

•セミパラメトリック統計推論の 情報幾何

外生変数に雑音がある場合

甘利俊一
理化学研究所脳科学総合研究センター

線形回帰: Semiparametrics

$$(x_1, y_1)$$

$$(x_2, y_2)$$

$$\vdots$$

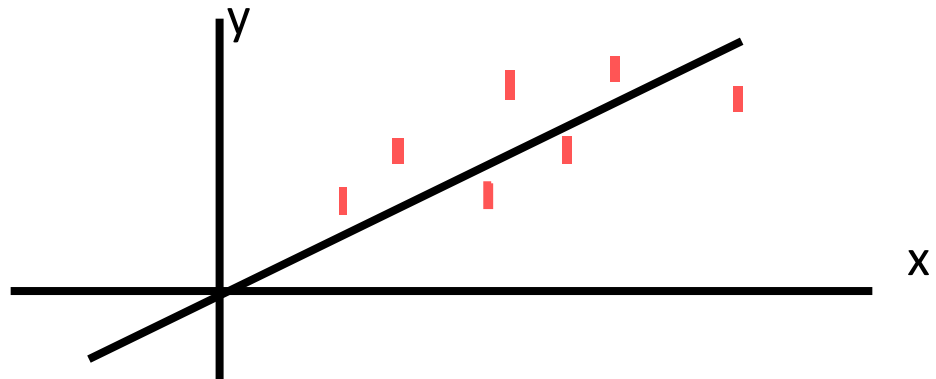
$$(x_n, y_n)$$

$$y = \theta x$$

$$x_i = \xi_i + \varepsilon_i$$

$$y_i = \theta \xi_i + \varepsilon'_i$$

$$\varepsilon_i, \varepsilon'_i \square N(0, \sigma^2)$$



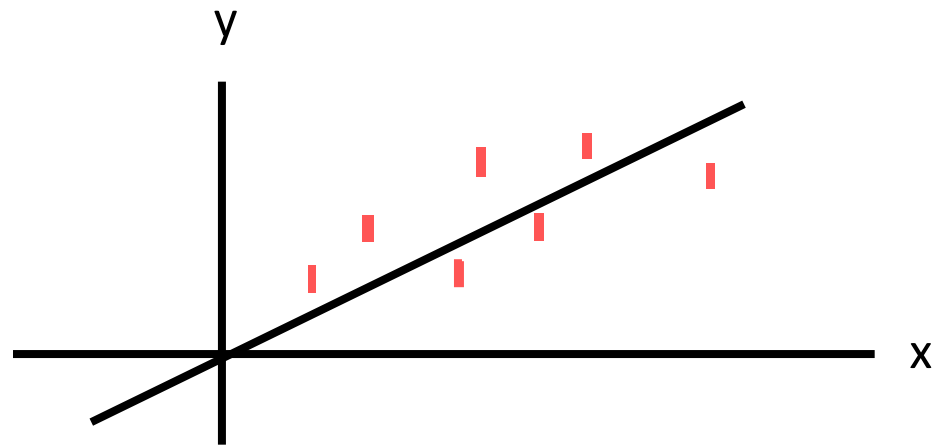
Semiparametric 統計モデル

$$M = \{p(x, \theta, r)\}$$

$$p(x, \theta) = r(x - \theta)$$

linear relation

$$y = \theta x$$



$$\begin{cases} y_i = \theta \xi_i + \varepsilon_i \\ x_i = \xi_i + \varepsilon_i' \end{cases}$$

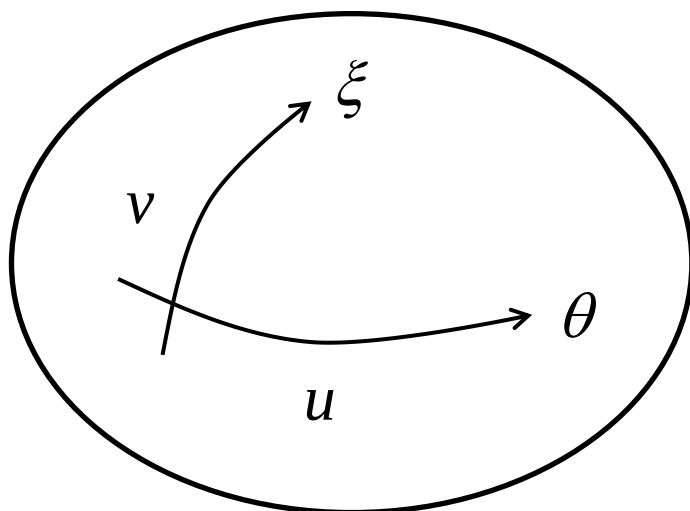
$$p(x, y; \theta) = \int p(x, y; \xi, \theta) r(\xi) d\xi$$

mle, least square, total least square

攪乱変数のある統計モデル

スコア関数

$p(x, \theta, \xi):$



$$u(x, \theta, \xi) = \frac{\partial}{\partial \theta} \log p(x, \theta, \xi)$$

$$v(x, \theta, \xi) = \frac{\partial}{\partial \xi} \log p(x, \theta, \xi)$$

同時推定とFisher情報量

$$x_1, \dots, x_n$$

最尤推定

$$\sum u(x_i, \theta, \xi) = 0$$

$$\sum v(x_i, \theta, \xi) = 0$$

Fisher情報行列

$$g = \begin{bmatrix} g_{\theta\theta} & g_{\theta\xi} \\ g_{\theta\xi} & g_{\xi\xi} \end{bmatrix} = \begin{bmatrix} E[u^2] & E[uv] \\ E[uv] & E[v^2] \end{bmatrix}$$

情報量損失と直交性

$$E[uv] = \langle u, v \rangle = 0 \quad \text{スコア直交の時}$$

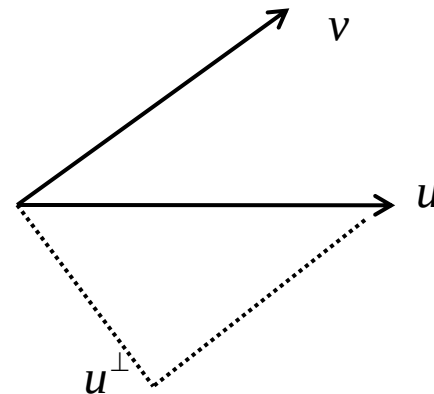
$$(g^{-1})_{\theta\theta} = 1 / g_{\theta\theta}$$

非直交:

$$(g^{-1})_{\theta\theta} = E[(u^\perp)^2]$$

$$u = u^\perp + cv \quad ; \quad c = \frac{E[uv]}{E[v^2]}$$

$$E[(u^\perp)^2] = E[u^2] - c^2 E[v^2]$$



無限個の外乱母数–Neyman–Scott問題

$$p(x, \theta, \xi)$$

$$x_1 \sim p(x, \theta, \xi_1)$$

$$x_2 \sim p(x, \theta, \xi_2)$$

$$\vdots$$

$$x_n \sim p(x, \theta, \xi_n)$$

未知パラメータ:

$$\theta, \xi_1, \dots, \xi_n$$

Semiparametric Statistical Model

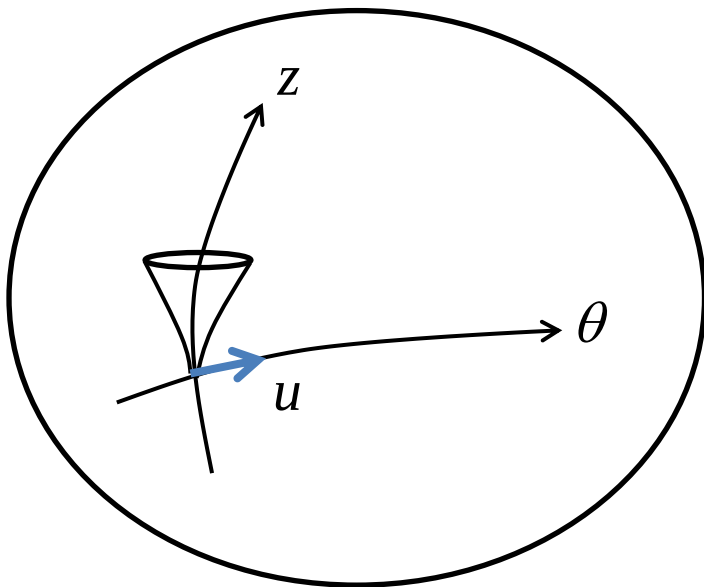
$$p(x, \theta, \xi) : x_i \in \xi_i \\ \xi_1, \dots, \xi_n \in z(\xi)$$

$$x_i \in p(x, \theta, z) = \int p(x, \theta, \xi) z(\xi) d\xi$$

θ : parameter of interest

z : nuisance parameter 関数次元

幾何学



$$u(x, \theta, z)$$

$$v(x, \theta, z)$$

$$v_t = \frac{\partial}{\partial t} \log p(x, \theta, z(\xi, t))$$

$$v(x, \theta, \xi) = \frac{\partial}{\partial \xi} \log p$$

直交スコア

$$V^N = \{v_t(x, \theta, z)\} = \{v(x, \theta, \xi)\} : \text{mixture model}$$

$$u^\perp = u - \int c_t v_t dt$$

Projected score 直交スコア

推定関数法 estimating function

$$f(x, \theta):$$

$$E_{\theta, Z} [f(x, \theta)] = 0: \text{unbiased}$$

$$\sum_{i=1}^n f(x_i, \theta) = 0: \text{推定方程式}$$

$$E \left[(\hat{\theta} - \theta)^2 \right] = \frac{1}{n} \frac{E[f^2]}{E[(\partial_{\theta} f)^2]}$$

平行移動

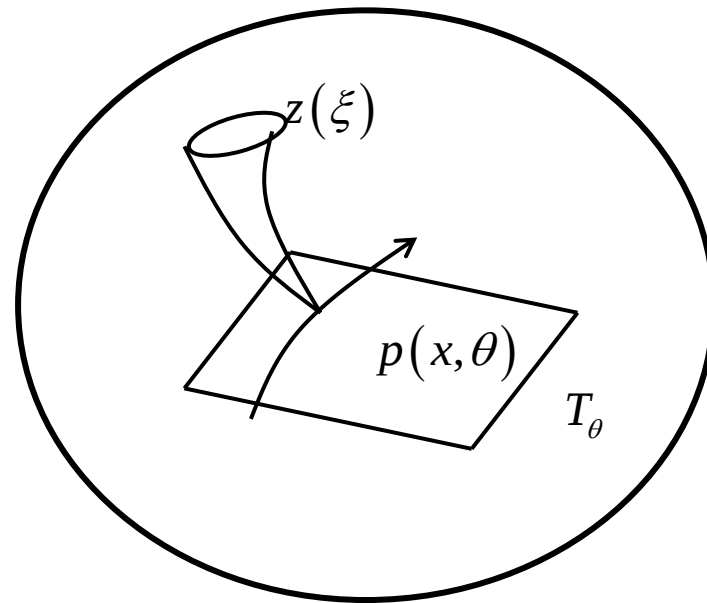
$$T_{\theta} = \left\{ r(x) \mid E_{\theta} [r(x)] = 0 \right\}$$

$$\text{内積} \langle r_1, r_2 \rangle = E_{\theta} [r_1(x) r_2(x)]$$

$$\prod_z^e r(x) = r(x) E_{\theta, z'} [r(x)]$$

$$\prod_z^m r(x) = \frac{p(x, \theta, z)}{x, \theta, z'} r(x)$$

$$\langle r_1, r_2 \rangle_{z'} = \left\langle \prod_z^e r_1, \prod_z^m r_2 \right\rangle$$



推定関数 $f(x, \theta)$

$$e\text{-不変}: E[f(x, \theta)] = 0$$

$$\prod_z^{z'} f(x, \theta) = f$$

$$m\text{-直交性}: \langle v, f \rangle = 0$$

$$\left\langle \prod_z^m v, f \right\rangle = 0$$

$$T_\theta = T_\theta^I \oplus T_\theta^N \oplus T_\theta^A$$

$u^I(x, \theta, z)$: 最適推定関数

悪い推定量

$$z = z(\xi, \eta)$$

$$p(x, \theta, z(\xi, \eta)) \rightarrow \hat{\theta}, \hat{\eta}$$

$$u^2(x, \theta, \hat{\eta}) \quad \times$$

$$u^I(x, \theta, \hat{\eta}) \quad \text{OK}$$

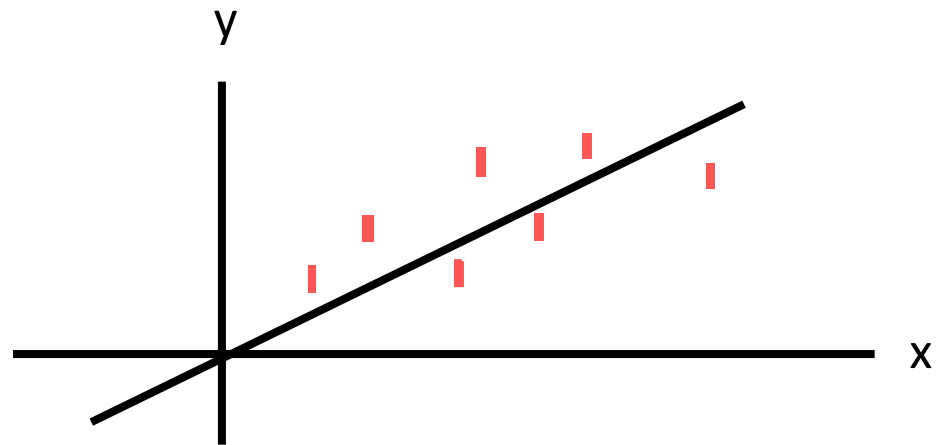
Semiparametric Statistics

$$M = \{p(x, \theta, r)\}$$

$$p(x, \theta) = r(x - \theta)$$

linear relation

$$y = \theta x$$



$$\begin{cases} y_i = \theta \xi_i + \varepsilon_i \\ x_i = \xi_i + \varepsilon_i' \end{cases}$$

$$p(x, y; \theta) = \int p(x, y; \xi, \theta) r(\xi) d\xi$$

mle, least square, total least square

Linear Regression: Semiparametrics

$$(x_1, y_1)$$

$$(x_2, y_2)$$

$$\vdots$$

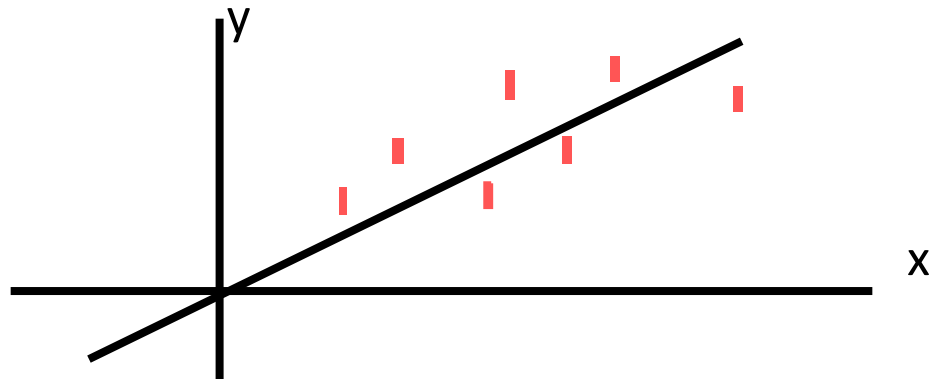
$$(x_n, y_n)$$

$$y = \theta x$$

$$x_i = \xi_i + \varepsilon_i$$

$$y_i = \theta \xi_i + \varepsilon_i'$$

$$\varepsilon_i, \varepsilon_i' \square N(0, \sigma^2)$$



Statistical Model

$$p(x, y | \theta, \xi) = c \exp \left\{ -\frac{1}{2} (x - \xi)^2 - \frac{1}{2} (y - \theta \xi)^2 \right\}$$

$$p(x_i, y_i | \theta, \xi_i) : \theta, \xi_1, \dots, \xi_n$$

$$p(x, y | \theta) = \int p(x, y | \theta, \xi) Z(\xi) d\xi$$

———— semiparametric

Semiparametric statistical model

$$x_1, x_2, \dots \square p(x, \theta, Z)$$

Estimating function

$$y(x, \theta) \quad E_{\theta, Z} [y(x, \theta)] = 0$$

$$E_{\theta', Z} [y(x, \theta)] \neq 0$$

Estimating equation

$$\sum y(x_i, \theta) = 0 \quad \Rightarrow \hat{\theta}$$

Example of estimating functions

$$f(x, y; \theta) = k(x + \theta y)(y - \theta x)$$

$$\int c \exp \left\{ -\frac{1}{2}(x - \xi)^2 - \frac{1}{2}(y - \theta' \xi)^2 \right. \\ \left. \times (y - \theta x) k(x + \theta y) Z(\xi) dx dy d \right\}$$

$$= 0, \quad \theta = \theta'$$

$$\neq 0, \quad \theta \neq \theta'$$

$$\sum k(x_i + \theta y_i)(y_i - \theta x_i) = 0$$

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum y(x_i, \theta) \\ = \frac{1}{\sqrt{n}} \sum y(x_i, \theta) + \frac{\sqrt{n}}{n} \sum y'(x_i, \theta) (\theta - \theta) \end{aligned}$$

$$\theta - \theta = \frac{1}{\sqrt{n}} \frac{\varepsilon}{E[y'(x, \theta)]}$$

$$\begin{aligned}
 p(x, y; \theta, \xi) &= \exp \left\{ -\frac{1}{2\sigma^2} (x - \xi)^2 - \frac{1}{2\sigma^2} (y - \theta\xi)^2 + c \right\} \\
 &= \exp \left\{ \frac{\xi}{\sigma^2} (x + \theta y) + c(x, y) - \psi(\theta, \xi) \right\}
 \end{aligned}$$

$$s(x, y; \theta) = x + \theta y$$

$$p(x, \theta, Z) = \int \exp \{ \xi s(x, y; \theta,) + c - \psi \} Z(\xi) d\xi$$

$$\sum f(x_i, y_i; \theta) = 0$$

$$f(x, y; \theta) = (x + \theta y + c)(y - \theta x)$$

$$c = \frac{\bar{\xi} \sigma^2}{\bar{\xi^2} - (\bar{\xi})^2} \quad \begin{cases} \bar{\xi} = 1 \\ \bar{\xi^2} = 2 \end{cases}$$

$$c = 0 : V = \frac{1}{n} \frac{(2 + \sigma^2) \sigma^2}{4} \quad : \frac{3}{4}$$

$$c = 1 : V = \frac{1}{n} \left(1 - \frac{1}{\sigma^2 + 2} \right) \sigma^2 \quad : \frac{2}{3}$$

$$c = \infty : v = \frac{1}{n} \sigma^2 \quad : 1$$

$$z(\xi) \sim N(\mu_\xi, \sigma_\xi^2)$$

$$u^I(x, \theta, Z) = (x + \theta y + c)(y - \theta x)$$

$$c = \mu_\xi / \sigma_\xi^2$$

$$f(x, \theta; c) = (x + \theta y + c)(y - \theta x)$$

$$c = \mu_\xi / \sigma_\xi^2$$

$$\mu_\xi = \frac{1}{n} \sum x_i$$

$$\sigma_\xi^2 = \frac{1}{n} \sum x_i^2 - \left(\mu_\xi \right)^2 - \sigma^2$$

$$u(x, y; \theta, Z) = \frac{\partial}{\partial \theta} \log p$$

$$= \partial_{\theta} s E[\xi | s] \dots$$

$$v(x, y; \theta, Z) = E[f(\xi) | s]$$

$$= k \{s(x, y; \theta)\}$$

$$u^I(x, y, \theta) = u - E[u | s]$$

$$= k(x + \theta y)(y - \theta x)$$

$$p(x, \theta, Z)$$

$$Z(\xi) = Z(\xi, \eta)$$

$$p(x, \theta, Z(\cdot, \eta)) \quad \text{does not work}$$

$$u^I(x, \theta, Z) \quad \text{estimating function}$$

Generalization

$$y = \boldsymbol{\theta} \cdot \boldsymbol{x}$$

$$\boldsymbol{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i$$

$$y_i = \boldsymbol{\theta} \cdot \boldsymbol{\xi}_i + \varepsilon_i'$$

$$f(\boldsymbol{x}, y; \boldsymbol{\theta}) = (y - \boldsymbol{\theta} \cdot \boldsymbol{x})k(\boldsymbol{x} + y\boldsymbol{\theta})$$

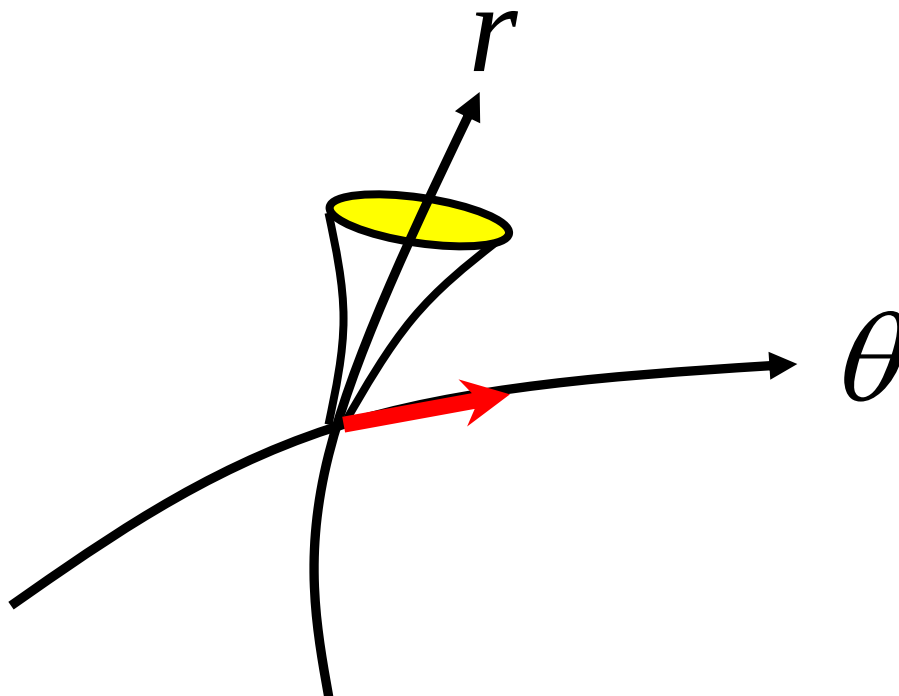
$$f_c(\boldsymbol{x}, y; \boldsymbol{\theta}) = (y - \boldsymbol{\theta} \cdot \boldsymbol{x})(\boldsymbol{x} + y\boldsymbol{\theta} + \boldsymbol{c})$$

$$\sum (y_i - \boldsymbol{\theta} \cdot \boldsymbol{x}_i)(\boldsymbol{x}_i + y_i\boldsymbol{\theta} + \boldsymbol{c}) = 0$$

$$\boldsymbol{c}^* = \sigma^2 V_{\boldsymbol{\xi}}^{-1} \overline{\boldsymbol{\xi}}$$

Fibre bundle of probability distributions

function space



Least squares?

$$L(\theta) = \sum (y_i - \theta x_i)^2 \rightarrow \min \quad : \hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\frac{1}{n} \sum \frac{y_i}{x_i}, \quad \frac{\sum y_i}{\sum x_i}$$

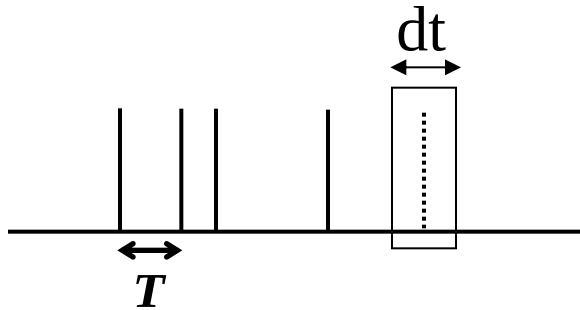
mle, TLS

$$\sum (y_i - \theta x_i)(\theta y_i + x_i) = 0$$

Neyman-Scott

Poisson process

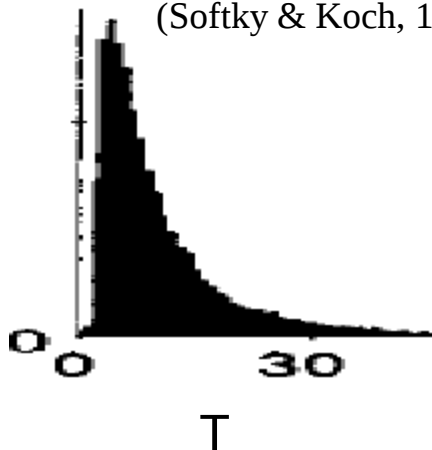
Poisson Process: Instantaneous firing rate is constant over time.



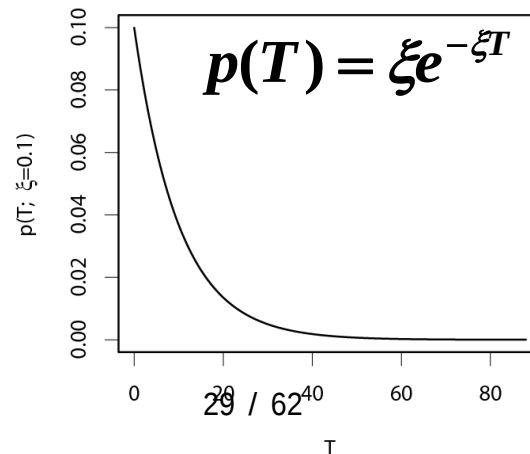
For every small time window dt , generate a spike with probability ξdt .

Cortical Neuron

(Softky & Koch, 1993)



Poisson Process



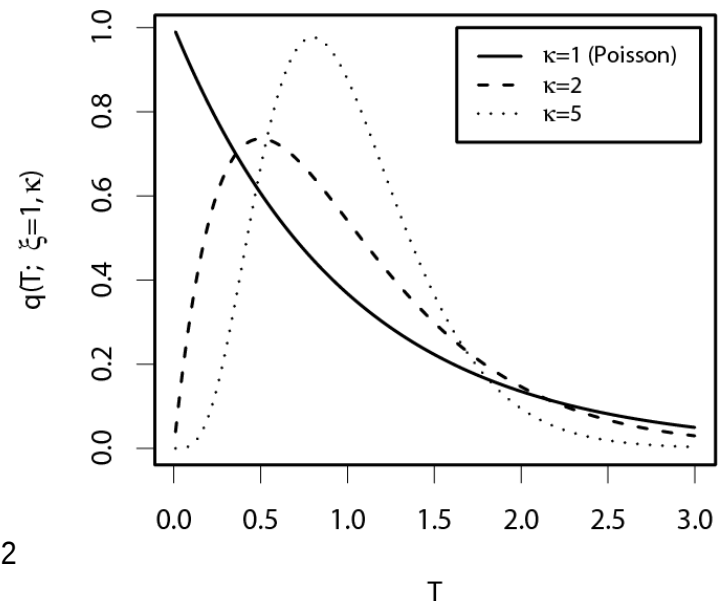
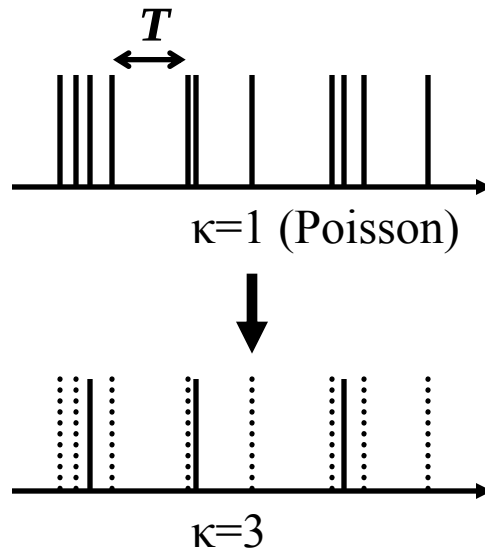
Poisson process cannot explain inter-spike interval distributions.

Gamma distribution

Gamma Distribution: Every κ -th spike of the Poisson process is left.

$$q(T; \xi, \kappa) = \frac{(\xi \kappa)^\kappa}{\Gamma(\kappa)} T^{\kappa-1} e^{-\xi \kappa T}.$$

Two parameters $\begin{cases} \xi: \text{Firing rate} \\ \kappa: \text{Irregularity} \end{cases}$



Gamma distribution

$$f(T) = \frac{(r\kappa)^\kappa}{\Gamma(\kappa)} T^{\kappa-1} \exp\{-r\kappa T\}$$

$\kappa = 1$: Poisson

$\kappa \rightarrow \infty$: regular

Integrate-and fire

Markov model

temporal correlations

$$\bigcirc: \quad x(1)x(2)\dots x(N)$$

independent with firing probability $r(t)$

→ spike counts: Poisson
ISI: exponential

renewal:

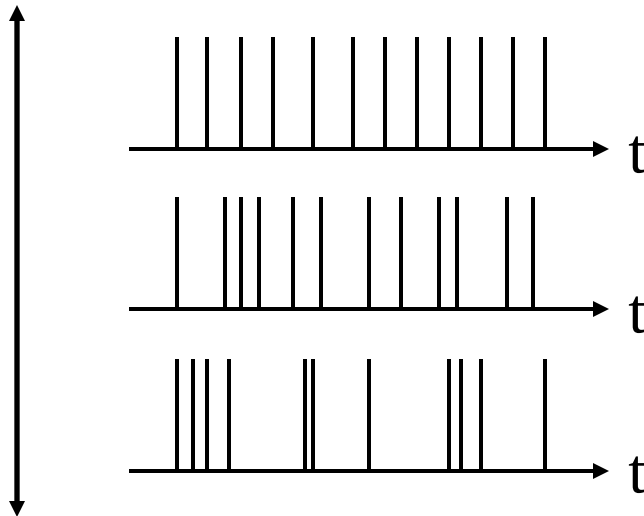
$$r(t) = k(t - t_i) \quad : \text{last spike}$$

ISI distribution $f(t)$

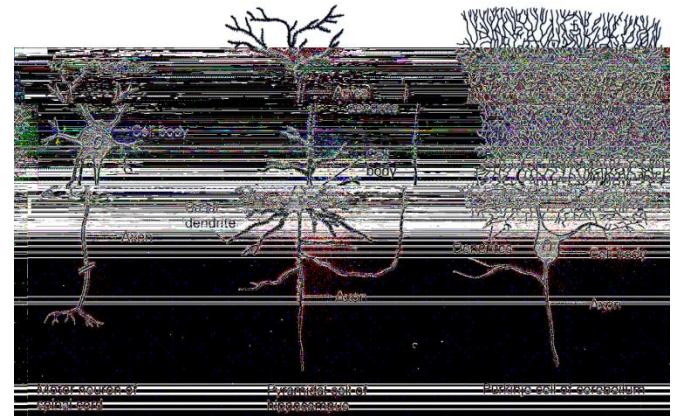
$$f(T) = ck(T) \exp \left\{ \int_0^T k(t) dt \right\}$$

Irregularity κ is unique to individual neurons.

Regular (large κ)



Irregular (small κ)

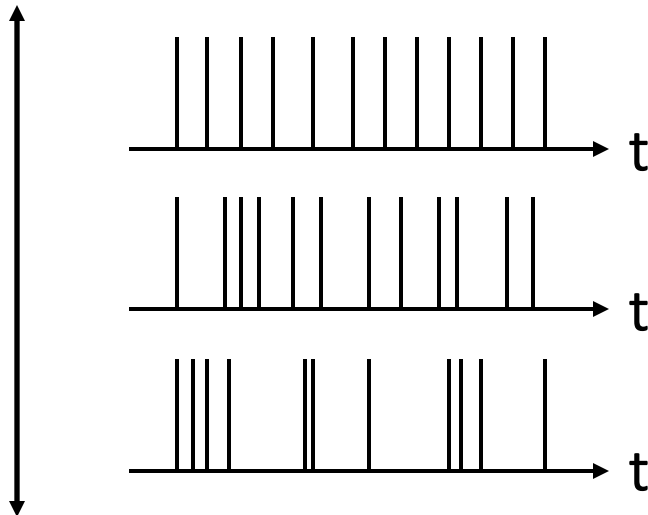


Irregularity varies among neurons.
(Baker & Lemon 2000; Shinomoto et.al., 2003)

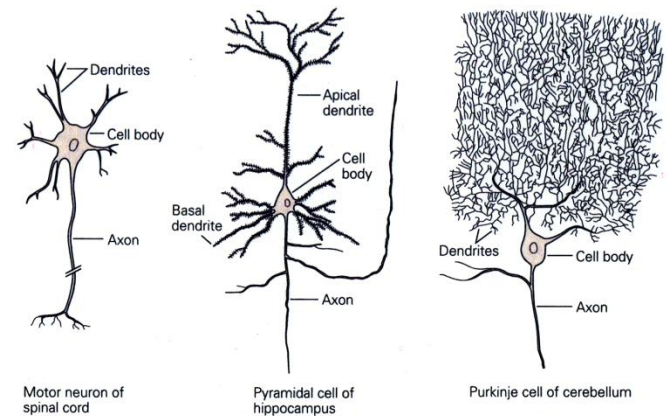
➔ **We assume that κ is independent of time.**

Irregularity κ is unique to individual neurons.

Regular (large κ)



Irregular (small κ)



Irregularity varies among neurons.

→ **We assume that κ is independent of time.**

Information geometry of estimating function

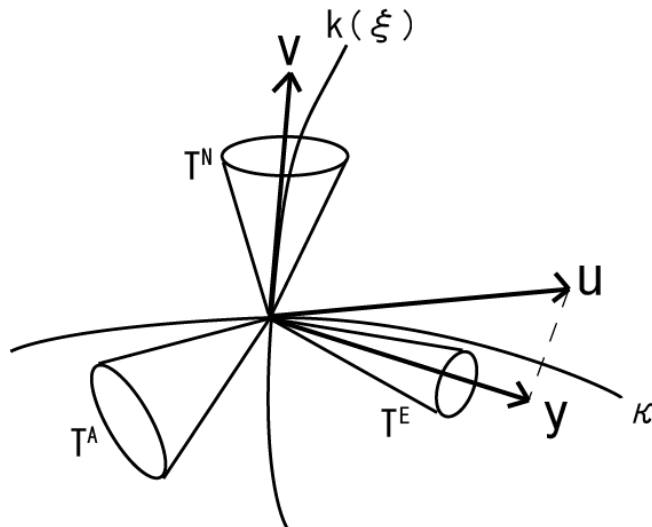
• Estimating function $y(T, \kappa)$:

$$\longrightarrow \sum_{l=1}^N y(T_l; \kappa) = 0$$

• Maximum likelihood Method:

$$\longleftrightarrow \frac{d}{d\kappa} \log p(T_1) \cdots p(T_N) = \sum_{l=1}^N u(T_l; \kappa) = 0$$

$$E[y(T, \kappa)] = 0$$



(See poster for details) 35 / 62

How to obtain an estimating function y :

$$\text{Score functions} \begin{cases} u(T; \kappa, k) \equiv \frac{d \log p(T; \kappa, k)}{d\kappa} \\ v(T; \kappa, k) \equiv \frac{\delta \log p(T; \kappa, k)}{\delta k(\xi)} \end{cases}$$

$$\longrightarrow y = u - \frac{\langle u \cdot v \rangle}{\langle v \cdot v \rangle} v$$

temporal correlations

$$\bigcirc: \quad x(1)x(2)\dots x(N)$$

independent with firing probability $r(t)$

→ spike counts: Poisson
ISI: exponential

renewal:

$$r(t) = k(t - t_i) \quad : \text{last spike}$$

ISI distribution $f(t)$

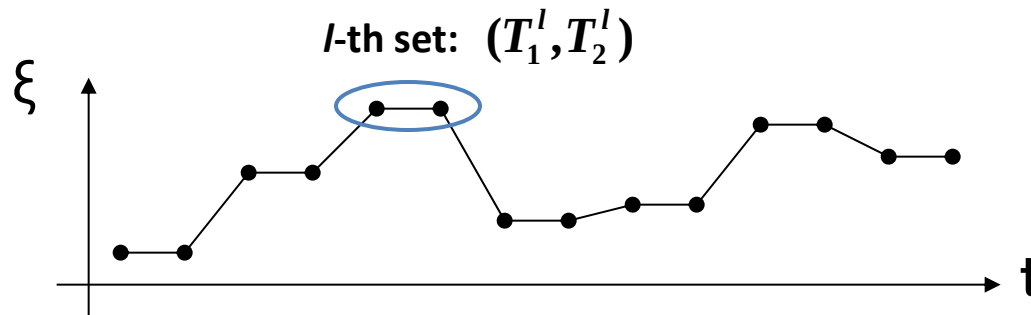
$$f(T) = ck(T) \exp \left\{ \int_0^T k(t) dt \right\}$$

Estimation of κ by estimating functions

1. No estimating function exists if the neighboring firing rates are different.
2. $m(\geq 2)$ consecutive observations must have the same firing rate.

Example: $m=2$

Model: $p(T_1, T_2; \kappa, k) = \int_0^\infty q(T_1; \xi, \kappa) q(T_2; \xi, \kappa) k(\xi) d\xi$

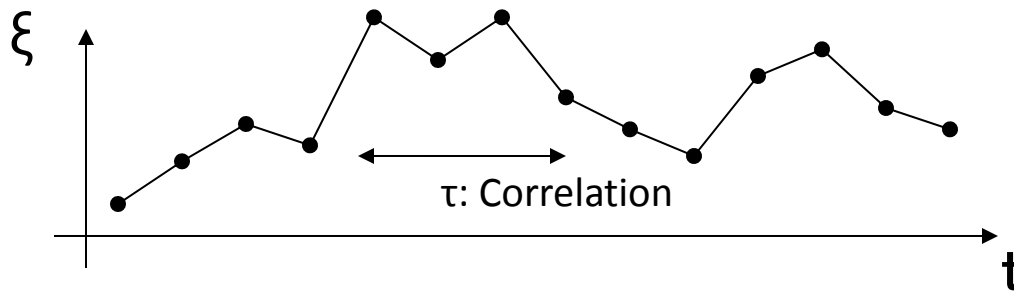


**Estimating function:
($E[y]=0$)**

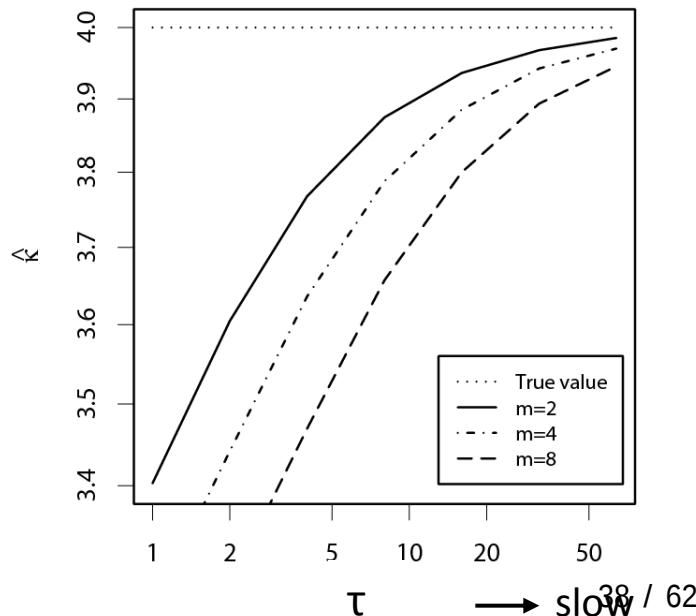
$$y = \log \frac{T_1 T_2}{(T_1 + T_2)^2} + 2\phi(2\kappa) - 2\phi(\kappa)$$

$$\bar{y} = \frac{1}{N} \sum_{l=1}^N \log \frac{T_1^l T_2^l}{(T_1^l + T_2^l)^2} + 2\phi(2\kappa) - 2\phi(\kappa) = 0$$

Case of $m=1$ (spontaneous discharge)



Firing rate continuously changes and is driven by Gaussian noise.

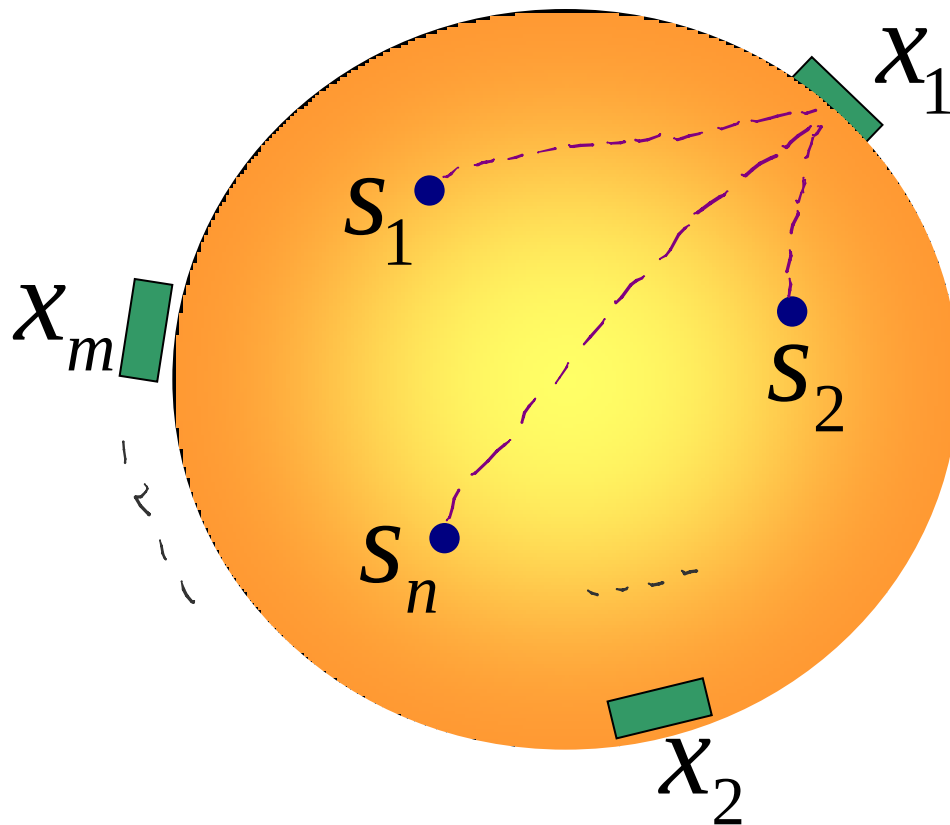


κ can be approximated if the firing rate changes slowly.

Estimating function ($m=2$):

$$\bar{y} = \frac{1}{N} \sum_{l=1}^N \log \frac{T_1^l T_2^l}{(T_1^l + T_2^l)^2} + 2\phi(2\kappa) - 2\phi(\kappa) = 0$$

mixture and unmixture of independent signals

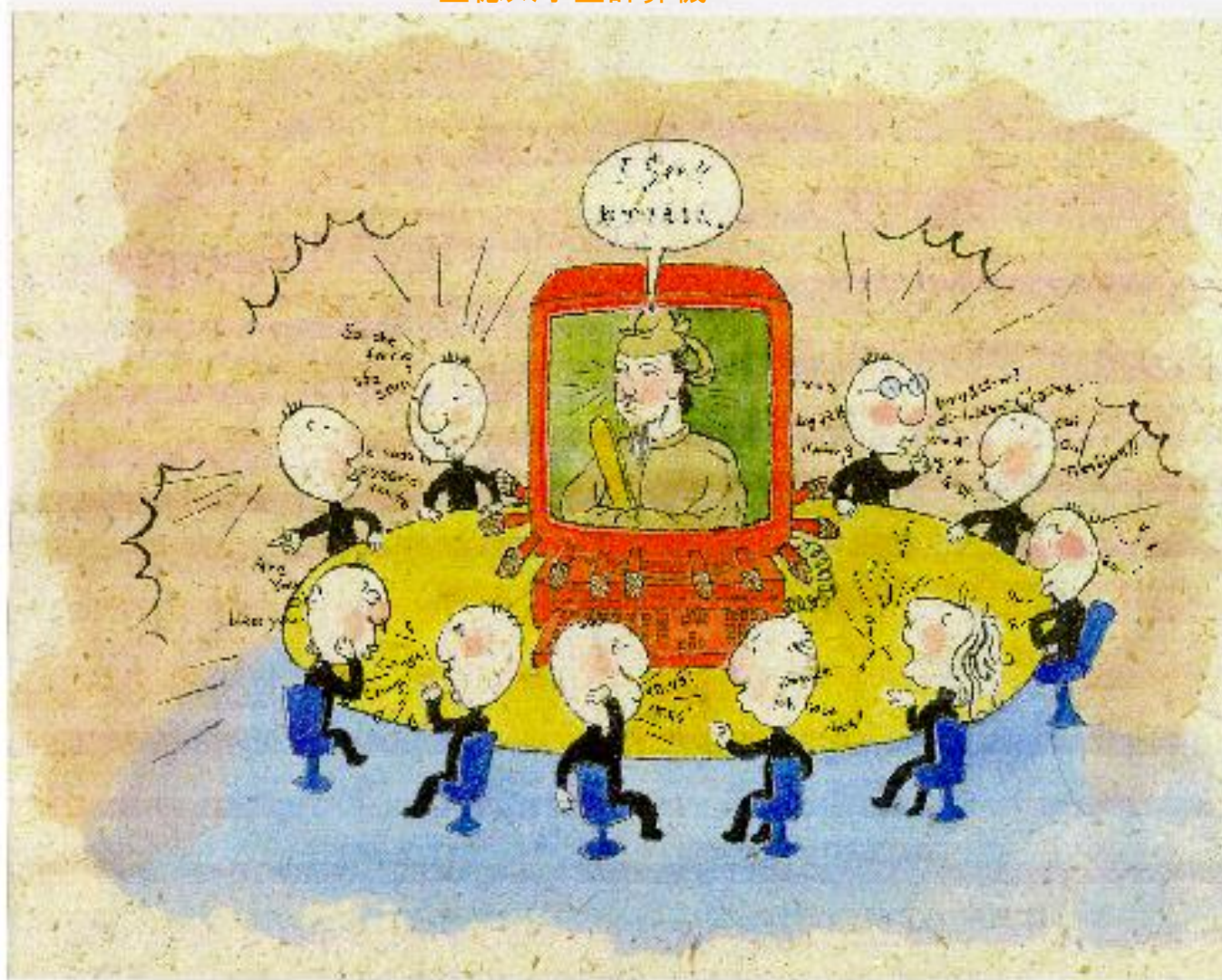


$$x_i = \sum_{j=1}^n A_{ij} s_j$$

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$



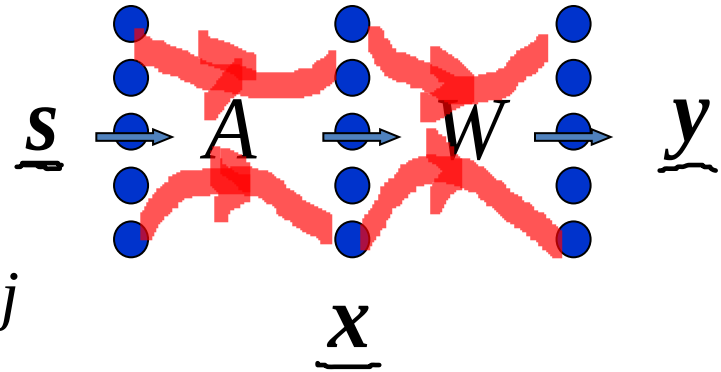
聖德太子型計算機



Independent Component Analysis

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

$$x_i = \sum A_{ij} s_j$$



$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad \mathbf{W} = \mathbf{A}^{-1}$$

observations: $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)$

recover: $\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(t)$

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \rightarrow \quad \mathbf{y} = \mathbf{W}\mathbf{x} \quad : \quad \mathbf{W} = \mathbf{A}^{-1}$$

observations: $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)$

$\mathbf{A}$: unknown matrix

$\mathbf{s}$: unknown

$$r(\mathbf{s}) = r_1(s_1)r_2(s_2)\dots r_n(s_n)$$

independent distribution

$r(\mathbf{s})$: unknown

$$E[\mathbf{s}] = 0$$

$$\Delta \mathbf{W} = -\eta \frac{\partial l(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}}$$

cost function:

degree of non-independence

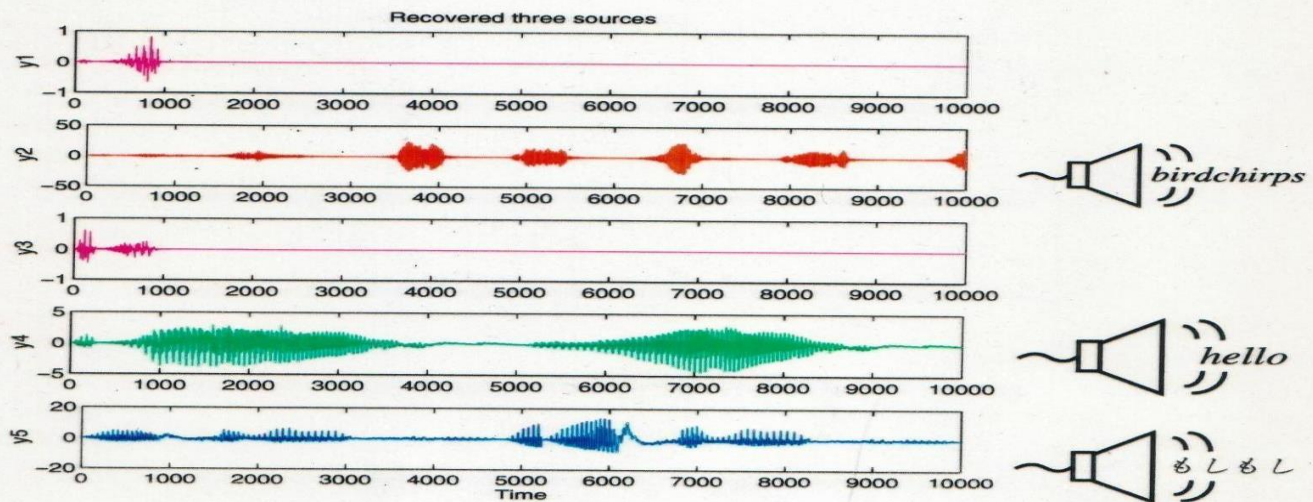
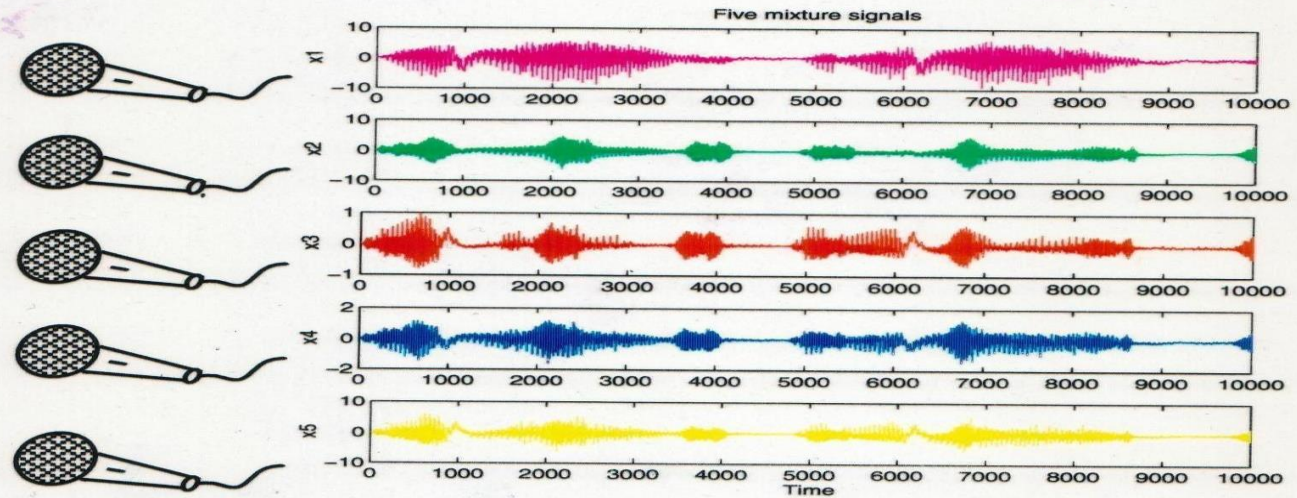
Semiparametric Statistical Model

$$p(\mathbf{x}; \mathbf{W}, r) = |\mathbf{W}| r(\mathbf{W}\mathbf{x})$$

$$\mathbf{W} = \mathbf{A}^{-1}, \quad r(s): \text{ unknown} \quad r = \prod r_i$$

$$x(1), x(2), \dots, x(t)$$

Cocktail party experiment



- 5 microphones (sensors) and only 3 speakers

Example of color image separation :



Five original images (but unknown to the neural net)



Five mixed images for separation

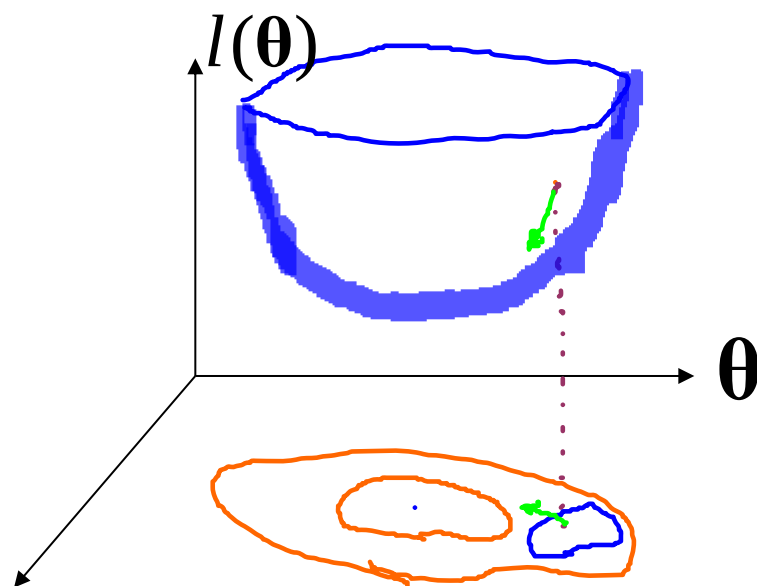


Natural Gradient

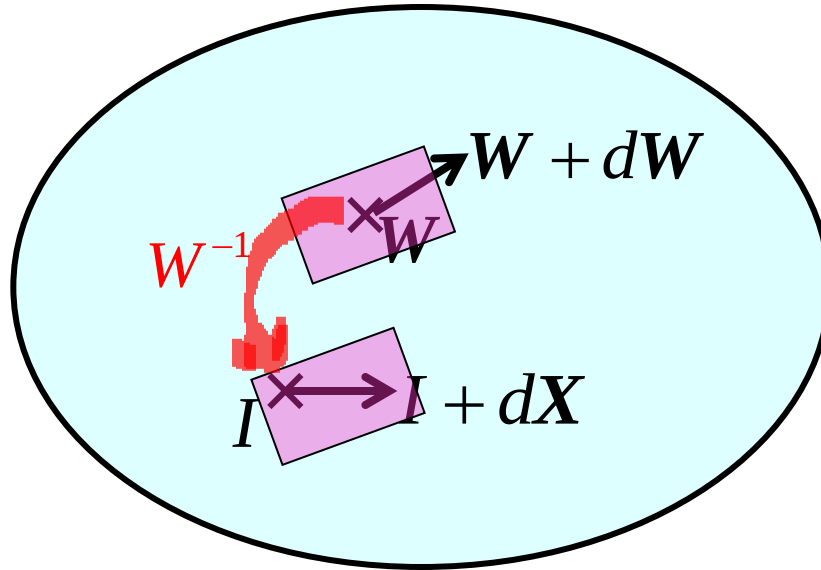
$$\max \quad dl = l(\boldsymbol{\theta} + d\boldsymbol{\theta}) - l(\boldsymbol{\theta})$$

$$|d\boldsymbol{\theta}|^2 = \varepsilon$$

$$\nabla l = G^{-1}(\boldsymbol{\theta}) \nabla l$$



Space of Matrices : Lie group



$$dX = dW W^{-1}$$

$$|dW|^2 = \text{tr}(dX dX^T) = \text{tr}(dW W^{-1} W^{-T} dW^T)$$

$$\nabla l = \frac{\partial l}{\partial W} W^T W$$

dX : non-holonomic basis

Natural Gradient

$$\Delta \mathbf{W} = -\eta \frac{\partial l(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}$$

Natural Gradient Learning Algorithm

$$\Delta \mathbf{W} = -\eta \mathbf{F}(\mathbf{y})$$

$$\mathbf{F}(\mathbf{y}) = \{I - \boldsymbol{\varphi}(\mathbf{y}) \mathbf{y}^T\} \mathbf{W} \quad ; \quad \mathbf{y} = \mathbf{W} \mathbf{x}$$

$$\Delta W_{ij} = -\eta \sum_k \{\delta_{ik} - \varphi_i(y_i) y_k\} W_{jk}$$

$$\varphi_i(y_i) = -\frac{d}{dy} \log q_i(y_i)$$

Estimating Functions

$$\Delta W = -\eta F(y, W)$$

$$E_{W,r} [F(y, W')] \begin{cases} = 0, & W' = W \\ \neq 0, & W' \neq W \end{cases}$$

estimating equation

$$\sum_t F(y_t, W) = 0 \quad y_t = Wx_t$$

learning

$$\Delta W_t = -\eta F(W_t x_t)$$

Admissible class

$$F(y, W) = \{I - \varphi(y)y^T\}W$$

$$\tilde{F} = R(W)F(y, W) \quad \{\alpha\varphi(y_i)y_j - \varphi(y_j)y_i\}W$$

$$\sum R(W)F(Wx_t) = 0: \text{ estimating equation}$$

$$\text{on-line learning} : \Delta W_t = -\eta R(W_t)F(W_t x_t)$$

$$\text{canonical estimating function: } \frac{\partial \tilde{F}}{\partial W} = I$$

Canonical Estimating Function ---Newton Method

$$\delta \mathbf{W}_t = -\eta \left(\frac{\partial \mathbf{F}}{\partial \mathbf{W}} \right)^{-1} \mathbf{F}(\mathbf{y}_t)$$

$$\mathbf{F} = \nabla l$$

$$(\nabla \nabla l)^{-1} \nabla l : \text{Newton Method}$$

Stabilization—Newton Method

$$\begin{aligned}\Delta W_{ij}(t) &= \eta_t \frac{1}{\sigma_i^2 \sigma_j^2 k_i k_j - 1} \left[\sigma_i^2 k_j \varphi(y_i) y_j - \varphi(y_j) y_i \right] W \\ &= \eta_t \{ \alpha \varphi(y_i) y_j - \varphi(y_j) y_i \} W\end{aligned}$$

$$\sigma_i^2 = E[y_i^2], \quad \kappa_i = E[\varphi_i'(y_i)]$$



efficient algorithm

adaptive method

Stability Analysis

$$\dot{\mathbf{W}}(t) = -\eta E[\mathbf{F}(\mathbf{y})]$$

$$\delta \dot{\mathbf{W}}(t) = -\eta E\left[\frac{\partial \mathbf{F}}{\partial \mathbf{W}}\right] \delta \mathbf{W}$$

$$\mathbf{R} = \frac{\partial \mathbf{F}}{\partial \mathbf{W}} \quad : \quad R_{ij,kl}$$

$$R_{AA'} = \begin{matrix} & A & A' \\ \begin{matrix} A \\ A' \end{matrix} & \begin{bmatrix} \kappa_i & 1 \\ 1 & \kappa_j \end{bmatrix} \end{matrix}$$

$$A = (ij)$$

$$A' = (ji)$$

$$\kappa_i = E[\varphi'_i(s_i)]$$

$$\kappa_i \kappa_j > 1 \quad , \quad \kappa_i > 0$$

Estimating function

$$y = \mathbf{w} \cdot \mathbf{x}$$

$$\mathbf{f}(\mathbf{x}, \mathbf{w}) = \varphi(y)\mathbf{x} - y\varphi(y)\mathbf{w}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \{ \varphi(y_t)\mathbf{x}_t - y_t\varphi(y_t)\mathbf{w}_t \}$$

Newton's Method

FastICA

Estimating Function in Noisy Case

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

$$\mathbf{F} = (F_{ab})$$

$$F_{ij} = y_i^3 y_j - 3v_{ii} y_i y_j - 3v_{ij} y_i^2 + 3v_{ii} v_{ij}$$

$$v_{ij} = E[n_i n_j]$$

$$\Delta v_{ij} = -c(v_{ij} - y_i y_j)$$

Temporal Correlation

$$\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(t), \dots$$

$$\mathbf{\Lambda}(\tau) = \overline{\mathbf{s}(t)\mathbf{s}(t-\tau)} = \text{diag}[\lambda_1, \dots, \lambda_n]$$

$$\begin{aligned}\mathbf{V}(\tau) &= \overline{\mathbf{x}(t)\mathbf{x}(t-\tau)^T} = \overline{A\mathbf{s}(t)\mathbf{s}(t-\tau)^T}A^T \\ &= A\mathbf{\Lambda}(\tau)A^T\end{aligned}$$

Simultaneous Diagonalization

$$W\mathbf{V}(\tau)W^T = \mathbf{\Lambda}(\tau) \quad \tau = 0, 1, 2, \dots$$

Estimating Function in Temporally Correlated Case

$$\mathbf{F}(\mathbf{y}, W) = \mathbf{I} - \boldsymbol{\varphi}(\tilde{\mathbf{y}}) \circ \mathbf{B}(z^{-1}) \mathbf{y}^T$$

$$: \varphi_i(\tilde{y}_i) B_i(z^{-1}) y_t$$

$$\tilde{y}_i = B_i(z^{-1}) y_i = \sum_k b_{ik} y_i(t-k)$$

**class of admissible estimating functions:
maximum likelihood; joint diagonalization**

Spatial mixing

$$\begin{cases} x_1 = a_1 s_1 + \cdots + a_n s_n \\ x_2 = a'_1 s_1 + \cdots + a'_n s_n \end{cases}$$

.....

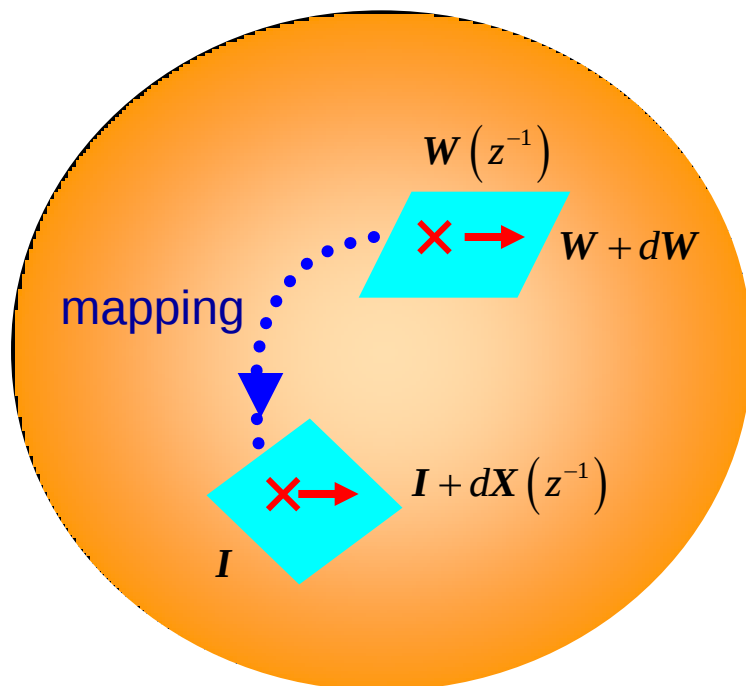
$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

Temporal mixing—convolution

$$x(t) = a_1 s(t) + a_2 s(t-1) + \cdots + a_n s(t-n)$$

$$\mathbf{x}(t) = \int \mathbf{A}(t-\tau) \mathbf{s}(\tau) d\tau$$

Manifold of Linear Systems



$$dX(z^{-1}) = dW(z^{-1})W^{-1}(z^{-1})$$

Metric Structure

{ Lie group
Fisher information

$$\tilde{\nabla} f(\mathbf{W}) = \frac{\partial f}{\partial \mathbf{W}} \circ W^T(z)W(z^{-1})$$

Manifold of Linear Systems

Topological Structure

$$\begin{cases} \mathbf{v}(t+1) = \mathbf{F}\mathbf{v}(t) + \mathbf{G}\mathbf{s}(t) \\ \mathbf{x}(t) = \mathbf{H}\mathbf{v}(t) \end{cases}$$

$$\mathbf{A}(z^{-1}) = \mathbf{H}(\mathbf{z}\mathbf{I} - \mathbf{F})^{-1}\mathbf{G}$$

$$\underbrace{\{\mathbf{F}, \mathbf{G}, \mathbf{H}\}}_{\text{parameter}} / T \quad \begin{cases} T\mathbf{F}T^{-1} \\ \mathbf{H}T, \quad T^{-1}\mathbf{G} \end{cases}$$

Non-Holonomic Basis

$$dX = dW W^{-1}$$

$$\Delta W = -\eta \frac{\partial l}{\partial W} W^T W$$



$$\Delta X_t = -\eta \frac{\partial l}{\partial X}$$

$$\Delta X = -\eta \left(\frac{\partial^2 l}{\partial X \partial X} \right)^{-1} \frac{\partial l}{\partial X}$$

