

ベイズ推論

電子情報通信学会ソサイエティ大会

AI-2 データ科学とコンピュータ科学の基礎理論と展開

2016年9月20日北海道大学

東京工業大学

渡辺澄夫

この講演の目的

統計的推論が命題論理の推論と異なる点を説明し、
ベイズ推論において解明されていることの概略を述べる。

もくじ

1. 統計的推論は命題論理の推論と何が違うのか
2. 統計的推論では何を知りたいのか
3. 予測誤差と交差検証誤差
4. 総和誤差と自由エネルギー

1. 統計的推論は命題論理の推論と何が本質的に違うのか

なぜ人間は「正しい統計的推論」を求めたのか

数学や物理学では一定の水準の厳密さにおいて「正しい推論」というものが存在している。

- 正しいモデルで正しく推論すれば正しい結論が得られる。
- 間違った結論は間違ったモデルか推論から生まれる。

(例) 連続関数の列が一様収束すれば極限の関数も連続である。

(例) シュレディンガー方程式を解いて得られた波動関数はその絶対値の二乗が量子の存在確率に比例する。

(理由) 統計学でも、「結論を信頼してよい」と言えるだけの厳密な基盤や正しい推論規則があることが強く望まれていた。

しかし「正しい統計的推論」は存在しない

データが与えられたとき、推論結果が必ず正しく
真の分布を推定しているような、そのような正しさの
推論が望まれていた。

(昔の人の意見)「最尤推定は上記の意味での正しい推論である」

(昔の人の意見)「ベイズ推定は上記の意味での正しい推論である」

「どちらが正しいか」という不毛な論争があった。今日でも
論争している人がある。しかし、どちらも正しくはない。

◎ 未知の分布の推定を行う統計学では基盤となる確率空間が確定しないので「正しさ」が定義できない(本質的に不良設定問題である)。

「正しい統計モデル・事前分布」も存在しない

データが与えられたとき、「正しい統計モデル」や「正しい事前分布」を定めるにはどうすべきかという議論も昔から行われてきた。しかしながら

真の分布は不明であり、真の分布についての仮説が何もない場合には、正しい統計モデルや正しい事前分布は存在しない。

◎ 「正しいモデル・正しい事前分布・正しい推論でなければ統計的推測はできない」という考え方の上では、何もできないことになる。

存在しない正しさを探すのは、もうやめよう。

得られた結論が、必ず正しく真の分布を推定できるような統計モデル・事前分布・統計推論は存在しない。

真の分布は不明であり、統計モデルも事前分布も人間が仮に与えたものに過ぎないから、得られた結果は確率1で間違っている。

「正しい統計モデル・事前分布・推論規則」は存在しない。
それにも関わらず

「正しいものは何か」と今日でも多くの人が繰り返し、繰り返し、問い続けている……。

……存在しない正しさを探すのは、もうやめよう。

統計学に何ができるのか(1)

それでは統計学には何もできないのか、といえばそうではない。

推測法を固定したとき、真の分布が不明であっても、
任意の(モデル・事前分布)を用いて得られた結論の
誤差に関しては普遍的な数学的法則が存在する。このことから

ある統計モデル・ある事前分布が与えられたとき、
得られた結論がどの程度間違っているかを数量的に知ることが
できる場合がある(※)。

(※) 真の分布が不明なのに誤差は推定できる(これが統計学の本領)。

(※) 1970年ころまで誰も気づかなかった。

- ◎ 確率空間が不明でも、任意の確率空間で同じ定理が成り立つ。
統計学とは、「どんな確率空間でも成り立つ普遍法則」を見つける学問。

統計学に何ができるのか(2)

統計学には、正しいモデル・正しい事前分布・正しい推論は存在しない。しかし「結論の誤差」については正しく推定する理論や方法を作ることができる。

(注) 数学における推論は、正しく行われているならば、結論も必ず正しいので、結論の誤差を調べるための理論や手続きは必要ない。

(注) 統計学における推論は常に間違っているので、結論の誤差を調べるための理論や手続きが必ず必要になる。

→ 誤差を小さくするモデル・事前分布・推論を見つけられることがある。

これが命題論理の推論と統計的推論の本質的な違いである。

(参考)「人工知能の正しさ」＝「統計学の正しさ」

命題論理で作られた人工知能は、設定された論理環境内では間違えることはなかったが、

今日の人工知能の推論はデータから推論規則を構築する統計的推論であって、実世界で利用できる可能性があるが、その正しさの意味は基本的に統計学と同じである。

つまり、今日の人工知能の推論は正しいのか、という質問に答えるには、統計学の基盤を正しく理解する必要がある。

◎統計学と同様に、正しい人工知能というものは存在しない。しかし人工知能がどのくらい間違えるかを理論的に調べることはできる。

2. 統計的推論では何を知りたいのか

ここではベイズ推論について述べる。

予測損失と総和損失について主に考察する。

記号と定義

真の分布 $q(x)$ から独立なデータ $X^n=(X_1, X_2, \dots, X_n)$ が得られたとする。

統計モデル $p(x|w)$ と事前分布 $\varphi(w)$ が与えられたとき

事後分布 $p(w|X^n)$ および **予測分布** $p(x|X^n)$ を次のように定義する。

事後分布による平均を $E_w [\]$ と書く。 Z は**周辺尤度**。

事後分布	$p(w X^n) = (1/Z) \varphi(w) \prod_{i=1}^n p(X_i w)$
予測分布	$p(x X^n) = E_w [p(x w)]$
周辺尤度	$Z = p(X^n) = \int \varphi(w) \prod_{i=1}^n p(X_i w) dw$

予測損失とKL情報量

ベイズ推測では

「 x の分布 $q(x)$ はおおよそ予測分布 $p(x|X^n)$ だろう」

と推測する。この推測の誤差を次の予測損失で測ることにする。

$$\text{予測損失 } G_n = - \int q(x) \log p(x|X^n) dx$$

真の分布のエントロピーを $S = - \int q(x) \log q(x) dx$ とすると

$$G_n = S + \int q(x) \log (q(x) / p(x|X^n)) dx$$

- ◎ 予測損失が小さいほど真の分布と予測分布のKL情報量が小さい。
しかし予測損失を求めるには真の分布が必要。

総和損失とKL情報量

ベイズ推測では

「 $x^n=(x_1, x_2, \dots, x_n)$ の分布 $q(x^n)$ はおおよそ $p(x^n)$ だろう」

と推測する。この推測の誤差を総和損失で測ることにする。

$$\begin{aligned}\text{総和損失} &= - \int q(x^n) \log p(x^n) \, dx^n \\ &= nS + \int q(x^n) \log (q(x^n) / p(x^n)) \, dx^n\end{aligned}$$

◎ 総和損失が小さいほど $q(X^n)$ と $p(X^n)$ のKL情報量が小さい。
しかし総和損失を求めるには真の分布が必要。

予測損失と総和損失の比較

予測損失:

過去の n 個のデータから、未来に出るデータを予測したときの損失。

総和損失:

与えた統計モデルと事前分布で、過去に得られたデータ n 個を説明するときの損失。これは、過去に得られたデータ n 個をベイズ符号化するときの平均符号長に等しい。

この二つは強い関係を持っているが、統計モデルと事前分布の適切さを測る規準としては、(漸近的にも)異なる性質を持つ。

推測のための確率変数

予測損失と総和損失は、どちらも真の分布がわからなければ求めることはできないが、次の量を用いることで推測することができる。

交差検証損失 C_n と自由エネルギー F_n を次のように定義する。

$$C_n = (1/n) \sum_{i=1}^n \log E_w [1/p(X_i|w)]$$

$$F_n = -\log \int \varphi(w) \prod_{i=1}^n p(X_i|w) dw$$

(注) C_n も F_n も確率変数。どちらも真の分布が不明でも原理的には計算できる(発散や演算量の問題は残る)。

基本的な性質

任意の(真の分布・統計モデル・事前分布)について次の等式が成り立つ。

$$E[G_{n-1}] = E[C_n]$$

$$E[F_n] = \text{総和損失} = \sum_{i=1}^{n-1} E[G_i]$$

これらの性質は定義から自然に導出される(ベイズ推測の自然さ)。

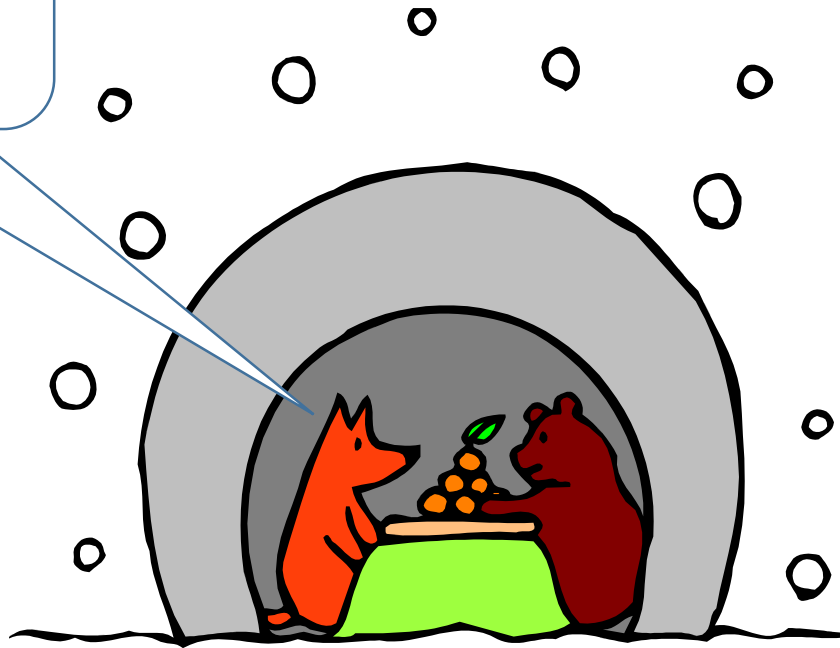
平均値が等しいだけでなく、さらに詳しい性質が解明されている。
具体的には G_n, C_n, F_n の確率変数としての漸近挙動がわかっている。
(数学的な準備と時間が必要なので本講演では述べない)。

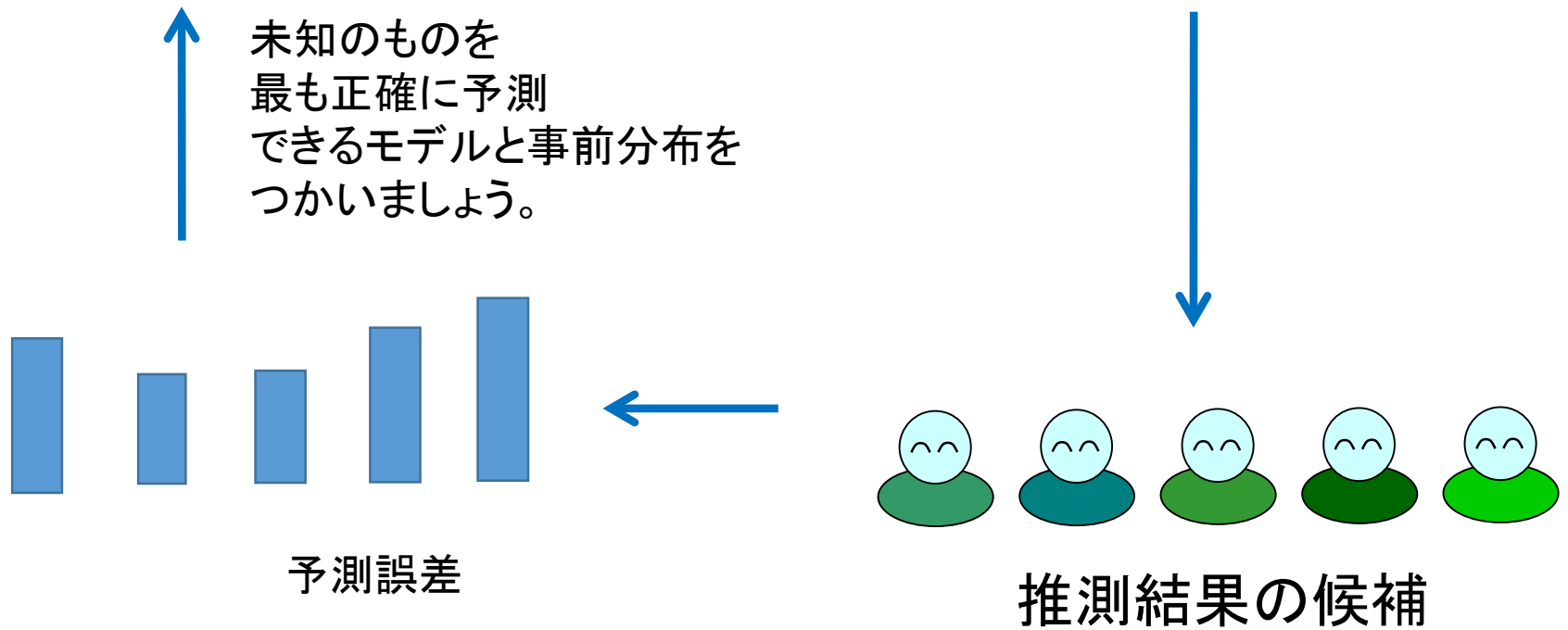
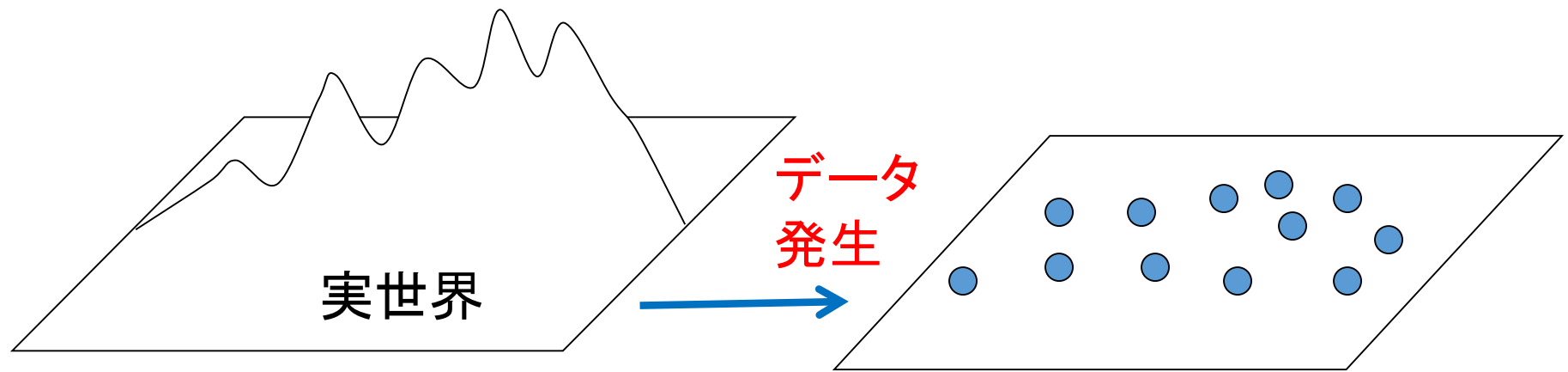
ベイズ推測の基礎:まとめ

- ◎ 予測損失と総和損失は、推測結果の誤差を表しているが真の分布がわからなければ計算できない。
- ◎ 交差検証損失と自由エネルギーは真の分布がわからなくても、データがあれば計算できる(演算量や発散の問題は残るが)。
- ◎ 予測損失と交差検証損失は、平均値は等しい。
- ◎ 総和損失と自由エネルギーは、平均値は等しい。
- ◎ 交差検証損失と自由エネルギーを計算することにより、ベイズ推測がどの程度の誤差を持つかを(平均値として)知ることができる。

3. 統計学で何がわかるか(予測編)

未来を正確に
予測できてこそ
統計学である



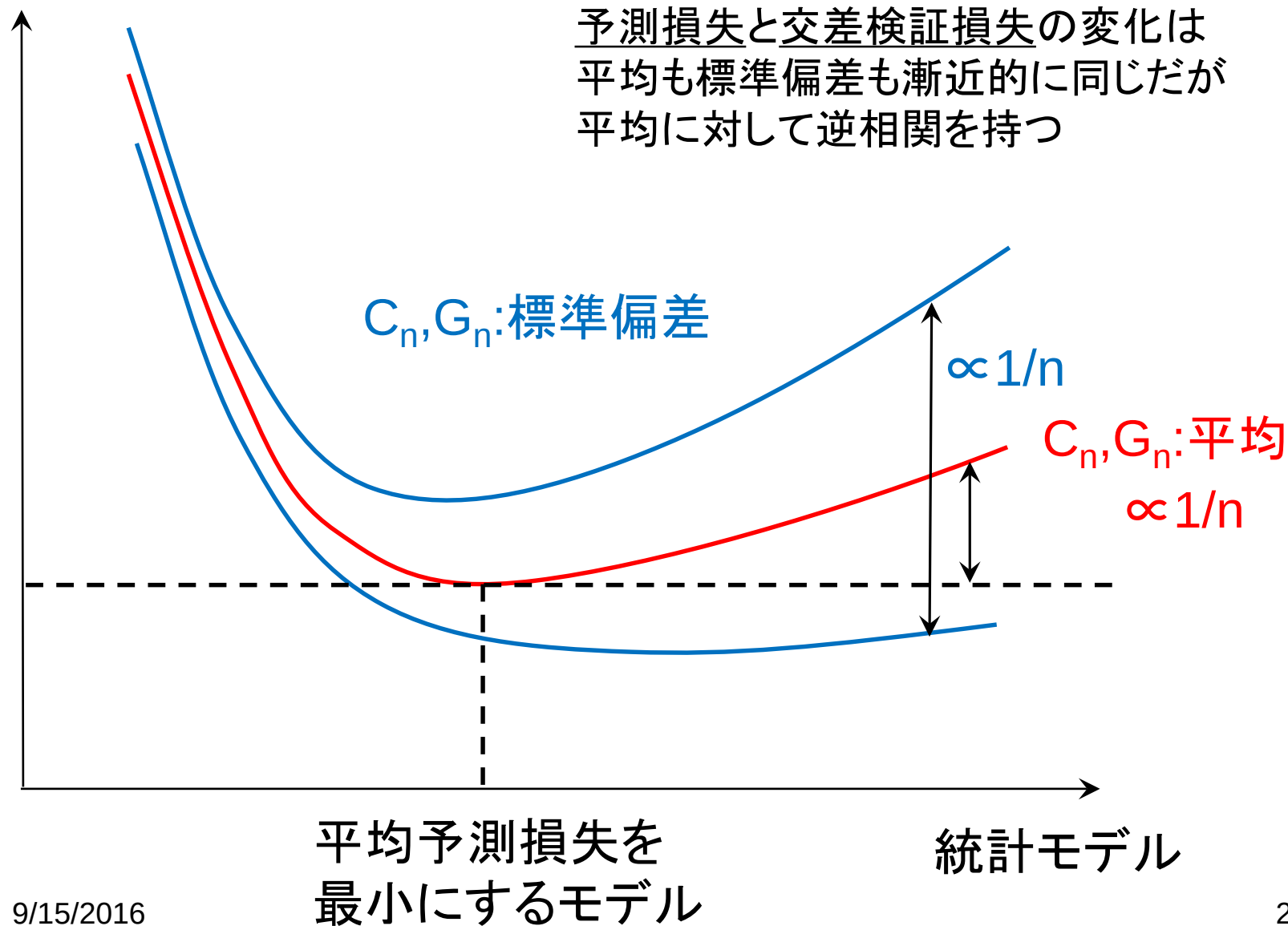


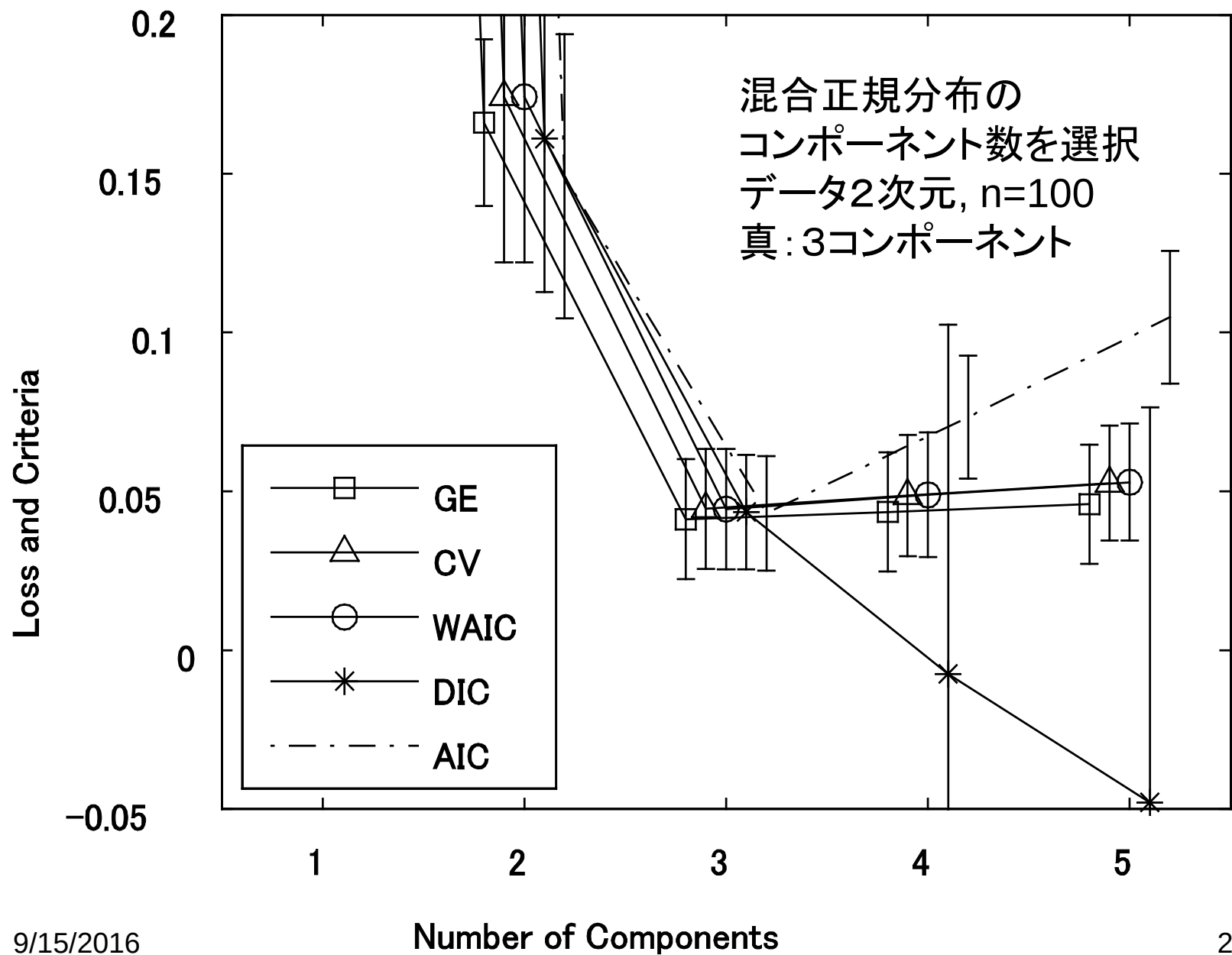
G_n : 予測誤差

C_n : 交差検証損失

(正則でも特異でも)

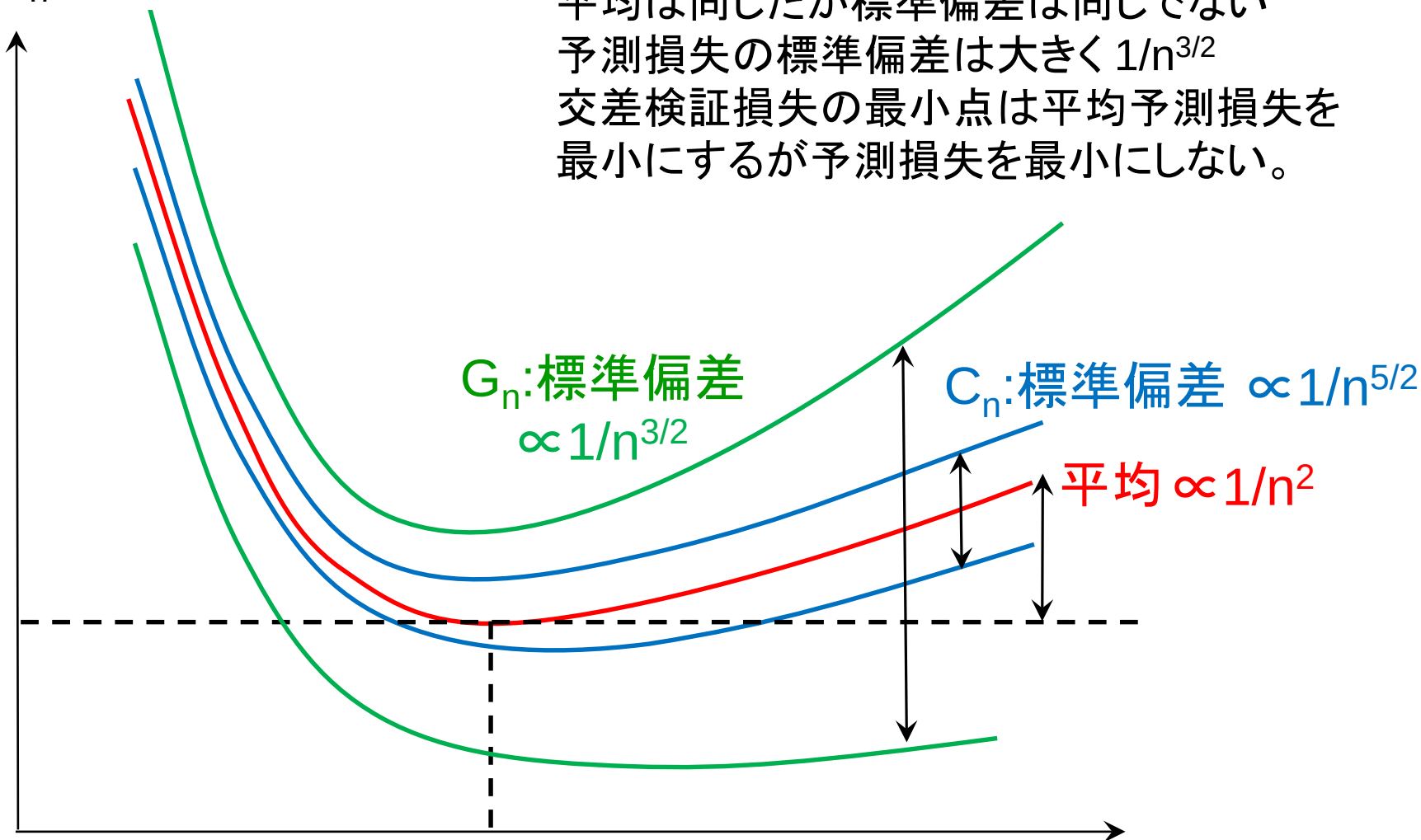
統計モデルを変化させたときの
予測損失と交差検証損失の変化は
平均も標準偏差も漸近的に同じだが
平均に対して逆相関を持つ





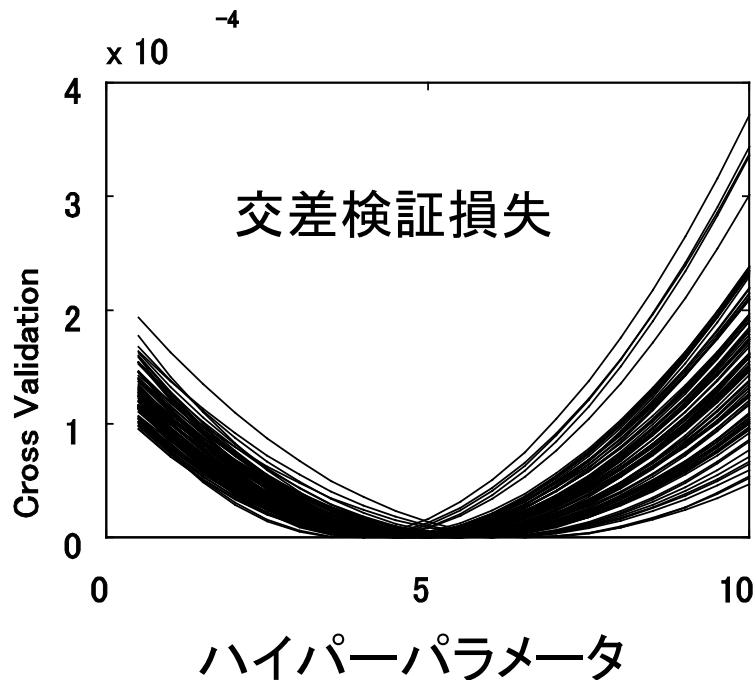
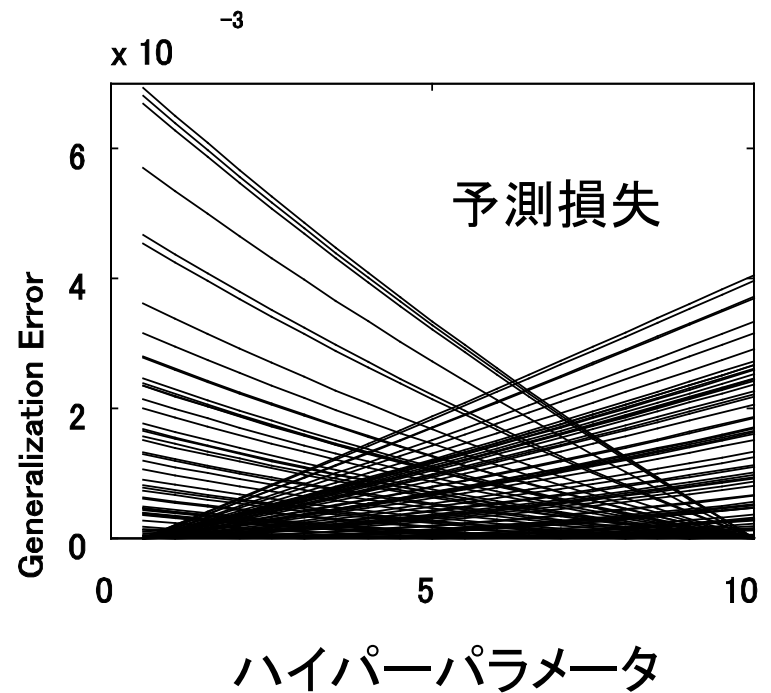
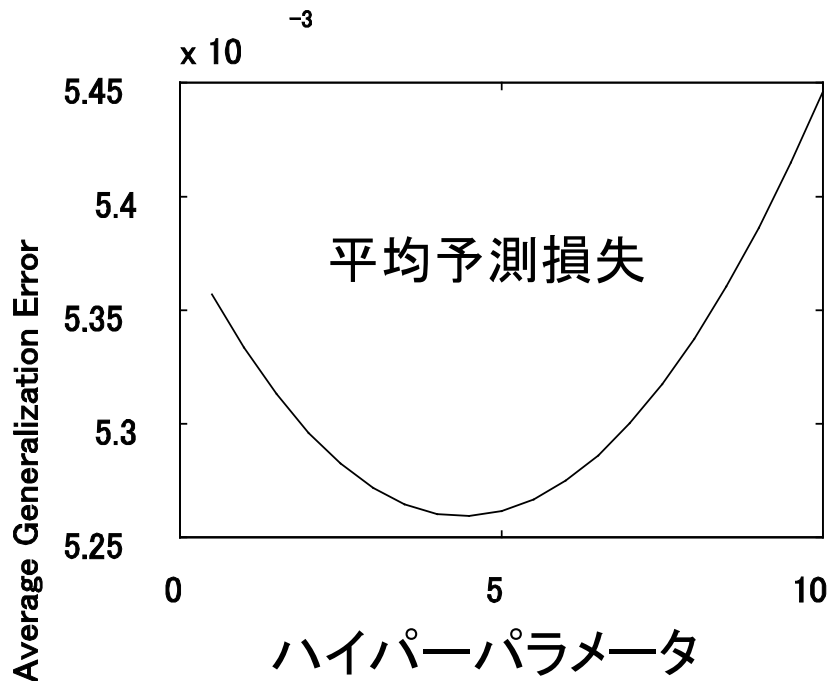
G_n : 予測誤差
 C_n : 交差検証損失

(正則な場合) 事前分布を変化させたときの
予測損失と交差検証損失の変化は
平均は同じだが標準偏差は同じでない
予測損失の標準偏差は大きく $1/n^{3/2}$
交差検証損失の最小点は平均予測損失を
最小にするが予測損失を最小にしない。



平均予測損失を最小にする事前分布
予測損失は最小にしない

事前分布



モデル

$$p(x|m,s) \propto s^{1/2} \exp(-(s/2)(x-m)^2)$$

事前分布

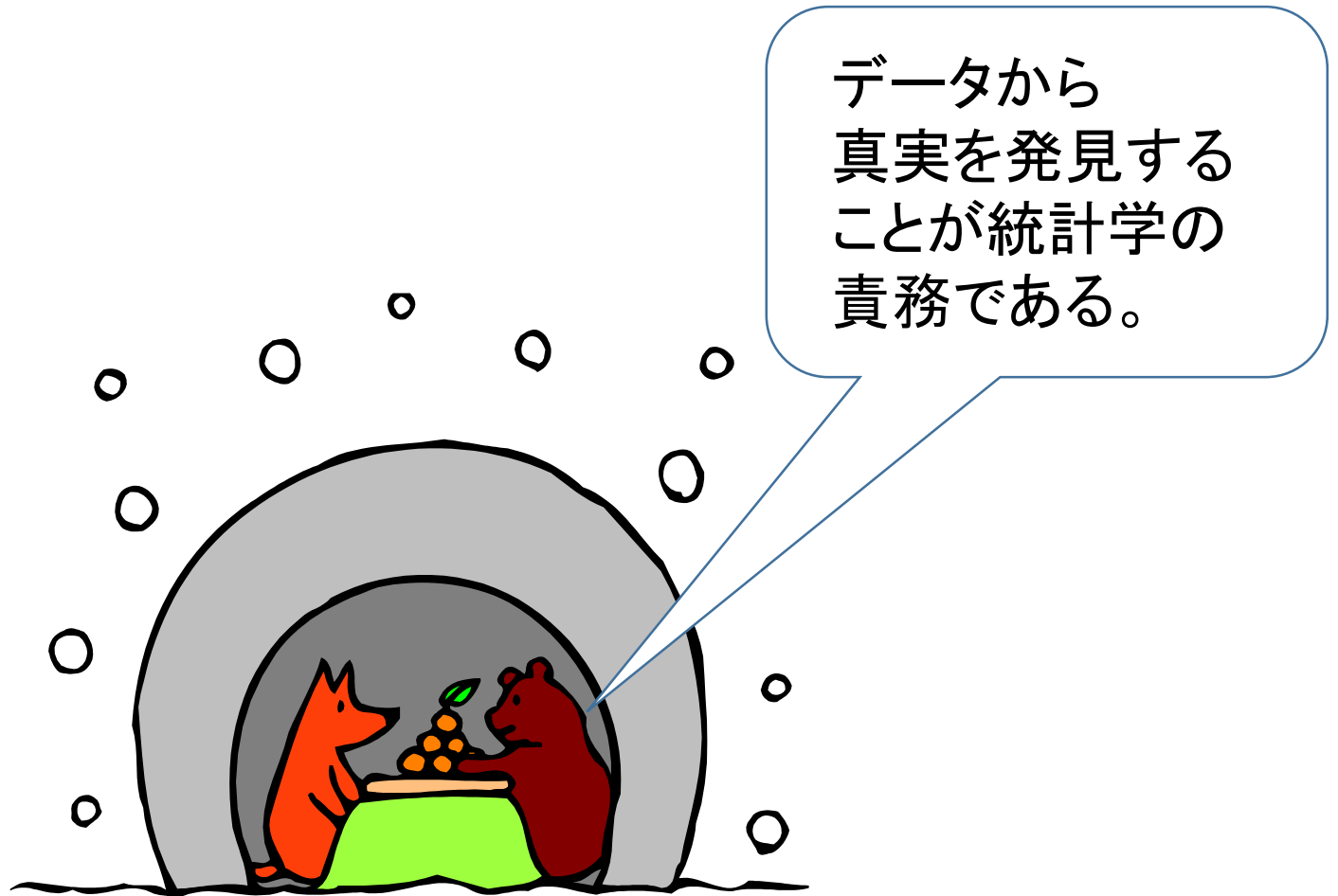
$$\varphi(m,s|h) \propto s^{1/2} \exp(-s(h+m^2)/2)$$

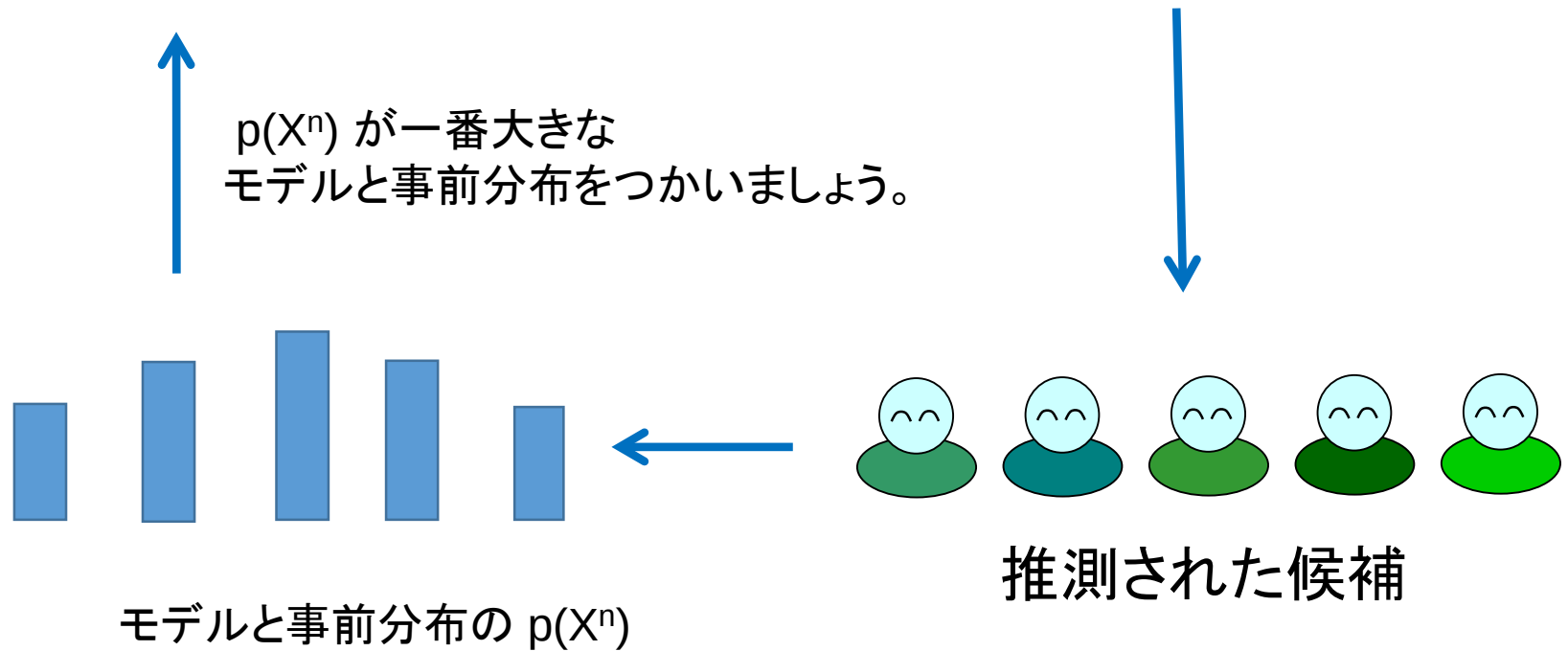
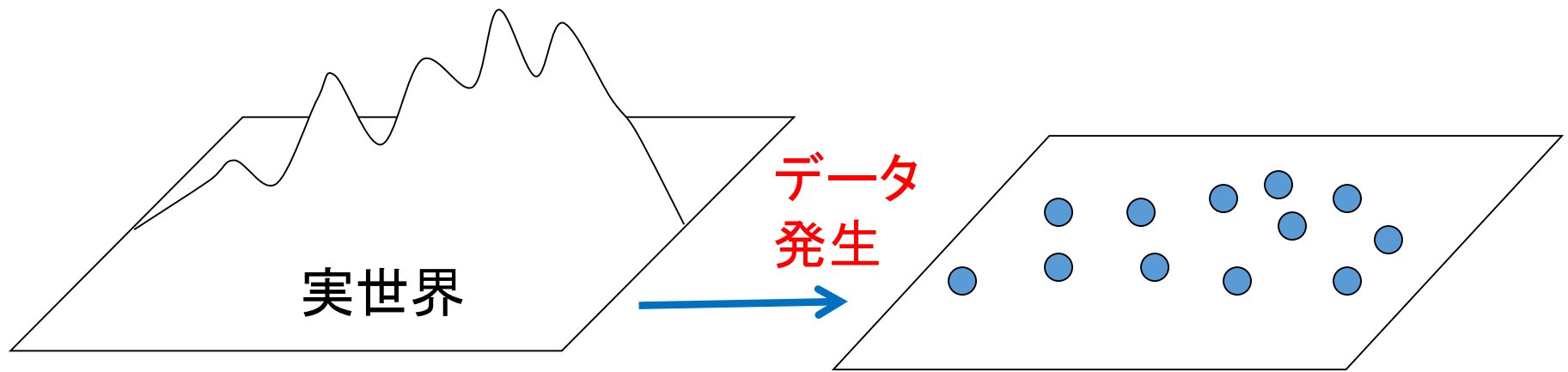
h : ハイパーパラメータ $[0,10]$

$n=200$

交差検証損失を最小にする
 ハイパーパラメータは平均予測
 損失をほぼ最小にするが、
 予測損失を最小にしない。

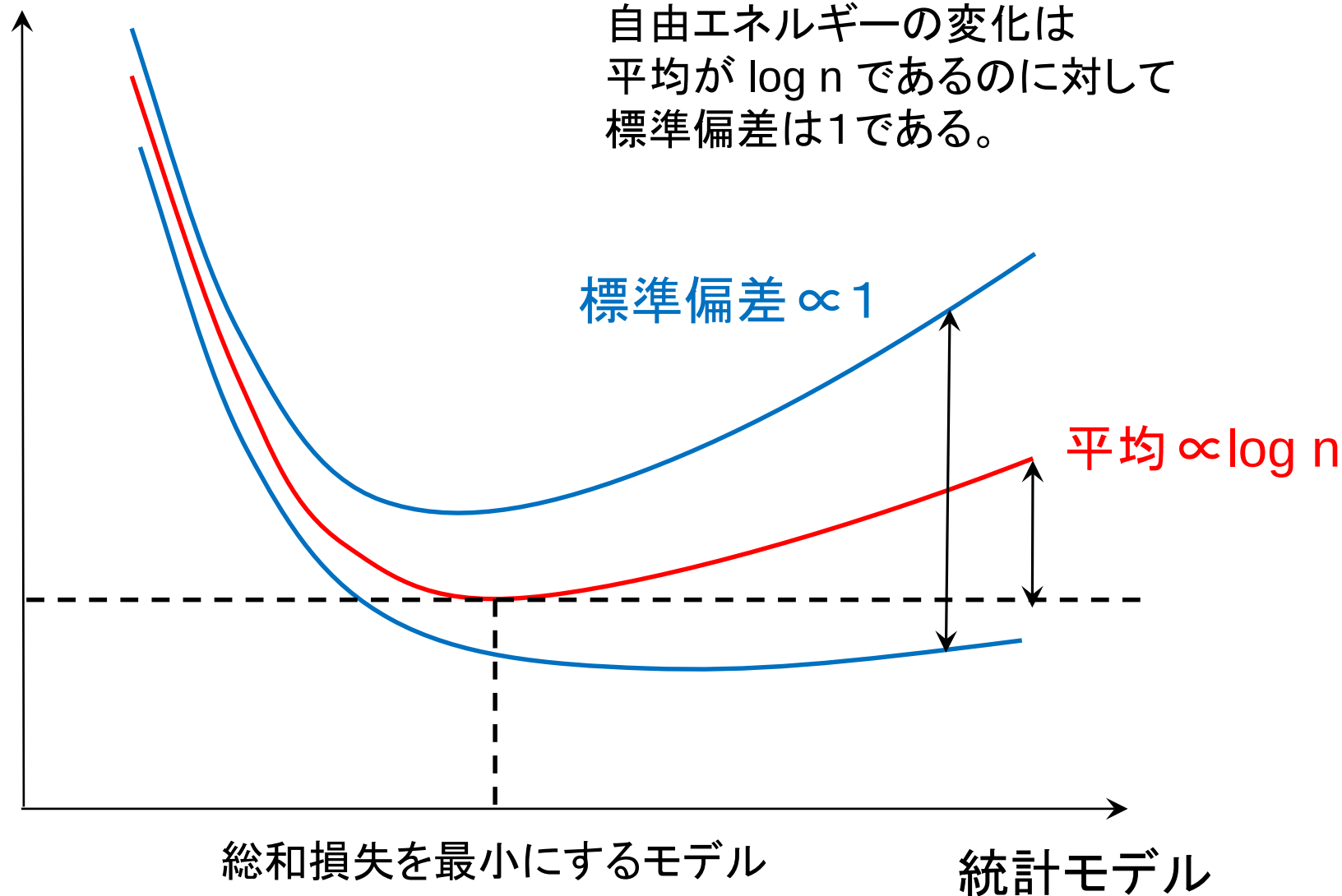
4. 統計学で何がわかるか(真実発見編)



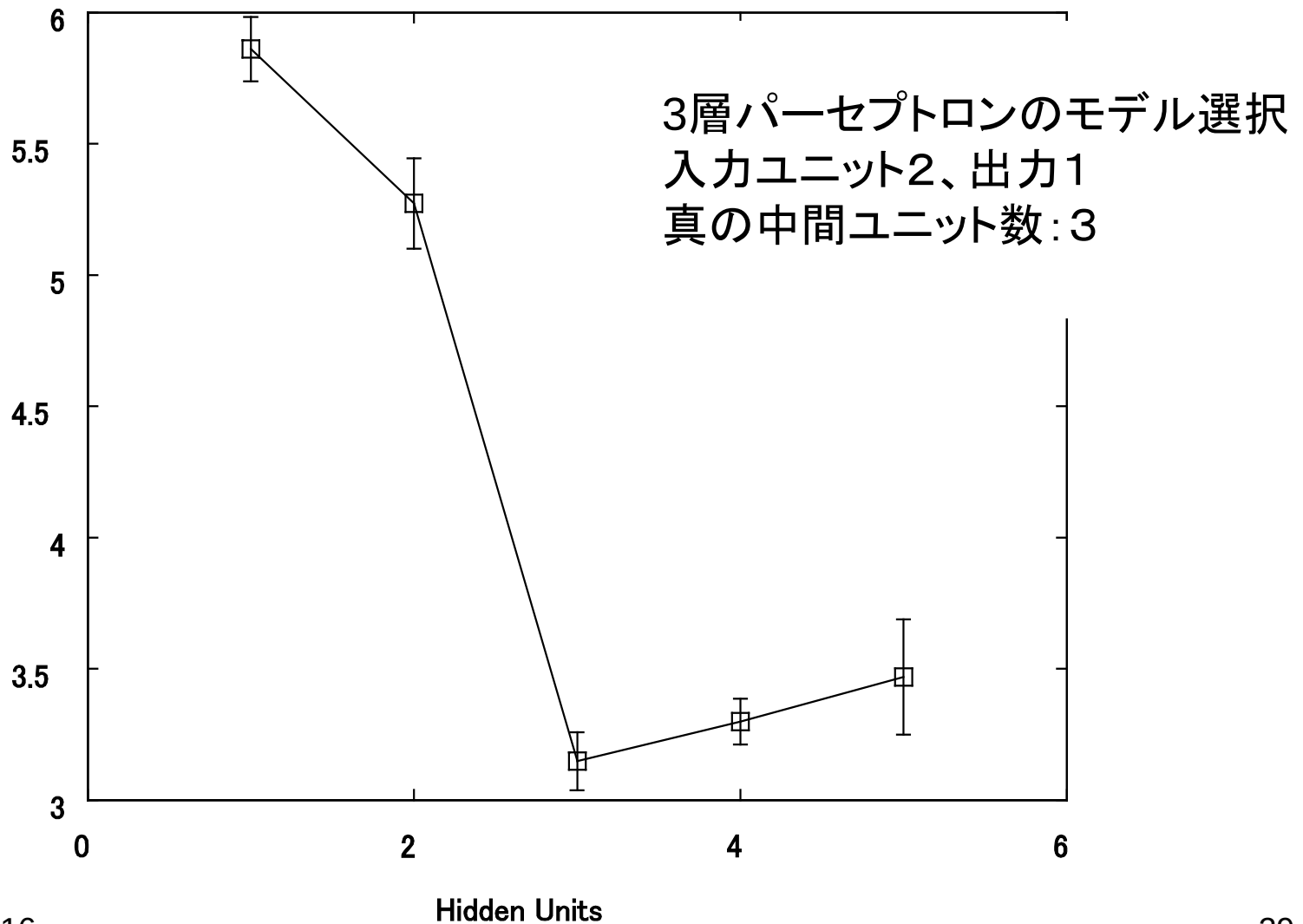


F_n : 自由エネルギー

(正則でも特異でも)
統計モデルを変化させたときの
自由エネルギーの変化は
平均が $\log n$ であるのに対して
標準偏差は1である。

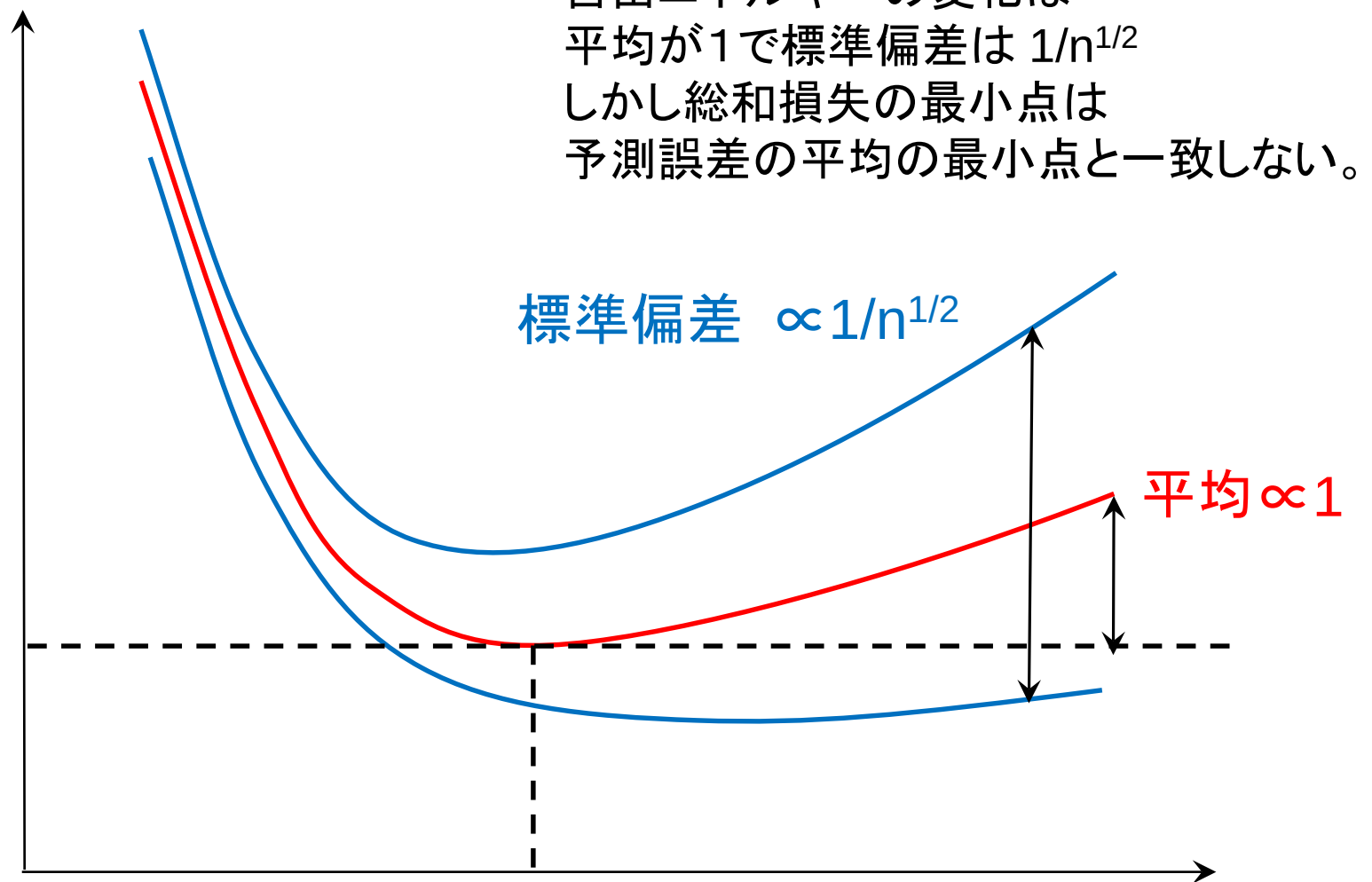


$\log(\text{自由エネルギーの推定値}) = (\log(\text{WBIC}))$



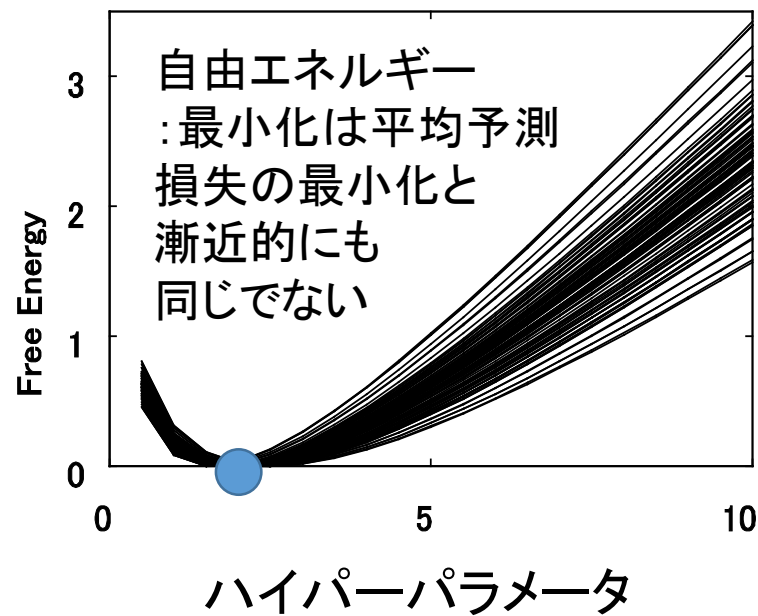
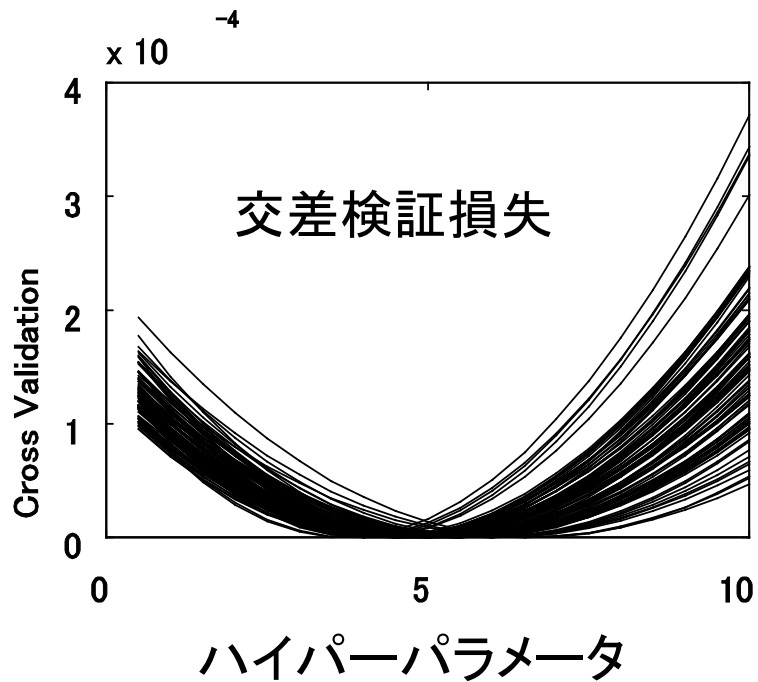
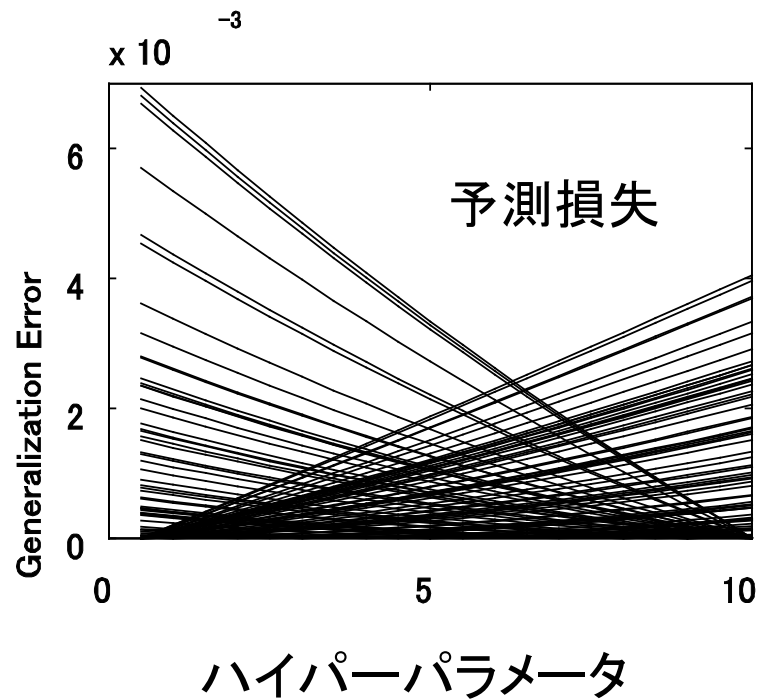
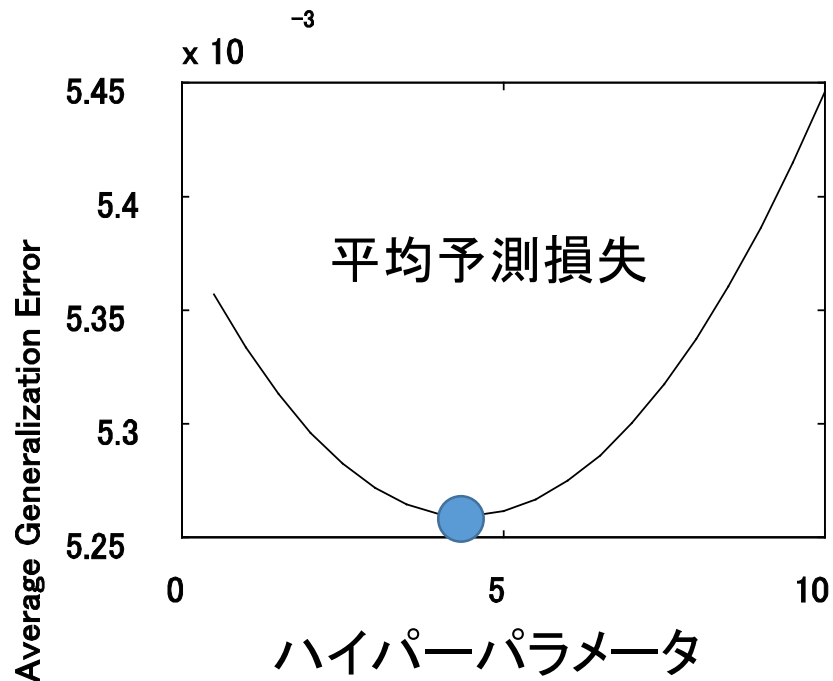
F_n : 自由エネルギー

(正則な場合) 事前分布を変化させたときの
自由エネルギーの変化は
平均が1で標準偏差は $1/n^{1/2}$
しかし総和損失の最小点は
予測誤差の平均の最小点と一致しない。



総和損失を最小にする事前分布:
予測損失の平均を最小にしない

事前分布



考察

1. 予測損失をできるだけ正確に知るにはどうしたらよいか。
2. 総和損失をできるだけ正確に知るにはどうしたらよいか。

結論

1. 統計学では、どんな確率空間でも成り立つ理論を作ること、推論の誤差を推定することができる。
2. ベイズ推論では、予測損失と総和損失が重要である。