「ベイズ統計の理論と方法」勉強会
「4. 一般理論」前半パート

Chihiro Mihara

テキスト

渡辺澄夫. ベイズ統計の理論と方法. コロナ社. 2012.

http://watanabe-www.math.dis.titech.ac.jp/users/swatanab/bayes-theory-method.html

※ 上のテキストの4章前半の内容の勉強会の資料ですが、勝手な説明を加えている箇所もあります。テキストの解釈の誤りや勝手な説明の変なところは私に帰属します。

テキスト4章までが目指すところ

「ベイズ推測する」って何をすること？

「真の分布 $q(x)$ はおおよそ $p^*(x) \equiv \int_W p(x|w)p(w|X^n)dw$ であろうと考える」こと。

→ といわれても、推測としてどういいのかよくわからない。

「ベイズ推測する」って結局どんな推測をすること？

真の分布 $q(x)$ と予測分布 $p^*(x)$ の誤差（汎化損失）

$$G_n \equiv -\int q(x) \log p^*(x)\, dx$$

（サンプルの選び方に依存する確率変数）

が［ ？ ］にしたがうような推測をすること。

→ ？ にあてはまる確率分布を特定するのがゴール！

テキスト3章までのあらすじ (1/2)

汎化損失 $G_n = -\mathbb{E}_X[\log \mathbb{E}_w[p(X|w)]]$ のしたがう分布を知りたい。

→ G_n は以下のようにキュムラント展開できる。

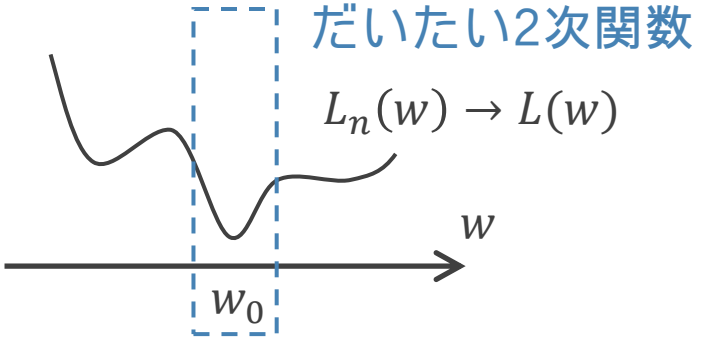
$$G_n = L(w_0) + \mathbb{E}_w[K(w)] - \frac{1}{2}\mathbb{E}_X[\mathbb{E}_w[f(X,w)^2] - \mathbb{E}_w[f(X,w)]^2] - \frac{1}{3!}\mathcal{G}^{(3)}(0) - \cdots$$

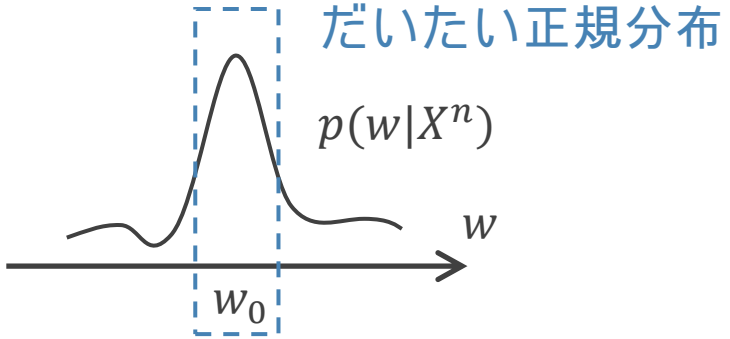
事後分布上の平均　事後分布上の分散

→ 事後分布 $p(w|X^n) \propto \exp\left(-n\beta L_n(w)\right)\varphi(w)$ の形が知りたい。

→ わからないので正規分布に近似できるように仮定をおきたい。

→ 平均対数損失 $L(w)$ は「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものと仮定してみる。





テキスト3章までのあらすじ (2/2)

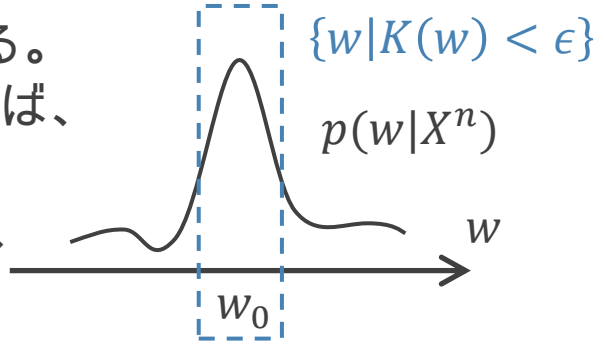
→ 平均対数損失 $L(w)$ は「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものと仮定してみると、G_n は以下のような確率変数であると示せる。ゴール達成！

$$G_n = L(w_0) + \frac{1}{n}\left(\frac{d}{2\beta} + \frac{1}{2}\|\xi_n\|^2 - \frac{1}{2\beta}\mathrm{tr}(IJ^{-1})\right) + o_p\left(\frac{1}{n}\right)$$

- J ： $L(w)$ の $w = w_0$ でのヘッセ行列。
- $I \equiv \left(\mathbb{E}_X\left[\nabla f(X,w)\left(\nabla f(X,w)\right)^T\right]\right)_{w=w_0}$
- $\xi_n \equiv \frac{J^{-\frac{1}{2}}}{\sqrt{n}}\left(\sum_{i=1}^n \nabla\left(K(w) - f(X_i,w)\right)_{w=w_0}\right)$

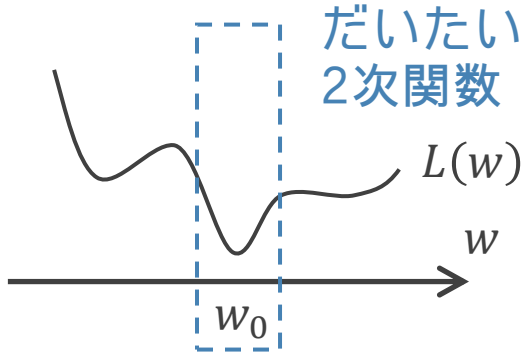
（サンプルの選び方に依存する確率変数）

上のことを示すには、w_0 の周り ϵ だけ切り取る。
ϵ を $n^{-1/2}$ よりゆっくり 0 に近づくようにとれば、
外側になる確率が n^{-1} より速く 0 に近づく。
平均値の定理を用いて内側を正規分布に近似し、
この正規分布上での平均や分散を求める。

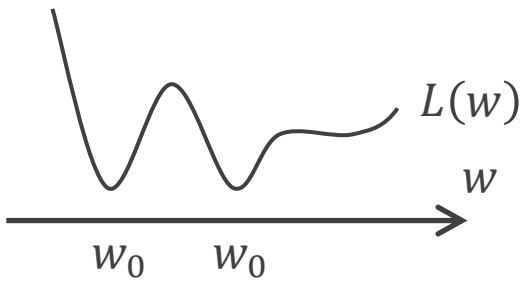


3章を終えて普通に気になること

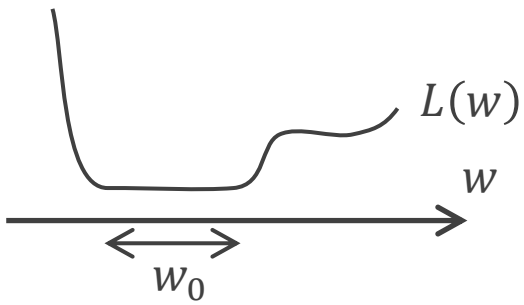
平均対数損失 $L(w)$ が「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものではない場合は、ベイズ推測は「推測としてこのようにいい」といえないの？？



こういうときはいえた。

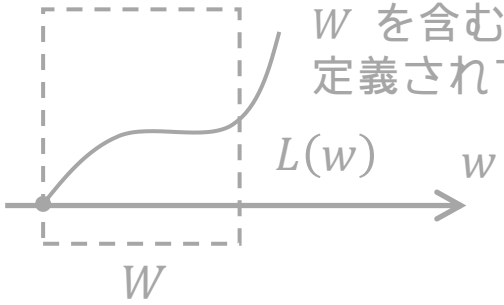


こういうときは？



こういうときは？

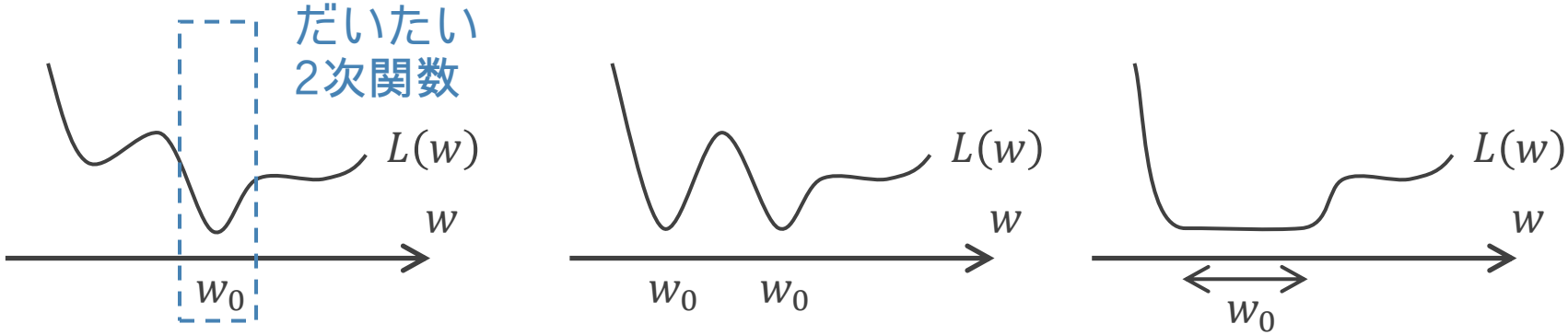
※ なお、以下のようなときについては考察しないことにする。



- W がコンパクトでない。
- 対数尤度比が相対的に有限な分散をもたない。

だから4章でやっていきたいこと

平均対数損失 $L(w)$ が「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものであるという仮定を外して、ベイズ推測は「推測としてこのようにいい」という！



こういうときはいえた。　こういうときも！　こういうときも！

ベイズ推測の汎化損失 G_n は［　？　］にしたがう！

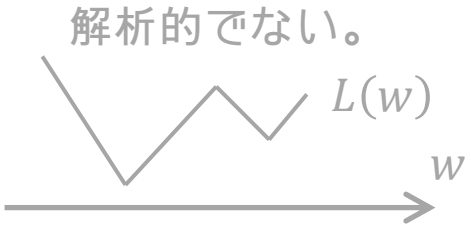
（確率分布）

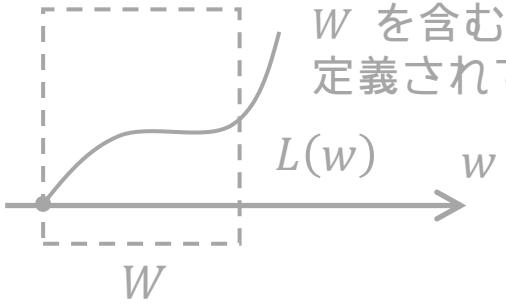
まず仮定

4章でこれは仮定すること

1. パラメータ集合 W はユークリッド空間のコンパクト部分集合であり、かつ、開集合 $W' \supset W$ が存在して、平均誤差 $K(w)$ が W' 上の解析関数である（95ページ 注意34）。
2. 対数尤度比 $f(x, w_0, w)$ が相対的に有限な分散をもつ（36ページ 注意12 (1)）。

※ なので、以下のようなときについては考察しない。





- W がコンパクトでない。
- 対数尤度比が相対的に有限な分散をもたない。

立ちはだかる壁

汎化損失 $G_n = -\mathbb{E}_X[\log \mathbb{E}_w[p(X|w)]]$ のしたがう分布を知りたい。

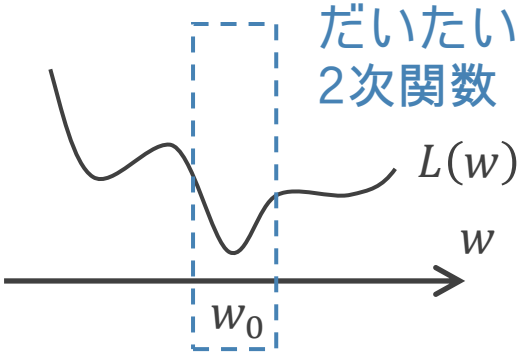
→ G_n は以下のようにキュムラント展開できる。

$$G_n = L(w_0) + \mathbb{E}_w[K(w)] - \frac{1}{2}\mathbb{E}_X[\mathbb{E}_w[f(X,w)^2] - \mathbb{E}_w[f(X,w)]^2]$$

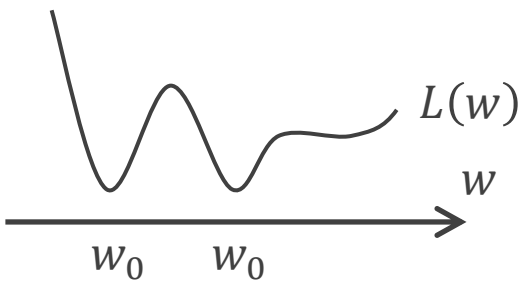
$$-\frac{1}{3!}\mathcal{G}^{(3)}(0) - \cdots$$

→ 事後分布 $p(w|X^n) \propto \exp\left(-n\beta L_n(w)\right)\varphi(w)$ の形が知りたい。

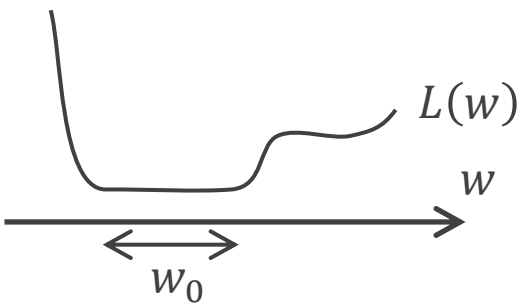
→ $L(w)$ に何も仮定をおかないとわからない！



こうであればいいが…。



こうかもしれない。



こうかもしれない。

こういう変換があればいいのに

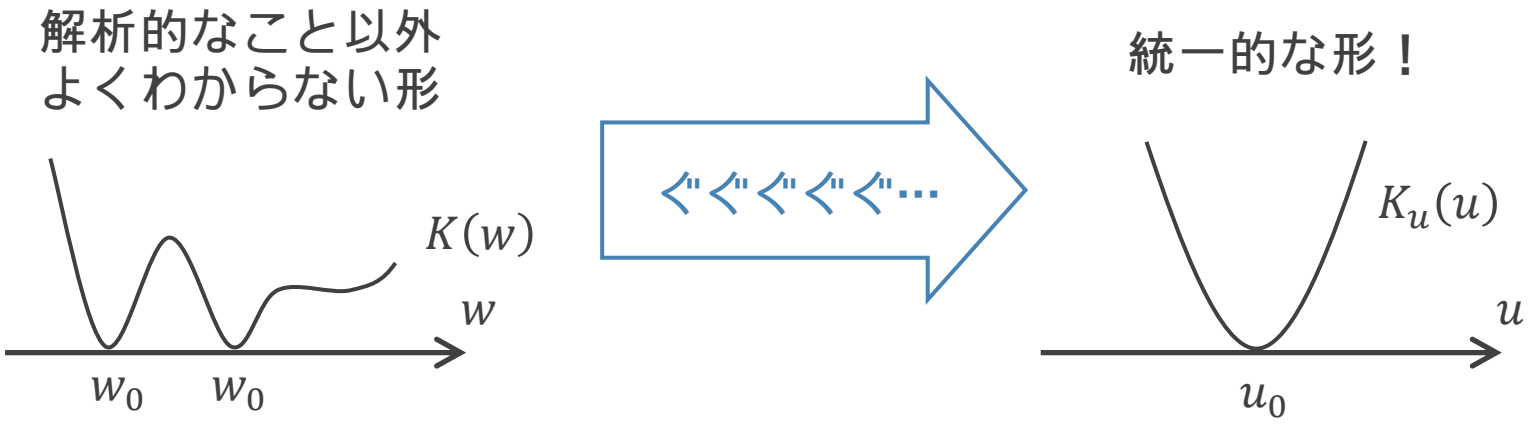
事後分布 $p(w|X^n) \propto \exp\left(-n\beta L_n(w)\right)\varphi(w)$ の形を知りたい。

→ $p(w|X^n) \propto \exp\left(-n\beta K_n(w)\right)\varphi(w)$ の形でもいいから知りたい。

※ $L_n(w)$ を最小値がゼロになるようオフセットしただけ。

→ $K(w)$ にしたところで何も仮定がないので全然わからない。

→ …パラメータ空間 W の方を何かぐにゃっと歪めて $K(w)$ を何か統一的な形にもっていくことはできないの？？



そういう変換があります

定理6（特異点解消定理のベイズ一般理論向け版）

$K(w) \geqq 0$ を開集合 $W \subset \mathbb{R}^d$ 上の非負解析関数とし、$K(w) = 0$ を満たす $w \in W$ が存在するとする。このとき、ある d 次元多様体 $\mathcal{M}$ と $\mathcal{M}$ 上の局所座標が取りうる値の集合 $\mathcal{U}$ からの解析写像 $g: \mathcal{U} \to W$ が存在して、$\mathcal{M}$ の局所座標ごとに、

$$K\big(g(u)\big) = u_1^{2k_1} u_2^{2k_2} \cdots u_d^{2k_d}$$

$$|g'(u)| = b(u) \left| u_1^{h_1} u_2^{h_2} \cdots u_d^{h_d} \right|$$

が成立するようにできる。ここで $|g'(u)|$ は $w = g(u)$ のヤコビアンであり、$b(u) > 0$ は 0 にならない解析関数であり、

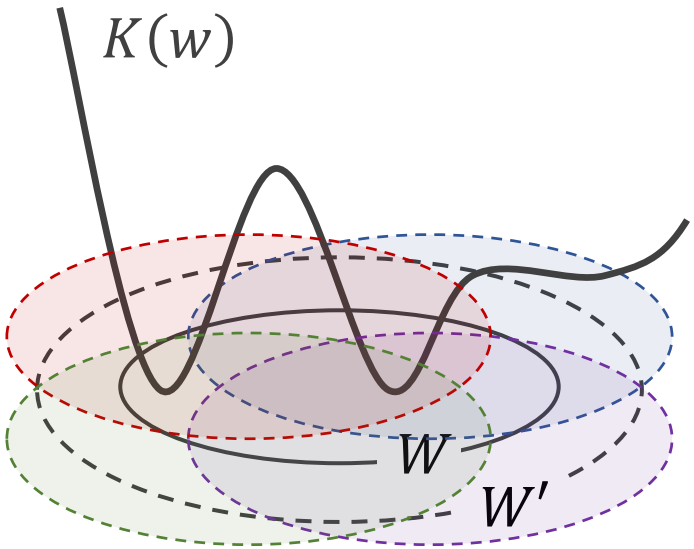
$$k = (k_1, k_2, \cdots, k_d), \quad h = (h_1, h_2, \cdots, h_d)$$

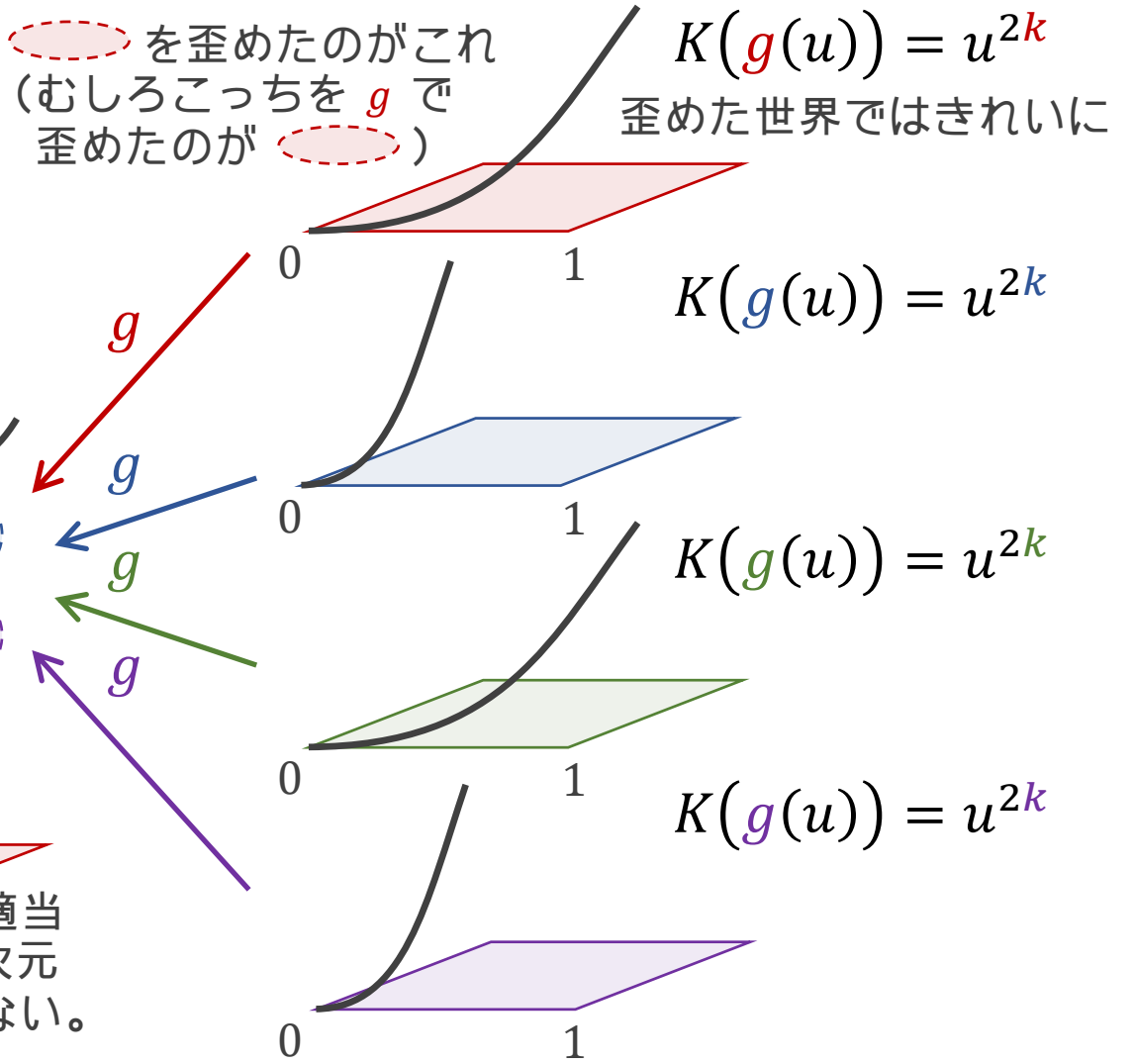
は非負の正数の集合である。但し $(k_1, k_2, \cdots, k_d)$ のうち少なくともどれか一つは 0 ではない。

つまり $u^{2k} \equiv u_1^{2k_1} u_2^{2k_2} \cdots u_d^{2k_d}$

有限個の統一的な形にできる
（ g, k はそれぞれ異なる）

よくわからない形を



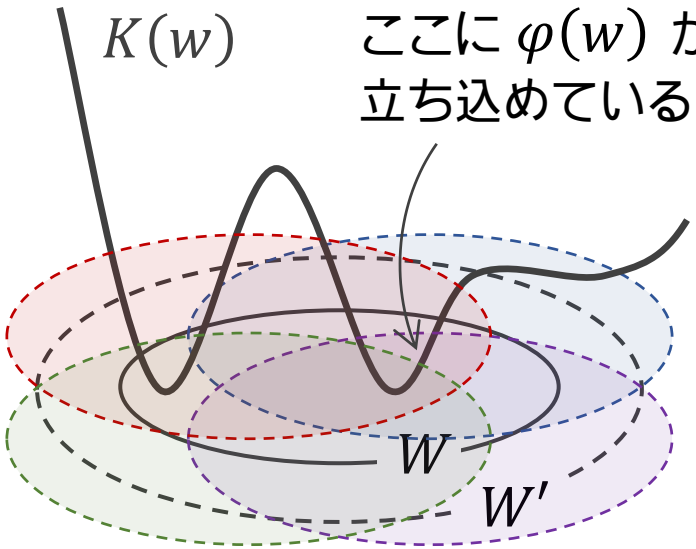


※ 新しいパラメータ空間  は、局所座標系を適当にとり、適当に切り分ければ $[0,1]^d$ の形（d次元超立方体）として一般性を失わない。

つまり

適当に分けてしまえばよい
（ b, h はそれぞれ異なる）

なので事前分布 $\varphi(w)$ も

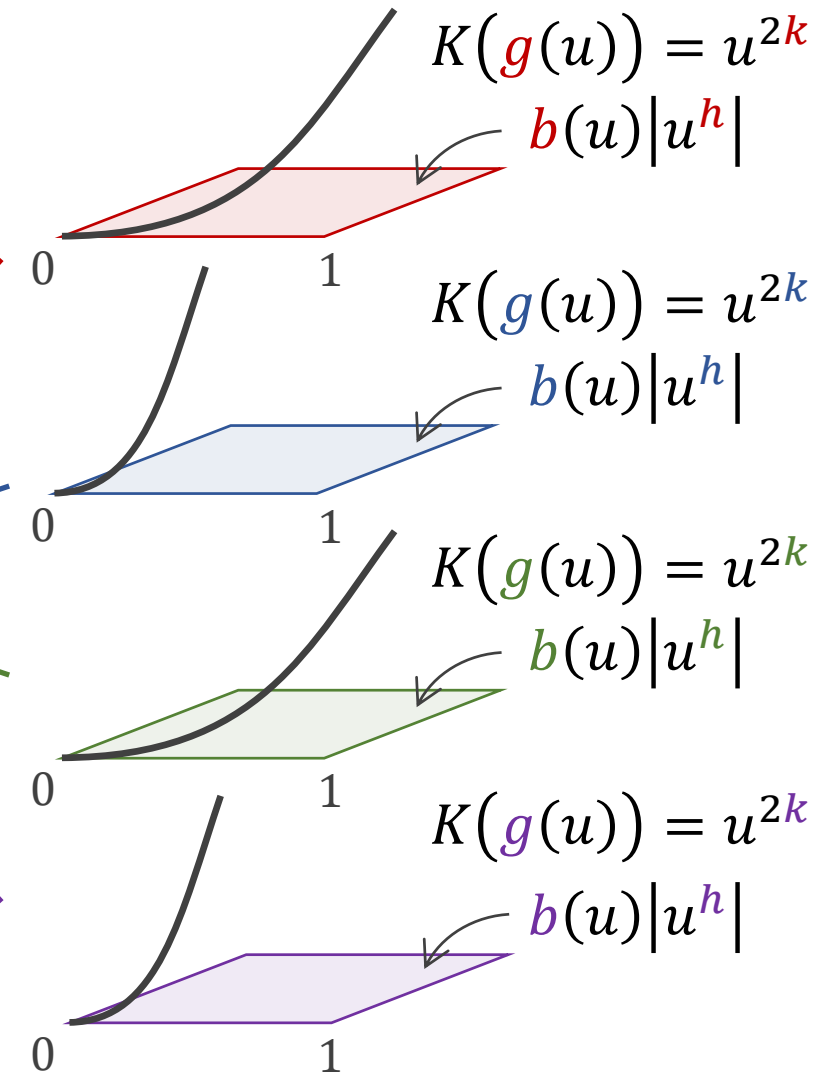


g

g

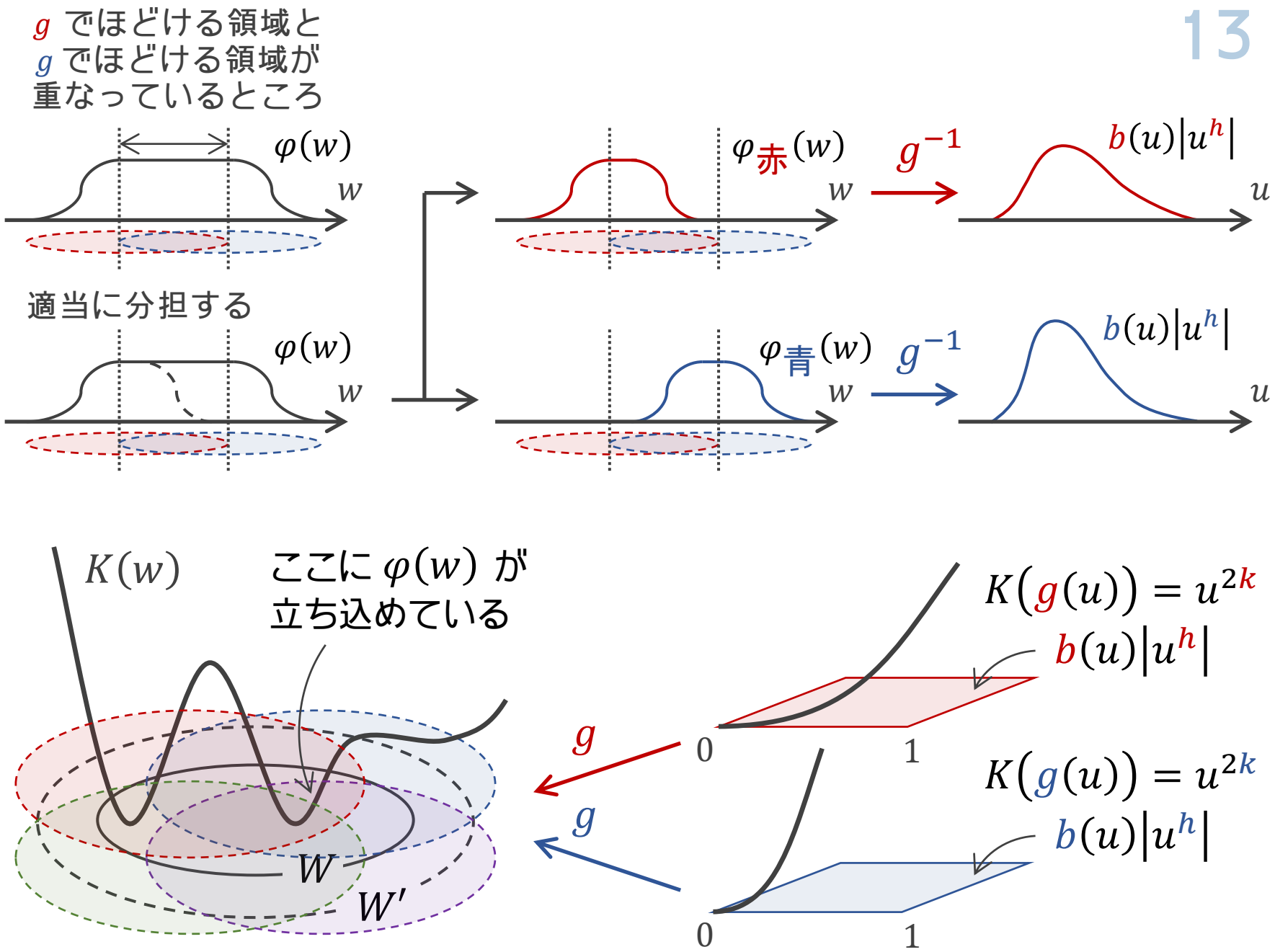
g

g



※ W がコンパクトなら適当に分ける
ことができる（参考. 1の分割）。

※ 新しい密度 $b(u)|u^h|$ は、事前分布を
分けた分 × 空間を g で歪めた分。


g でほどける領域と
g でほどける領域が
重なっているところ
$\varphi(w)$
w
適当に分担する
$\varphi(w)$
w
$\varphi_{赤}(w)$
w
g^{-1}
$b(u)|u^h|$
u
$\varphi_{青}(w)$
w
g^{-1}
$b(u)|u^h|$
u
$K(w)$
ここに $\varphi(w)$ が
立ち込めている
W
W'
g
g
$K(g(u)) = u^{2k}$
$b(u)|u^h|$
0
1
$K(g(u)) = u^{2k}$
$b(u)|u^h|$
0
1

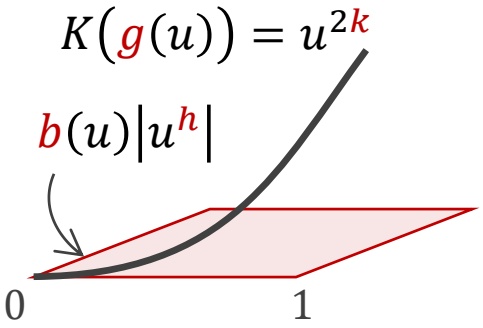
事後分布の標準形

事後分布 $p(w|X^n) \propto \exp\left(-n\beta K_n(w)\right)\varphi(w)$ の形を知りたい。

→ 定理6を適用し、w の分布から u の分布に。

$$p_{\text{赤}}(u|X^n) \propto \exp\left(-n\beta K_n\left(g(u)\right)\right) b(u)\left|u^h\right|$$

この形を考えればよい。



→ $K_n(g(u))$ は $K(g(u)) = u^{2k}$ とはずれるが、

$n \to \infty$ であるガウス過程に法則収束する $\xi_n(u)$ を用いて、

$$K_n(g(u)) = K(g(u)) - (\sqrt{n})^{-1} u^k \xi_n(u)$$

とかける。

→ つまり、事後分布の標準形はこうなる。

$$p_{\text{赤}}(u|X^n) \propto \exp\left(-n\beta u^{2k} + \sqrt{n}\beta u^k \xi_n(u)\right) b(u)\left|u^h\right|$$

確率的にゆらがない　　確率的にゆらぐ

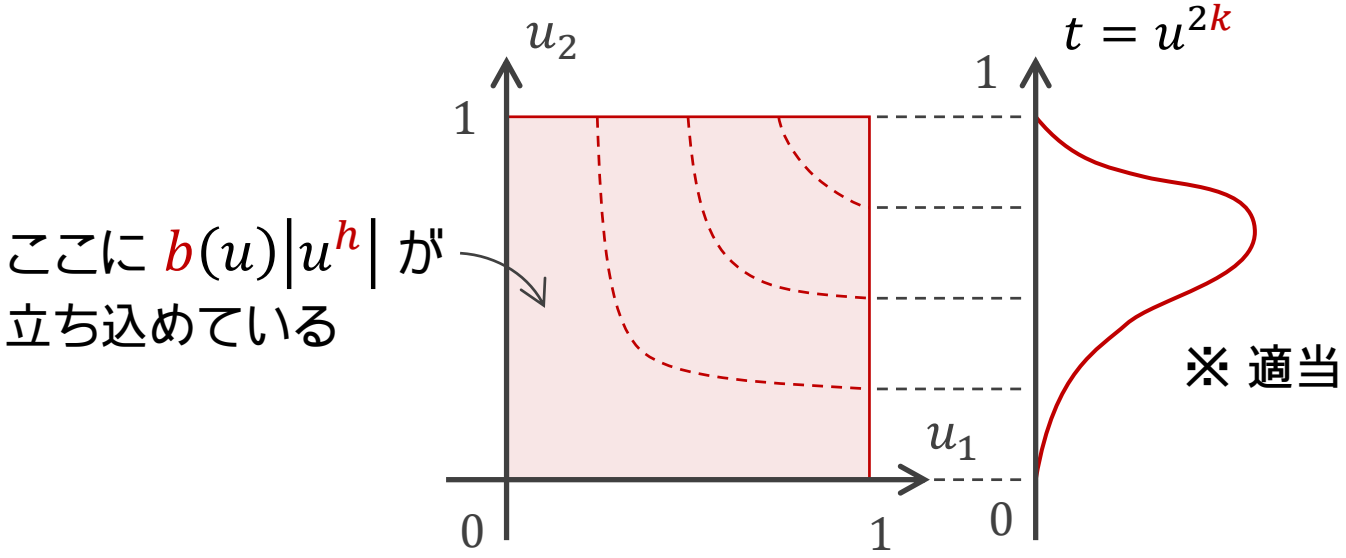
事後分布の標準形

じゃあ、$p_{赤}(u|X^n) \propto \exp\left(-n\beta u^{2k} + \sqrt{n}\beta u^k \xi_n(u)\right) b(u)\left|u^h\right|$ 上で $K\left(g(u)\right) = u^{2k}$ の平均や $f\left(x, g(u)\right) = u^k a(x,u)$ の分散がほしい。

→ このままだとやりづらい。

→ u^{2k} を主役にした方が捗る。

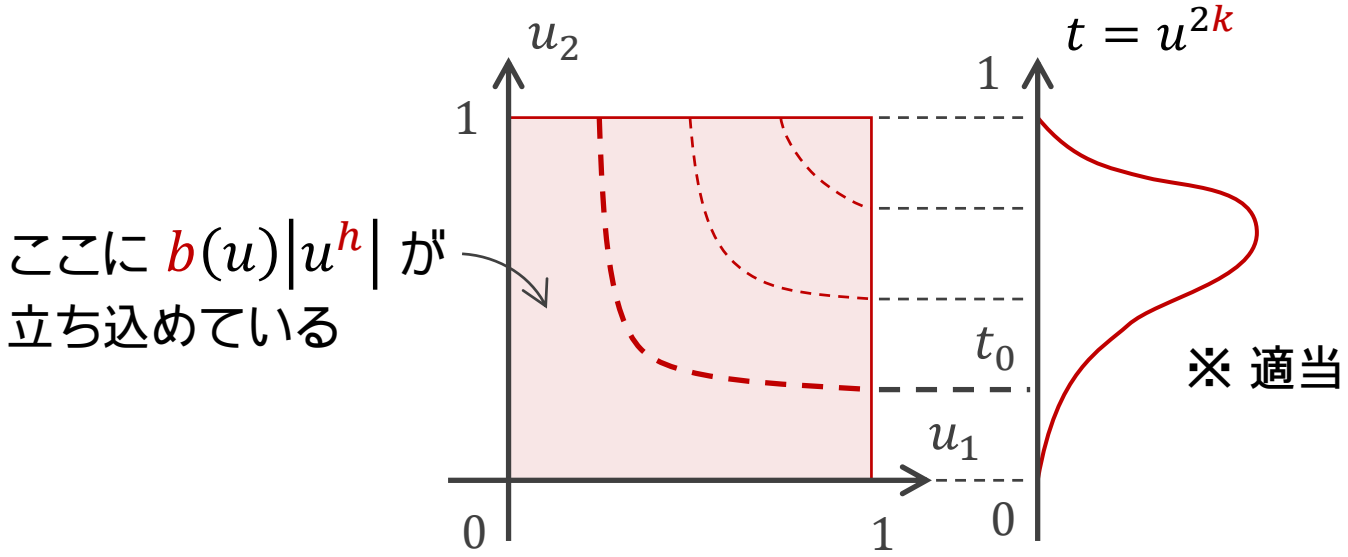
→ u の密度 $b(u)\left|u^h\right|$ 上での $t = u^{2k}$ の密度を考えたい（こっちの密度を状態密度という）。



状態密度

u の密度 $b(u)\left|u^h\right|$ 上での $t = u^{2k}$ の密度を知りたい。

→ $t_0 = u^{2k}$ を満たす u はたくさんある。



→ $t = f(u)$ を満たす u の密度を集めてくる記号がある。

$$p_t(t) = \int \delta(t - f(u))p_u(u)du$$

これ

→ $\delta(t - u^{2k})b(u)\left|u^h\right|$ をかけて積分するとどうなるか知りたい。

状態密度

メリン変換 ― t^z をかけて $t=[0,+\infty)$ で積分（$z\in\mathbb{C}$）。
複素関数が正則 ― どこから近づいても値が定まっている。
複素関数の極 ― $(z-\alpha)^m$ をかけて正則になるなら $z=\alpha$ が m 位極。

$\delta(t-u^{2k})b(u)|u^h|$ をかけて積分するとどうなるか知りたい。

→ $t\to 0$ で支配的な成分が大事（事後分布は $n\to\infty$ では $K(g(u))=0$ で一番濃くなるから）。

→ 実はメリン変換という変換で複素数の関数にすると、極の位置から、 $t\to 0$ で支配的な成分が取り出せる。

補題22

$\lambda>0$ を実数、 $m>0$ を自然数とする。

$$f_m(t)=\begin{cases} t^{\lambda-1}(-\log t)^{m-1} \\ 0 \end{cases}$$

$t\to 0$ で関数が 0 に近づく速さ

↕ 対応

極の位置と位数

のメリン変換は以下である。

$$(Mf_m)(z)=\frac{(m-1)!}{(z+\lambda)^m}$$

$z=-\lambda$ で値が定まらないが $(z+\lambda)^m$ をかければ定まる。

状態密度

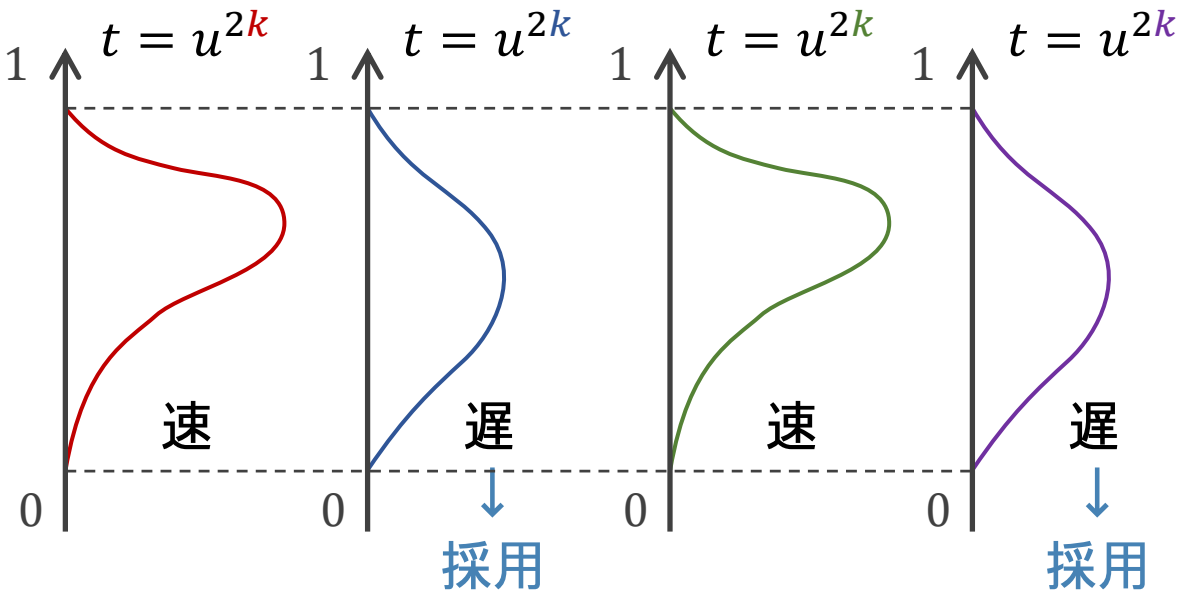
定理8

ある微小積分 du^* が存在して $t \to 0$ で以下が成り立つ。

$$\delta\left(t - u^{2k}\right) b(u) \left|u^h\right| du = t^{\lambda-1}(-\log t)^{m-1} du^* + o(\text{より速く}0\text{へ})$$

メリン変換して、実部が最大の極（ $z = -\lambda$ ）を取り出せばよい。

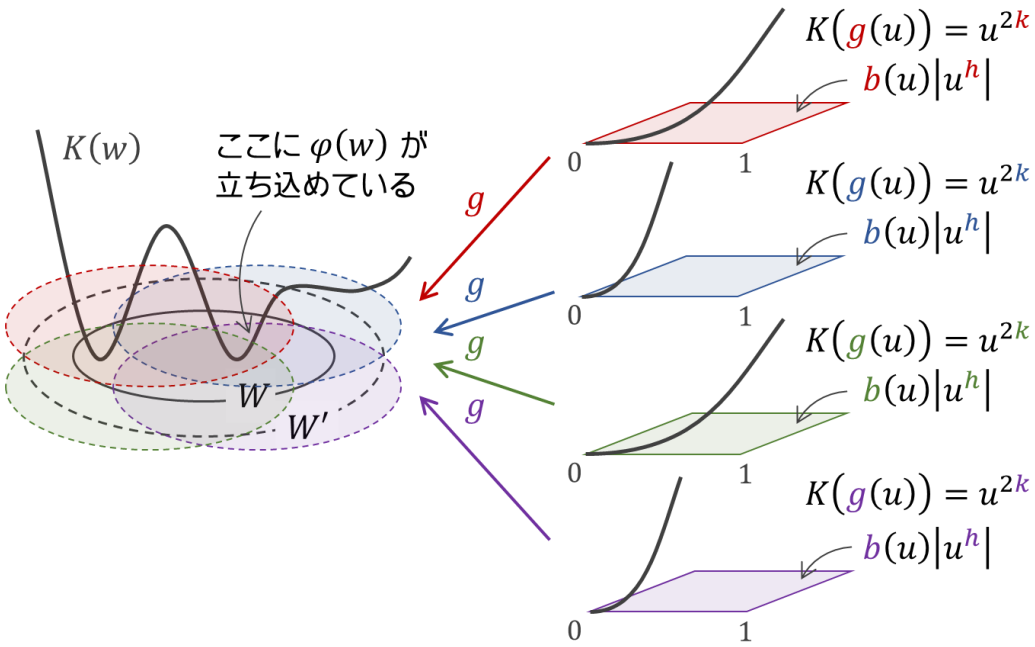
λ, m は局所座標ごとに異なるが、0 に近づくのが一番遅いものの値にすればよい。

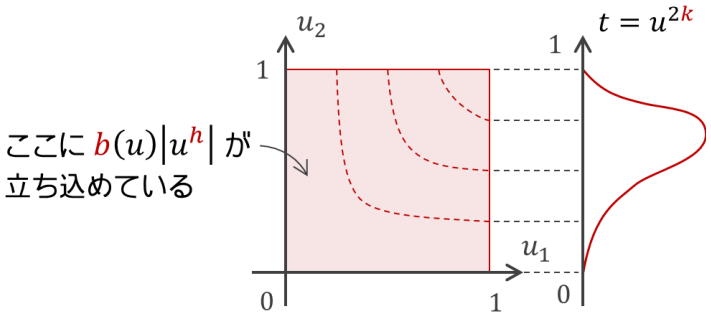


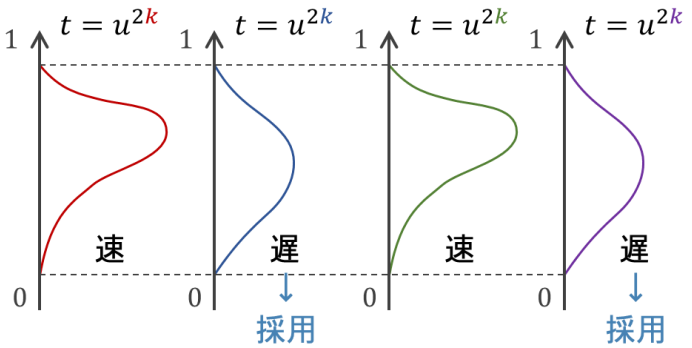
4章前半のまとめ

平均誤差 $K(w)$ が一般的な場合を取り扱うために、定理6の変数変換によって $K\big(g(u)\big) = u^{2k}$ の形に標準化し、それを用いて事後分布も標準化した。

事後分布上で期待値や分散をとる準備として、事後分布を u の密度から $t = u^{2k}$ の密度（$t \to 0$ での）に直した。







多様体

※ 特異点解消につかわれるのは代数多様体という多様体です。

※ テキスト89ページで説明しているのと、これから説明するのは代数多様体ではなく微分多様体です。ただ、代数多様体は微分多様体の構造をもっています。

※ 代数多様体を導入するのはこの勉強会の範囲を超えるのと、4章で何をしているかの理解は微分多様体でかなうと思われるため、微分多様体を説明しています。

多様体

パラメータ空間 W の方を歪めることで平均誤差 $K(w)$ を標準形にしたい。まず，パラメータを移す先 $\mathcal{M}$ を用意したい。$\mathcal{M}$ は集合だが，各点の周りに座標（つまり，変換したパラメータ）はほしい。ただの集合は「この点の周り」という概念がないので，まず位相というものを入れる。

定義（開集合系，位相空間）

集合 $\mathcal{M}$ の部分集合族 $\mathcal{O}$ が以下の3つの条件を満たすとき， $\mathcal{O}$ は $\mathcal{M}$ の**開集合系**であるという。

1. $\emptyset, \mathcal{M} \in \mathcal{O}$
2. $O_1, O_2 \in \mathcal{O} \Rightarrow O_1 \cap O_2 \in \mathcal{O}$
3. 任意の集合 Λ に対し，各元 $\lambda \in \Lambda$ から $\mathcal{O}$ の元 O_λ への対応を与えたとき，$\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathcal{O}$

集合 $\mathcal{M}$ にある開集合系 $\mathcal{O}$ が与えられているとき， 集合 $\mathcal{M}$ を $\mathcal{O}$ を開集合系とする**位相空間**といい，$\mathcal{O}$ を $\mathcal{M}$ の**位相**という。また，$\mathcal{O}$ の元を $\mathcal{M}$ の**開集合**という。

多様体

次に位相空間 $\mathcal{M}$ の各点の周り（各点を含む開集合）に d 次元の座標を割り当てるが，位相が保たれるようにする。つまり，$\mathcal{M}$ の開集合の像が $\mathbb{R}^d$ の開集合になり，$\mathbb{R}^d$ の開集合の逆像も $\mathcal{M}$ の開集合になるようにする。

定義（連続写像）

X, Y を位相空間とする。写像 $f : X \to Y$ に対して，Y の任意の開集合 V の逆像 $f^{-1}(V)$ が X の開集合であるとき，f は**連続写像**であるという。

定義（同相写像）

X, Y を位相空間とする。写像 $f : X \to Y$ に対して，f が全単射で f も f^{-1} も連続写像であるとき，f は**同相写像**であるという。

参考.

- ドーナツ $\in \mathbb{R}^3$ を取っ手付きのコップ $\in \mathbb{R}^3$ に変形する同相写像はある。
- ドーナツ $\in \mathbb{R}^3$ をビー玉 $\in \mathbb{R}^3$ に変形する同相写像はない。

多様体

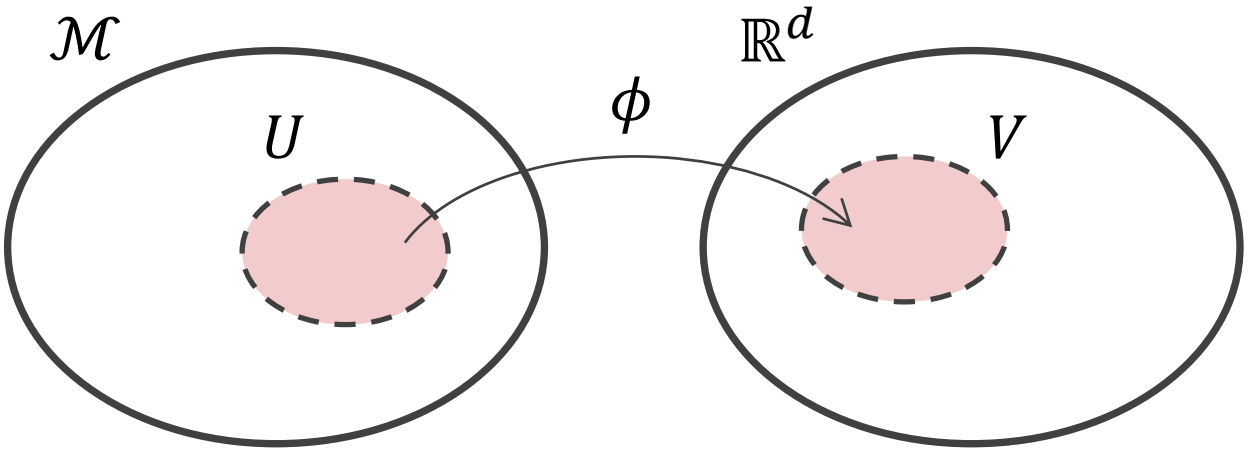
定義（座標近傍系）

位相空間 $\mathcal{M}$ の開集合 U から $\mathbb{R}^d$ の開集合 V への写像 $\phi : U \to V$ が同相写像であるとき，(U, ϕ) の対を $\mathcal{M}$ の d **次元座標近傍（チャート；地図）**という（※）。また，d 次元座標近傍の族 $S = \{(U_\lambda, \phi_\lambda)\}_{\lambda \in \Lambda}$ が $\bigcup_{\lambda \in \Lambda} U_\lambda = \mathcal{M}$ を満たすとき，S を $\mathcal{M}$ の d **次元座標近傍系（アトラス；地図帳）**という。

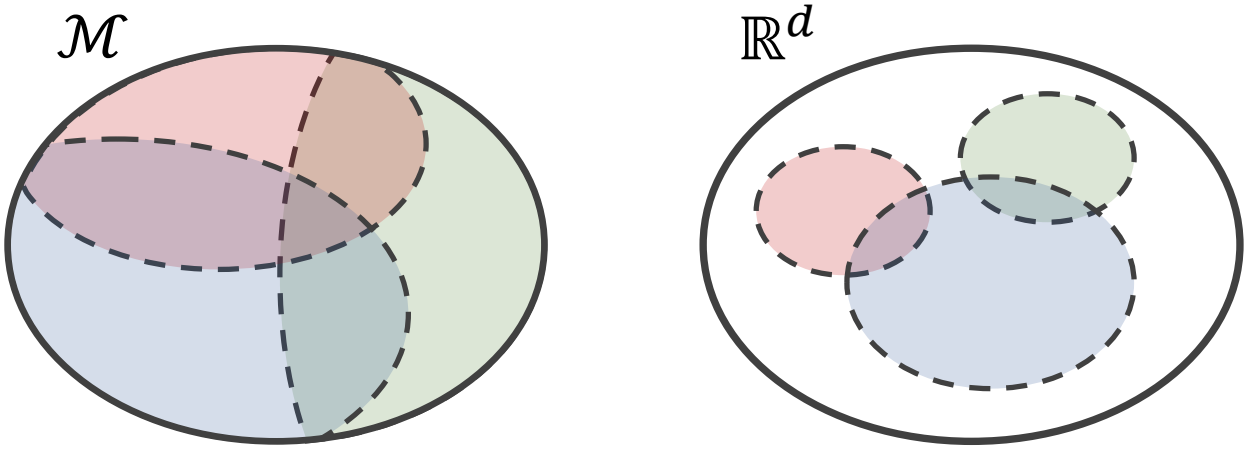
※ このとき ϕ を $U \subset \mathcal{M}$ の**局所座標系**といい，$x \in U$ に対して $\phi(x)$ を**局所座標**という。

多様体

座標近傍
（チャート）



座標近傍系
（アトラス）



多様体

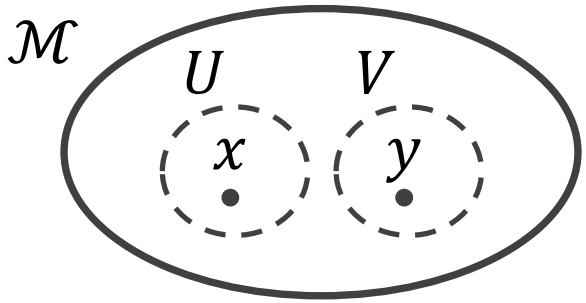
位相空間 $\mathcal{M}$ に座標近傍系さえ定義されていれば、パラメータの変換先にはなる。

ただし、多様体には通常、さらに以下のハウスドルフ性という性質も課される。

- これがないと1の分割ができない。

定義（ハウスドルフ空間）

位相空間 $\mathcal{M}$ の任意の2点 $x, y\ (x \neq y)$ に対してある開集合 U, V が存在して $x \in U,\ y \in V,\ x \cap y \neq \emptyset$ とできるとき、$\mathcal{M}$ をハウスドルフ空間という。



多様体

ハウスドルフ空間であって座標近傍系が定義できるものを多様体（位相多様体）という。

特に、以下を満たす多様体を微分多様体という。

- だから何？というか、特異点解消につかうのはこのような多様体。

定義（微分可能多様体）

ハウスドルフ空間 $\mathcal{M}$ に d 次元座標近傍系 $\{(U_\alpha, \phi_\alpha)\}$ が定義されていて、以下が満たされているとき、 $\mathcal{M}$ を d 次元 C^r 級微分可能多様体という。

- $U_\alpha \cap U_\beta \neq \emptyset$ であるような任意の α, β に対して、$\phi_\alpha \circ \phi_\beta^{-1}$ が C^r 級写像である。

標準形

定理6（特異点解消定理のベイズ一般理論向け版）

$K(w) \geqq 0$ を開集合 $W \subset \mathbb{R}^d$ 上の非負解析関数とし、$K(w) = 0$ を満たす $w \in W$ が存在するとする。このとき、ある d 次元多様体 $\mathcal{M}$ と $\mathcal{M}$ 上の局所座標が取りうる値の集合 $\mathcal{U}$ からの解析写像 $g: \mathcal{U} \to W$ が存在して、$\mathcal{M}$ の局所座標ごとに、

$$K\big(g(u)\big) = u_1^{2k_1} u_2^{2k_2} \cdots u_d^{2k_d}$$

$$|g'(u)| = b(u) \left| u_1^{h_1} u_2^{h_2} \cdots u_d^{h_d} \right|$$

が成立するようにできる。ここで $|g'(u)|$ は $w = g(u)$ のヤコビアンであり、$b(u) > 0$ は 0 にならない解析関数であり、

$$k = (k_1, k_2, \cdots, k_d), \quad h = (h_1, h_2, \cdots, h_d)$$

は非負の正数の集合である。但し $(k_1, k_2, \cdots, k_d)$ のうち少なくともどれか一つは 0 ではない。

標準形

定理6にまつわる大事な注意：

- 定理6を満たす g はコンパクト集合の引き戻しがコンパクトである（注意34）。
  - ので、1の分割（というか事前分布の分割）ができる（注意37）。
- 平均誤差と事前分布は同時特異点解消が可能である（注意36）。
  - もし事前分布にゼロとなる点があった場合は、定理8で困るので、同時特異点解消しておくこと。