, h) is a Hessian manifold
$$\Longleftrightarrow \quad R^{\nabla} = 0 \text{ and } T^{\nabla} = 0$$
$\Longleftrightarrow$ (M, h, ∇, ∇^*) is a dually flat space.

(2) (M, ∇, h) is a space of distant parallelism
$$\Longleftrightarrow \quad R^{\nabla} = 0 \text{ and } T^{\nabla} \neq 0 \qquad (R^{\nabla^*} = 0, \quad T^{\nabla^*} = 0).$$

SMAT with the SLD Fisher metric (Kurose 2007)

$\mathrm{Herm}(d)$: the set of all Hermitian matrices of degree d.
$\mathcal{S}$: a space of quantum states

$$\mathcal{S} = \{P \in \mathrm{Herm}(d) \mid P > 0,\ \mathrm{trace} P = 1\}$$

$T_P\mathcal{S} \cong \mathcal{A}_0$ $\qquad\qquad$ $\mathcal{A}_0 = \{X \in \mathrm{Herm}(d) \mid \mathrm{trace} X = 0\}$

We denote by $\widetilde{X}$ the corresponding vector field of X.

For $P \in \mathcal{S},\ X \in \mathcal{A}_0$, define $\omega_P(\widetilde{X})$ $(\in \mathrm{Herm}(d))$ by

$$X = \frac{1}{2}(P\omega_P(\widetilde{X}) + \omega_P(\widetilde{X})P)$$

The matrix $\omega(\widetilde{X})$ is the "symmetric logarithmic derivative".

A Riemannian metric and an affine connection are defined as follows:

$$h_P(\widetilde{X}, \widetilde{Y}) = \frac{1}{2}\mathrm{trace}\left(P(\omega_P(\widetilde{X})\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})\omega_P(\widetilde{X}))\right),$$

$$\left(\nabla_{\widetilde{X}}\widetilde{Y}\right)_P = h_P(\widetilde{X}, \widetilde{Y})P - \frac{1}{2}(X\omega_P(\widetilde{Y}) + \omega_P(\widetilde{Y})X).$$

The SMAT $(\mathcal{S}, \nabla, h)$ is a space of distant parallelism.

$(R = R^* = 0,\ T^* = 0,\ \text{but } T \neq 0)$

3 Pre-contrast functions

M : a manifold (an open domain in R^n)
$\rho(p, Z_q)$: a function on $M \times TM$
$\underline{\rho[X_1 \cdots X_i | Y_1 \cdots Y_j Z] : \text{a function on } M}$ defined by

$$\rho[X_1 \cdots X_i | Y_1 \cdots Y_j Z](r) := (X_1)_{(p)} \cdots (X_i)_{(p)} (Y_1)_{(q)} \cdots (Y_j)_{(q)} \rho(p, Z_q)||_{\substack{p=r \\ q=r}}$$

For example,

$$\begin{aligned}
\rho[\,|XZ](r) &= X_{(q)} \rho(p, Z_q)||_{\substack{p=r \\ q=r}} \\
\rho[XY|Z](r) &= X_{(p)} Y_{(q)} \rho(p, Z_q)||_{\substack{p=r \\ q=r}} \\
\rho[XY|ZV](r) &= X_{(p)} Y_{(p)} Z_{(q)} \rho(p, Z_q)||_{\substack{p=r \\ q=r}} \\
&\vdots
\end{aligned}$$

Definition 3.1
$\rho : M \times TM \to \mathrm{R}$: a pre-contrast function on M

$$\overset{\text{def}}{\Longleftrightarrow} \quad \begin{aligned}
&(1)\ \rho(p, f_1 X_1 + f_2 X_2) = f_1 \rho(p, X_1) + f_2 \rho(p, X_2), \\
&(2)\ \rho[|X] = 0 \ (\text{i.e. } \forall r \in M, \rho(r, X_r) = 0), \\
&(3)\ h(X, Y) := -\rho[X|Y] \quad \text{is non-degenerate.}
\end{aligned}$$

Example 3.2
$\rho(p, q)$: *contrast function* $\Longrightarrow X_q \rho(p, q)$: *pre-contrast function*

Proposition 3.3
We can define affine connections ∇ and ∇^* by

$$\begin{aligned} h(\nabla^*_X Y, Z) &= -\rho[XY|Z], \\ h(Y, \nabla_X Z) &= -\rho[Y|XZ]. \end{aligned}$$

Moreover, ∇, ∇^* *: mutually dual with respect to h.*
∇^* *: torsion-free*

(Proof)

$$\begin{aligned} Xh(Y,Z) &= -X\rho[Y|Z] = -\rho[XY|Z] - \rho[Y|XZ] \\ &= h(\nabla^*_X Y, Z) + h(Y, \nabla_X Z) \\ h(\nabla^*_X Y - \nabla^*_Y X, Z) &= -\rho[XY|Z] + \rho[YX|Z] \\ &= -\rho[[X,Y]|Z] \;=\; h([X,Y], Z) \end{aligned}$$

Lemma 3.4
$\rho(X_p, q)$: *a pre-contrast function on M*
$\Longrightarrow$ (M, ∇, h) **is a quasi-statistical manifold.**

Tensor fields B and B^* are defined by

$$\begin{aligned} h(D^*(X,Y)Z,V) &:= -\rho[XYZ|V] \\ h(V,D(X,Y)Z) &:= -\rho[V|XYZ] \end{aligned}$$

$$\begin{aligned} B(X,Y)Z &:= D_{X,Y}Z - \nabla_X\nabla_Y Z \\ B^*(X,Y)Z &:= D^*_{X,Y}Z - \nabla^*_X\nabla^*_Y Z \end{aligned}$$

$$(\ h(V,B(X,Y)Z) \ = \ -\rho[V|XYZ] + \rho[V|\nabla_X\nabla_Y Z]\ )$$

B : **the Bartlett tensor**
B^* : **the dual Bartlett tensor**

Theorem 3.5
R, R^* : the curvature tensors of ∇, ∇^*, respectively.

$$\Longrightarrow \qquad \begin{aligned} R(X,Y)Z &= B(Y,X)Z - B(X,Y)Z, \\ R^*(X,Y)Z &= B^*(Y,X)Z - B^*(X,Y)Z. \end{aligned}$$

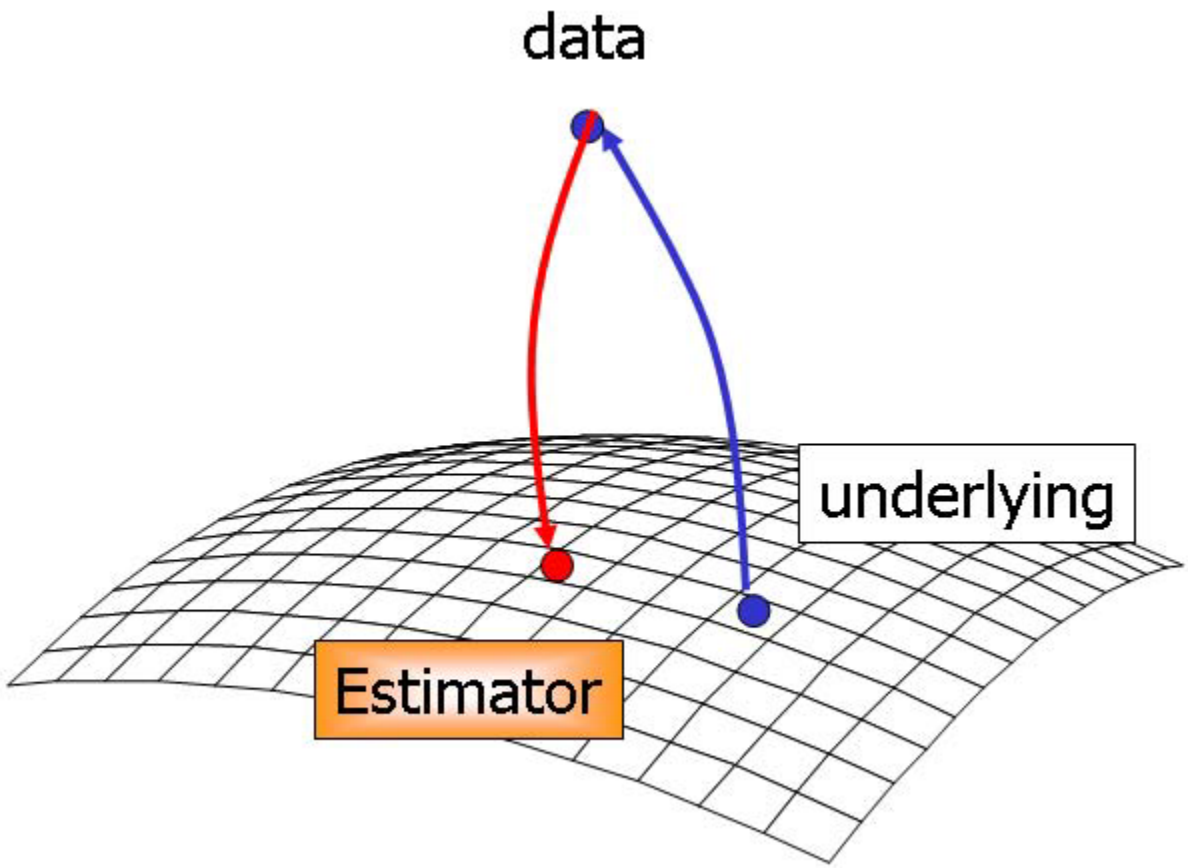
4 Geometry of non-conservative estimating functions

Statistical inference for curved exponential families

S : an exponential family

M : a curved exponential family embedded into S

$x_1, \cdots, x_N$: N independent observations of the random variable x distributed to $p(x; u) \in M$



Statistical inference for curved exponential families

S : an exponential family

M : a curved exponential family embedded into S ($\dim M = m$)

$x_1, \cdots, x_N$: N independent observations of the random variable x distributed to $p(x; u) \in M$

Given $x^N = (x_1, \cdots, x_N)$, a function L on U can be defined by

$$L(u) = p(x_1; u) \cdots p(x_N; u) = \prod_{i=1}^{n} p(x_i; u)$$

We call L a likelihood function.

$$\log L(u) = \log p(x_1; u) + \cdots + \log p(x_N; u) = \sum_{i=1}^{N} \log p(x_i; u)$$

We say that a statistic is the maximum likelihood estimator if it maximizes the likelihood function:

$$L(\hat{u}) = \max_{u \in U} L(u) \qquad \left(\Longleftrightarrow \qquad \log L(\hat{u}) = \max_{u \in U} \log L(u) \right)$$

Hence, the estimating equation is

$$\frac{\partial}{\partial u^a} \log L(u) = 0 \qquad (a = 1, \ldots, m)$$

$$
\begin{aligned}
\bar{x} &= \frac{1}{N}\sum x_i \quad \text{(the sample mean of } x^N\text{)} \\
\hat{\eta}_i &= \frac{1}{N}\sum_{j=1}^{N} F_i(x_j) \quad \text{(the sample mean of the random variable } F_i\text{.)} \\
\phi(\theta) &= E_\theta[\log p(\theta)] \quad (-\phi(\theta) \text{ is the entropy of } p(\theta))
\end{aligned}
$$

Then the log likelihood is given by

$$
\begin{aligned}
\log L(u) &= \sum_{j=1}^{N} \log p(x_j; u) = \sum_{j=1}^{N} \left\{ \sum_{i=1}^{n} F_i(x_j)\theta^i(u) - \psi(\theta(u)) \right\} \\
&= \sum_{i=1}^{n} N \left\{ \hat{\eta}_i \theta^i(u) - \psi(\theta(u)) \right\}.
\end{aligned}
$$

On the other hand, the Kullback-Leibler divergence is given by

$$
\begin{aligned}
\rho_{KL}(p(\hat{\eta}), p(u)) &= \phi(\hat{\eta}) + \psi(\theta(u)) - \sum_{i=1}^{n} \hat{\eta}_i \theta^i(u) \\
&= \phi(\hat{\eta}) - \frac{1}{N}\log L(u).
\end{aligned}
$$

The maximum likelihood estimation $\hat{u}$ is the point in M which minimizes the divergence from $p(\hat{\eta})$.

Estimation of voter transition probabilities

$\boxed{\text{Votes cast}}$ (in the k-th constituency, $k = 1, \ldots, N$)

		Election 2		
	Party	C	L	Total
Election 1	C	X_{1k}	$m_{1k} - X_{1k}$	m_{1k}
	L	X_{2k}	$m_{2k} - X_{2k}$	m_{2k}
	Total	Y_k	$m_k - Y_k$	m_k

$X_{1k} \sim B(m_{1k}, \theta^1)$

$X_{2k} \sim B(m_{2k}, \theta^2)$

$X_{1k} \perp\!\!\!\perp X_{2k}$

X_{1k} and X_{2k} are not observed

We want to estimate the voter transition probabilities,

θ^1, θ^2 : the probabilities that a voter who votes for parties C, L in Election 1, votes for C in Election 2, respectively,

from the observed total Y_k of voters who vote for party C in Election 2.

Each cast is carried out individually, but we can observe the marginals only.

The standard maximum likelihood method does not work.

Remark: $B(m, \theta)$ is the binomial distribution. $P(x_k) = {}_m\mathrm{C}_k \theta^k (1-\theta)^{m-k}$

Regular parametric estimation

$$\rho_{KL}\left(p(y;\theta')||p(y;\theta)\right) = \int p(y;\theta')\log\frac{p(y;\theta')}{p(y;\theta)}dy :$$

KL-divergence (contrast function)

$$s(y;\theta) = \{s^i(y;\theta)\}, s^i(y;\theta) = \frac{\partial}{\partial\theta^i}\log p(y;\theta) : \text{score function for } \theta$$

$$\rho\left((\partial_i)_\theta, p(y;\theta')\right) = -\int s^i(y;\theta)p(y;\theta')dy :$$ (trivial) pre-contrast function

		Election 2		
	Party	C	L	Total
Election 1	C	X_{1k}	$m_{1k} - X_{1k}$	m_{1k}
	L	X_{2k}	$m_{2k} - X_{2k}$	m_{2k}
	Total	Y_k	$m_k - Y_k$	m_k

$(k = 1, \ldots N)$

$X_{1k} \sim B(m_{1k}, \theta^1)$

$X_{2k} \sim B(m_{2k}, \theta^2)$

$X_{1k} \perp\!\!\!\perp X_{2k}$

X_{1k} and X_{2k} are not observed

<u>Quasi-score functions</u>

$$q^i(y;\theta) = \sum_{k=1}^{N}\frac{m_{ik}\{y_k - \mu_k(\theta)\}}{V_k(\theta)} \qquad (i = 1, 2)$$

$$\mu_k(\theta) = E[Y_k] = m_{1k}\theta^1 + m_{2k}\theta^2,$$

$$V_k(\theta) = V[Y_k] = m_{1k}\theta^1(1-\theta^1) + m_{2k}\theta^2(1-\theta^2)$$

Regular parametric estimation

$\rho\left((\partial_i)_\theta, p(y;\theta')\right) = -\int s^i(y;\theta)p(y;\theta')dy$: (trivial) pre-contrast function

Pre-contrast function

$$\rho\left((\partial_i)_\theta, p(y;\theta')\right) = -\sum_{k=1}^{N} q^i(y_k;\theta)p(y_k;\theta') \qquad (\partial_i)_\theta = \left(\frac{\partial}{\partial\theta^i}\right)_{p(y;\theta)}$$

Induced geometric structure (SMAT)

Riemannian metric: $(g_{ij}(\theta)) = \sum_{k=1}^{N} \frac{1}{V_k(\theta)} \begin{pmatrix} m_{1k}^2 & m_{1k}m_{2k} \\ m_{1k}m_{2k} & m_{2k}^2 \end{pmatrix}$

Dual affine connections:

$$\Gamma_{ij,l}(\theta) = E_\theta\left[\{\partial_i q^j(y;\theta)\}s^l(y;\theta)\right] = \sum_{k=1}^{N} \frac{1-2\theta^i}{V_k(\theta)^2} m_{ik}m_{jk}m_{lk}, \left(\frac{\partial q^1}{\partial\theta^2} \neq \frac{\partial q^2}{\partial\theta^1}\right)$$

$$\Gamma^*_{ij,l}(\theta) = \sum_y \{\partial_i\partial_j p(y;\theta)\}q^l(y;\theta) = \sum_{k=1}^{N} \frac{m_{lk}}{V_k(\theta)}\{\partial_i\partial_j\mu_k(\theta)\} = 0$$

$$(R = R^* = 0, \quad T^* = 0, \quad \text{but} \quad T \neq 0)$$

5 A toy example: questionnaire to students

The survey was carried out in my linear algebra class and calculus class.
(We regard each class as a constituency, $N = 2$)

Question 1: Where is your home town? (First ballot)

Nagoya city	suburbs of Nagoya	Gifu, Mie	somewhere else
Nagoya city	outside of Nagoya city		

Question 2: Where is your place of residence? (Second ballot)

Nagoya city	suburbs of Nagoya	Gifu, Mie	somewhere else
Nagoya city	outside of Nagoya city		

We infer transposition probabilities θ^1: city $\Longrightarrow$ city
θ^2: outside $\Longrightarrow$ city

		Election 2		
	Party	C	L	Total
Election 1	C	X_{1k}	$m_{1k} - X_{1k}$	m_{1k}
	L	X_{2k}	$m_{2k} - X_{2k}$	m_{2k}
	Total	Y_k	$m_k - Y_k$	m_k

$(k = 1, 2)$

$X_{1k} \sim B(m_{1k}, \theta^1)$

$X_{2k} \sim B(m_{2k}, \theta^2)$

$X_{1k} \perp\!\!\!\perp X_{2k}$

X_{1k} and X_{2k} are not observed

linear algebra		place of residence		
		city	outside	total
home town	city	*	*	14
	outside	*	*	37
	total	23	28	51

calculus	place of residence		
	city	outside	total
city	*	*	6
outside	*	*	45
total	19	32	51

estimations using quasi-score functions

$$\hat{\theta}^1 = \frac{83}{102} \simeq 0.8137, \qquad \hat{\theta}^2 = \frac{32}{102} \simeq 0.3137$$

Since we used clickers, we could observe each cast.

linear algebra		place of residence		
		city	outside	total
home town	city	14	0	14
	outside	9	28	37
	total	23	28	51

calculus	place of residence		
	city	outside	total
city	5	1	6
outside	14	31	45
total	19	32	51

Sample ratios from observed data

$$\bar{\theta}^1 = \frac{19}{20} = 0.95, \qquad \bar{\theta}^2 = \frac{23}{82} \simeq 0.2805$$

Appendix: Optimal transport estimator

Suppose that we have no information about constituencies.

		Election 2			
	Party	C	L	Total	$X_1 \sim B(m_1, \theta^1)$
Election 1	C	X_1	$m_1 - X_1$	m_1	$X_2 \sim B(m_2, \theta^2)$
	L	X_2	$m_2 - X_2$	m_2	$X_1 \perp\!\!\!\perp X_2$
	Total	$\bar{m}_1$	$\bar{m}_2$	m	X_1 and X_2 are not observed

We want to estimate the voter transition probabilities,

θ^1, θ^2 : **the probabilities that a voter who votes for parties C, L in Election 1 votes for C in Election 2, respectively.**

Appendix: Optimal transport estimator

		Election 2		
	Party	C	L	Total
Election 1	C	$*$	$*$	m_1
	L	$*$	$\min(m_2, \bar{m}_2)$	m_2
	Total	$\bar{m}_1$	$\bar{m}_2$	m

$X_1 \sim B(m_1, \theta^1)$

$X_2 \sim B(m_2, \theta^2)$

$X_1 \perp\!\!\!\perp X_2$

X_1 and X_2 are not observed

We suppose that $\bar{m}_2 < m_2$.

		Election 2		
	Party	C	L	Total
Election 1	C	m_1	0	m_1
	L	$m_2 - \bar{m}_2$	$\bar{m}_2$	m_2
	Total	$\bar{m}_1$	$\bar{m}_2$	m

$$\hat{\theta}^1 = 1$$

$$\hat{\theta}^2 = \frac{m_2 - \bar{m}_2}{m_2}$$

This 2×2 contingency table implies the optimal transport mapping form the marginal distribution $\{m_1, m_2\}$ to $\{\bar{m}_1, \bar{m}_2\}$.

linear algebra		place of residence		
		city	outside	total
home town	city	14	0	14
	outside	9	28	37
	total	23	28	51

calculus	place of residence		
	city	outside	total
city	6	0	6
outside	13	32	45
total	19	32	51

Optimal transport estimations

$$\hat{\theta}^1 = \frac{20}{20} = 1, \qquad \hat{\theta}^2 = \frac{22}{82} \simeq 0.2683$$

Since we used clickers, we could observe each cast.

linear algebra		place of residence		
		city	outside	total
home town	city	14	0	14
	outside	9	28	37
	total	23	28	51

calculus	place of residence		
	city	outside	total
city	5	1	6
outside	14	31	45
total	19	32	51

Sample ratios from observed data

$$\bar{\theta}^1 = \frac{19}{20} = 0.95, \qquad \bar{\theta}^2 = \frac{23}{82} \simeq 0.2805$$

Statistical inferences

Dually flat spaces

$(x_1, x_2, \ldots, x_N)$: N-independent observations
$L(\theta) = p(x_1; \theta)p(x_2; \theta) \cdots p(x_N; \theta)$
$\Longrightarrow$ Maximum likelihood estimator, Dually flat spaces

Non-integrable geometry

$(x_1, \ldots, x_N)$: N-independent events, but we cannot observe.
Likelihood functions do not exist in the sense above.
$\Longrightarrow$ Non-conservative estimating function
Statistical manifolds admitting torsion