平成24年度　東京大学理学部情報科学科

月曜2限10：30～12：00

理7-214

知識処理論

2012年5月7日（月）

井元清哉
東京大学医科学研究所
ヒトゲノム解析センター
DNA情報解析分野
http://bonsai.hgc.jp/~imoto
imoto@ims.u-tokyo.ac.jp

正則化法とベイズ

$$
\begin{aligned}
S_\lambda(\boldsymbol{\beta}) &= (\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) + \lambda\boldsymbol{\beta}^T\boldsymbol{\beta} \\
&\Rightarrow -\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta}) - \frac{\lambda'}{2}\boldsymbol{\beta}^T\boldsymbol{\beta} \\
&\Rightarrow \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right\}\cdot\left(\frac{\lambda'}{2\pi}\right)^{(p+1)/2}\exp\left(-\frac{\lambda'}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right) \\
&= L(\boldsymbol{\beta},\sigma^2)\cdot p(\boldsymbol{\beta}) \\
&\propto p(\boldsymbol{\beta}\mid\boldsymbol{x})
\end{aligned}
$$

$$L(\boldsymbol{\beta},\sigma^2)\cdot p(\boldsymbol{\beta}) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right\}\cdot\left(\frac{\lambda}{2\pi}\right)^{(p+1)/2}\exp\left(-\frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)$$

$$\propto p(\boldsymbol{\beta}\,|\,\boldsymbol{x})$$

$$-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})-\frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}$$

$$=-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta}-2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y}+\boldsymbol{y}^T\boldsymbol{y}+\sigma^2\lambda\boldsymbol{\beta}^T\boldsymbol{\beta})$$

$$=-\frac{1}{2\sigma^2}\{\boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{X}+\sigma^2\lambda I)\boldsymbol{\beta}-\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y}+\boldsymbol{y}^T\boldsymbol{y}\}$$

$$=-\frac{1}{2\sigma^2}\{(\boldsymbol{\beta}-\boldsymbol{R}^{-1}\boldsymbol{X}^T\boldsymbol{y})^T\boldsymbol{R}(\boldsymbol{\beta}-\boldsymbol{R}^{-1}\boldsymbol{X}^T\boldsymbol{y})-\boldsymbol{y}^T(\boldsymbol{XR}^{-1}\boldsymbol{X}^T-I)\boldsymbol{y}\}$$

$$p(\boldsymbol{\beta}\,|\,\boldsymbol{x})\sim N_p(\boldsymbol{R}^{-1}\boldsymbol{X}^T\boldsymbol{y},\sigma^2\boldsymbol{R}^{-1})$$

事後分布は正規分布

$$p(\boldsymbol{\beta}\,|\,\boldsymbol{x}) \propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right\} \cdot \left(\frac{\lambda}{2\pi}\right)^{(p+1)/2} \exp\left(-\frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)$$

$$\hat{\boldsymbol{\beta}}_{\mathrm{P}} = \boldsymbol{R}^{-1}\boldsymbol{X}^T\boldsymbol{y} = (\boldsymbol{X}^T\boldsymbol{X} + \sigma^2\lambda I)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$p(\boldsymbol{\beta}\,|\,\boldsymbol{x}) \sim N_p(\boldsymbol{R}^{-1}\boldsymbol{X}^T\boldsymbol{y}, \sigma^2\boldsymbol{R}^{-1})$$

分散 σ^2

データの質・情報量

分散 λ^{-1}

事前知識の確信度

データは高品質、事前知識は曖昧　σ^2: 小、λ^{-1}: 大（λ: 小）

データは情報量小、事前知識は確度高　σ^2: 大、λ^{-1}: 小（λ: 大）

自然共役事前分布

パラメータの事前分布 $p(\boldsymbol{\theta})$ が属する分布族を $\Im$ と書く。

例えば、$\Im = \left\{ N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \Theta_{\boldsymbol{\mu}}, \boldsymbol{\Sigma} \in \Theta_{\boldsymbol{\Sigma}} \right\}$

このとき、ある尤度に対してパラメータの事後分布

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}) \qquad \left(= \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\boldsymbol{x})} \right)$$

もまた $\Im$ に属するとき、$p(\boldsymbol{\theta})$ を自然共役事前分布と呼ぶ。

対応表

尤度	自然共役事前分布
正規分布	正規分布（平均パラメータ） 逆カイ2乗分布（分散パラメータ） $p(\sigma^2) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-\nu/2-1}\exp\left(-\frac{1}{2\sigma^2}\right)$ $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt, \quad (x>0)$
二項分布 $L(\pi) = \begin{pmatrix} n \\ x \end{pmatrix}\pi^x \cdot (1-\pi)^{n-x}$	ベータ分布　$\mathrm{Beta}(\alpha, \beta)$ $p(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha)}\pi^{\alpha-1} \cdot (1-\pi)^{\beta-1}$
多項分布	ディリクレ分布

Bayesian networks で重要（「情報科学とバイオインフォマティクス」で講義）

$$p(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} \cdot (1-\pi)^{\beta-1} \sim \mathrm{Beta}(\alpha, \beta)$$

$$\int L(\pi)p(\pi)d\pi = \int \binom{n}{n} \pi^{x} \cdot (1-\pi)^{n-x} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} \cdot (1-\pi)^{\beta-1} d\pi$$

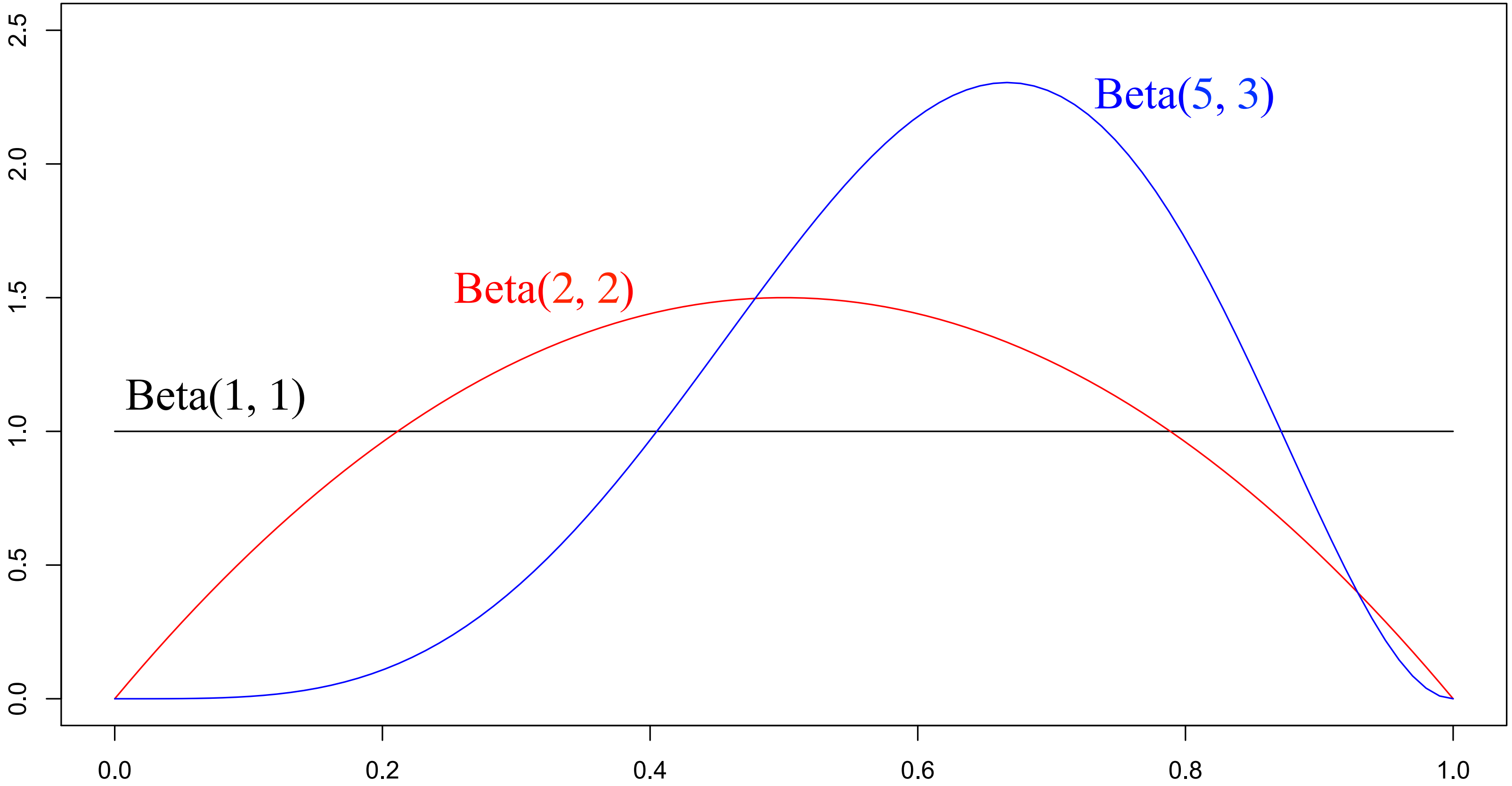
$$= \binom{n}{n} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \int \pi^{x+\alpha-1} \cdot (1-\pi)^{n-x+\beta-1} d\pi$$

$$= \binom{n}{n} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$p(\pi \mid \boldsymbol{x}) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)} \pi^{x+\alpha-1} \cdot (1-\pi)^{n-x+\beta-1} \sim \mathrm{Beta}(x+\alpha, n-x+\beta)$$

データの情報が
入った

ベータ分布



線形重回帰モデル

データ: $(x_{i1}, x_{i2}, \ldots, x_{ip}, y_i)$

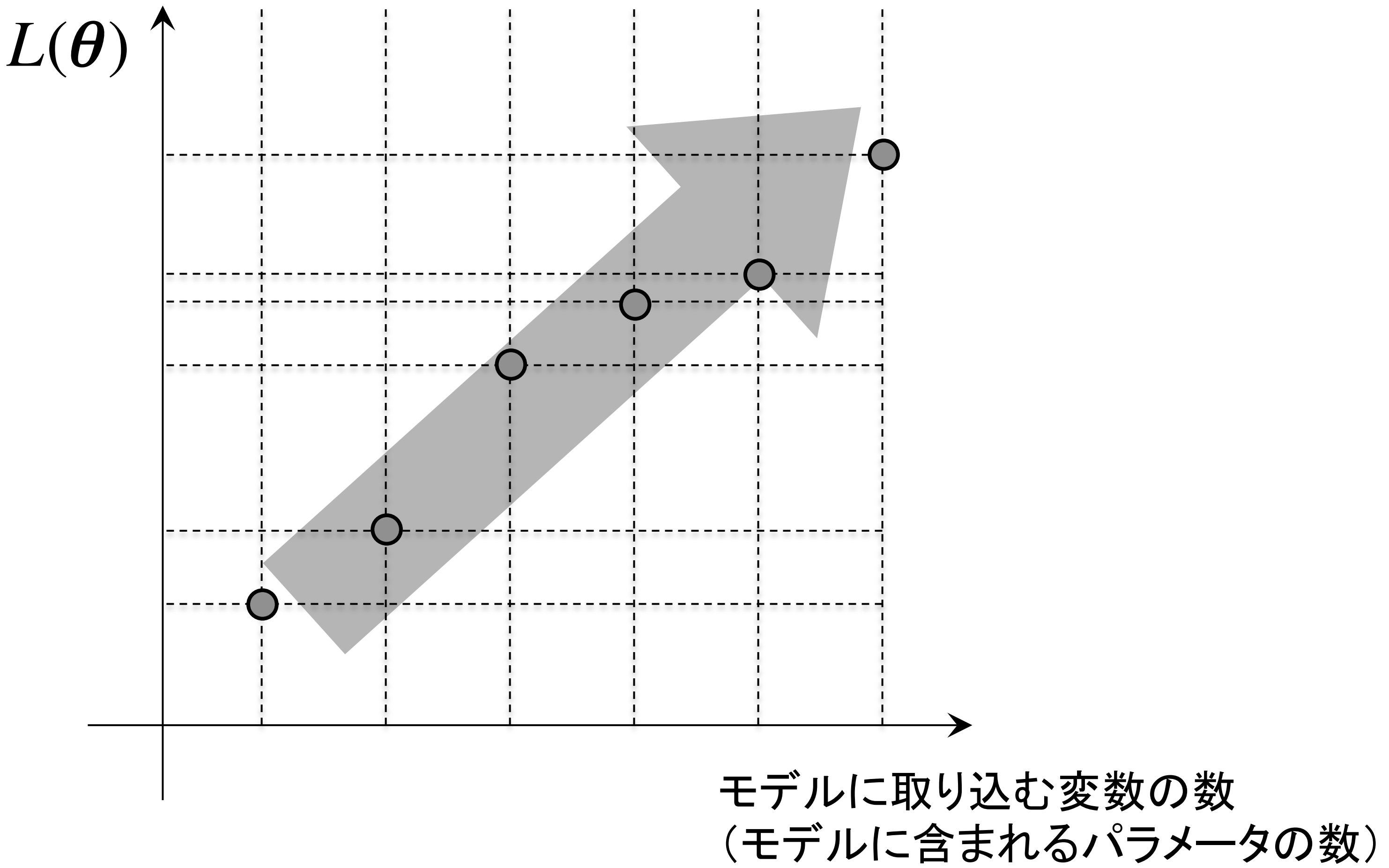
線形重回帰モデル:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{22} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

yを説明(予測)するための変数が p 個あるが
どれが必要なものであろうか？

統計的モデル選択

尤度とモデルの複雑さ



情報量規準

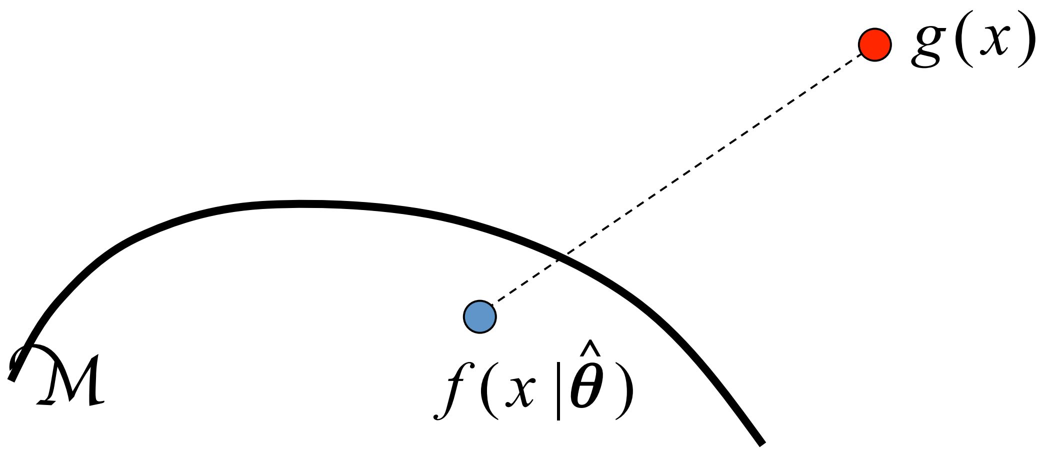
基本となる考え方：

データ： $x_1, x_2, \ldots, x_n \sim f(x \mid \boldsymbol{\theta})$ ー 想定したモデル

データを生成した真の分布： $g(x)$ ー 未知

最尤法により推定されたモデル： $f(x \mid \hat{\boldsymbol{\theta}})$

$f(x \mid \hat{\boldsymbol{\theta}})$ と $g(x)$ の距離が近いほど良いモデル



AIC、TIC

$$\mathrm{AIC} = -2\left\{\sum_{i=1}^{n}\log f(x_i \mid \hat{\boldsymbol{\theta}}) - (\#\ \text{parameters})\right\}$$

Akaike (1973)

－ $\hat{\boldsymbol{\theta}}$ は最尤推定量

－ $\exists \boldsymbol{\theta}^* \in \Theta;\quad f(x \mid \boldsymbol{\theta}^*) = g(x)$

$$\mathrm{TIC} = -2\left\{\sum_{i=1}^{n}\log f(x_i \mid \hat{\boldsymbol{\theta}}) - \mathrm{tr}\{I(\hat{\boldsymbol{\theta}})J(\hat{\boldsymbol{\theta}})^{-1}\}\right\}$$

$$I(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(x_i \mid \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\frac{\partial \log f(x_i \mid \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^T}$$

$$J(\hat{\boldsymbol{\theta}}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(x_i \mid \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}$$

竹内 (1976)

－ $\hat{\boldsymbol{\theta}}$ は最尤推定量

Note: $\exists \boldsymbol{\theta}^* \in \Theta;\quad f(x \mid \boldsymbol{\theta}^*) = g(x) \;\Rightarrow\; \mathrm{TIC} = \mathrm{AIC}$

第22回京都賞（2006年）

赤池博士は、情報数理の基礎概念に基づく、実用性と汎用性の両方を兼ね備えた、統計モデル選択のための規準Akaike Information Criterion(AIC)を提唱し、データとモデル双方の世界を結びつける新しいパラダイムを打ち立てた。AICは現在、数学・統計学以外に、医学・生物学・制御工学・経済学・環境学・地球物理学・社会科学などの広範な分野に利用され、その実用性と信頼性は高く評価され続けている。