＿人人人人人人人＿
＞　まだ作成中　＜
￣Y^Y^Y^Y^Y^Y￣

「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 $\boxed{\quad ？\quad}$ 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。

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

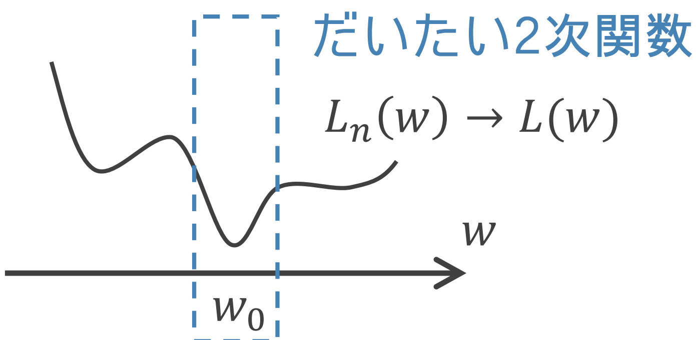
事後分布上の平均

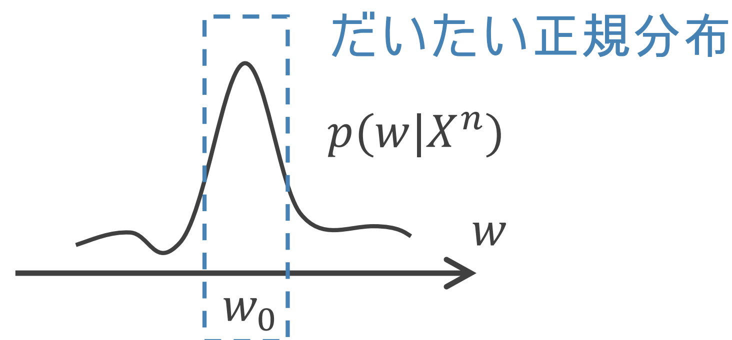
事後分布上の分散

→ 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)

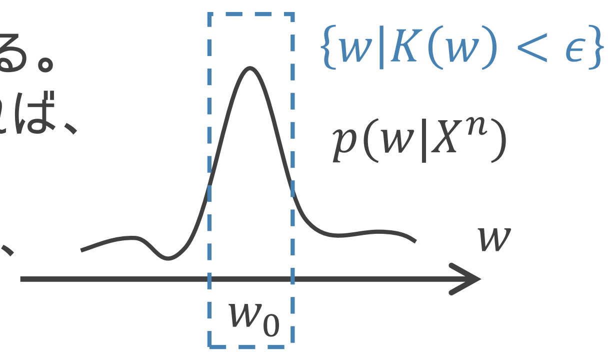
→ 平均対数損失 $L(w)$ は「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものと仮定してみると、G_n は以下のような確率変数であると示せる。ゴール達成！

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

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

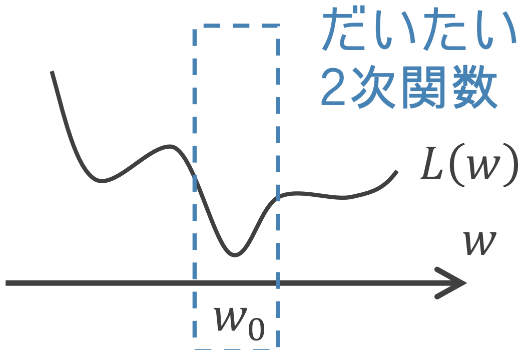
（サンプルの選び方に依存する確率変数）

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。

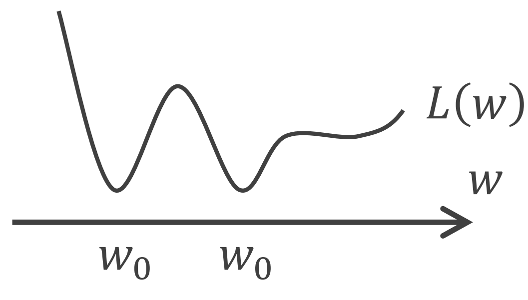


3章を終えて普通に気になること

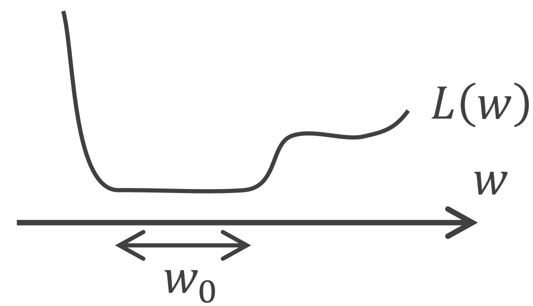
平均対数損失 $L(w)$ が「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものではない場合は、ベイズ推測は「推測としてこのようにいい」といえないの？？



こういうときはいえた。



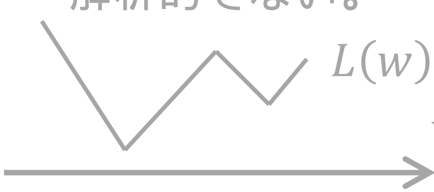
こういうときは？

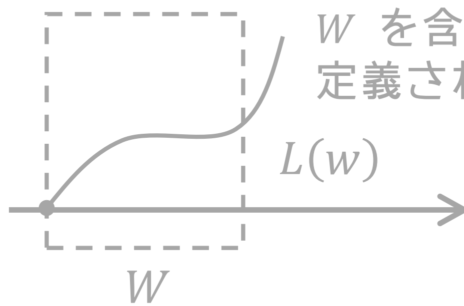


こういうときは？

※ 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。

解析的でない。



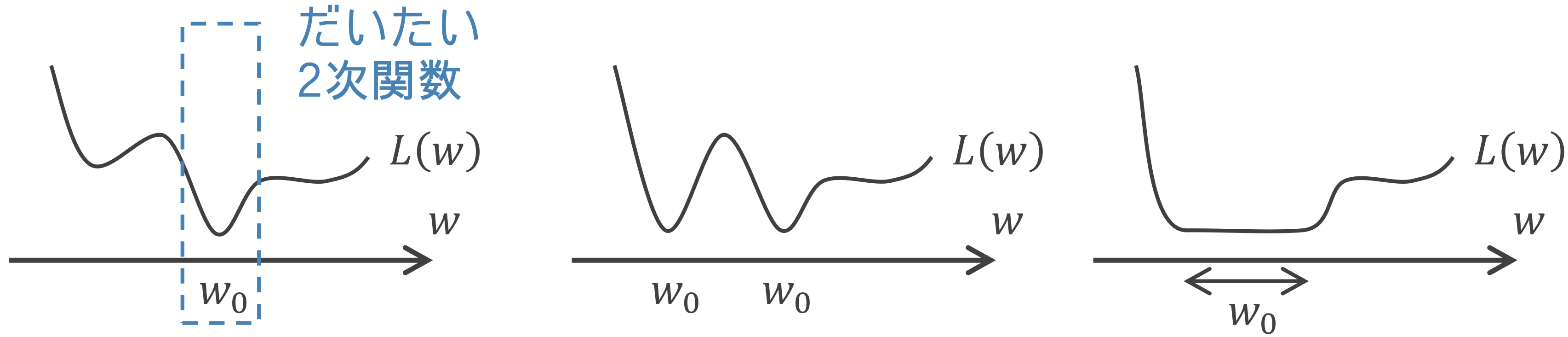


W を含む開集合で定義されていない。

- 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。

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 $\boxed{\quad ？ \quad}$ 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)）。

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。

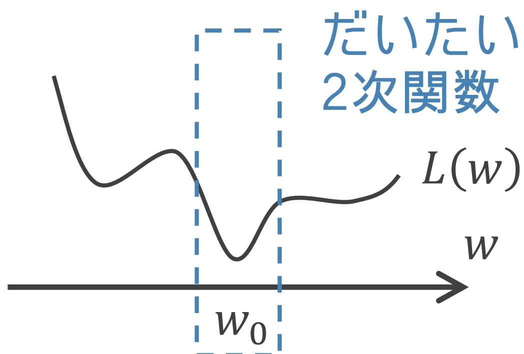
→ G_n は以下のようにキュムラント展開できる。

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

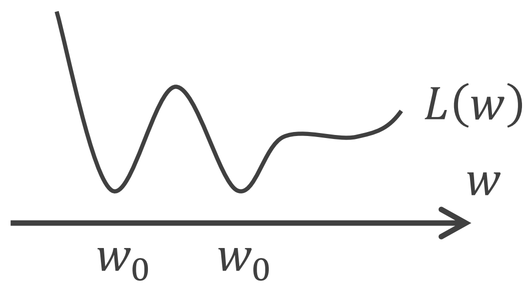
$$-\frac{1}{3!}\mathcal{G}^{(3)}(0) - \cdots$$

→ 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。

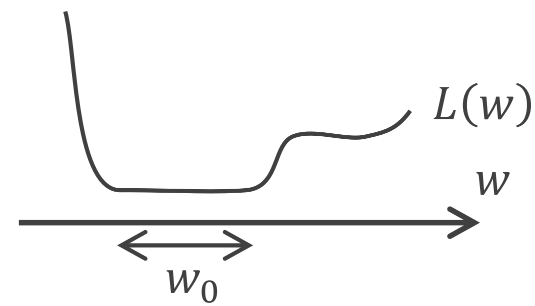
→ $L(w)$ に何も仮定をおかないとわからない！



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

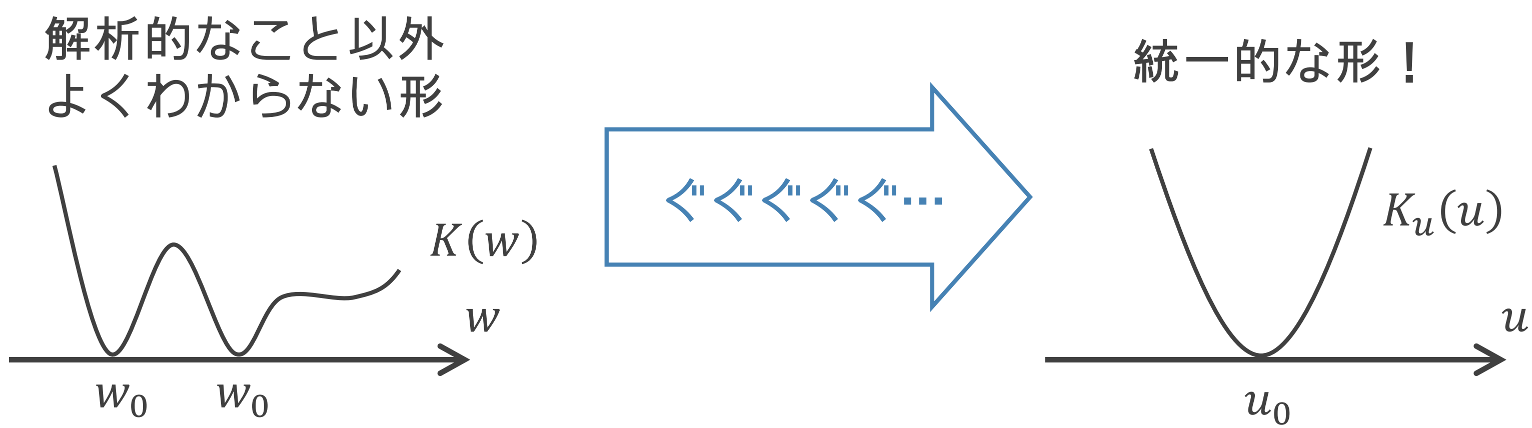
事後分布 $p(w|X^n) \propto \exp\left(-n\beta L_n(w)\right)\varphi(w)$ の形を知りたい。

→ $p(w|X^n) \propto \exp\left(-n\beta K_n(w)\right)\varphi(w)$ 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 $g\colon \mathcal{U} \to W$ 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、

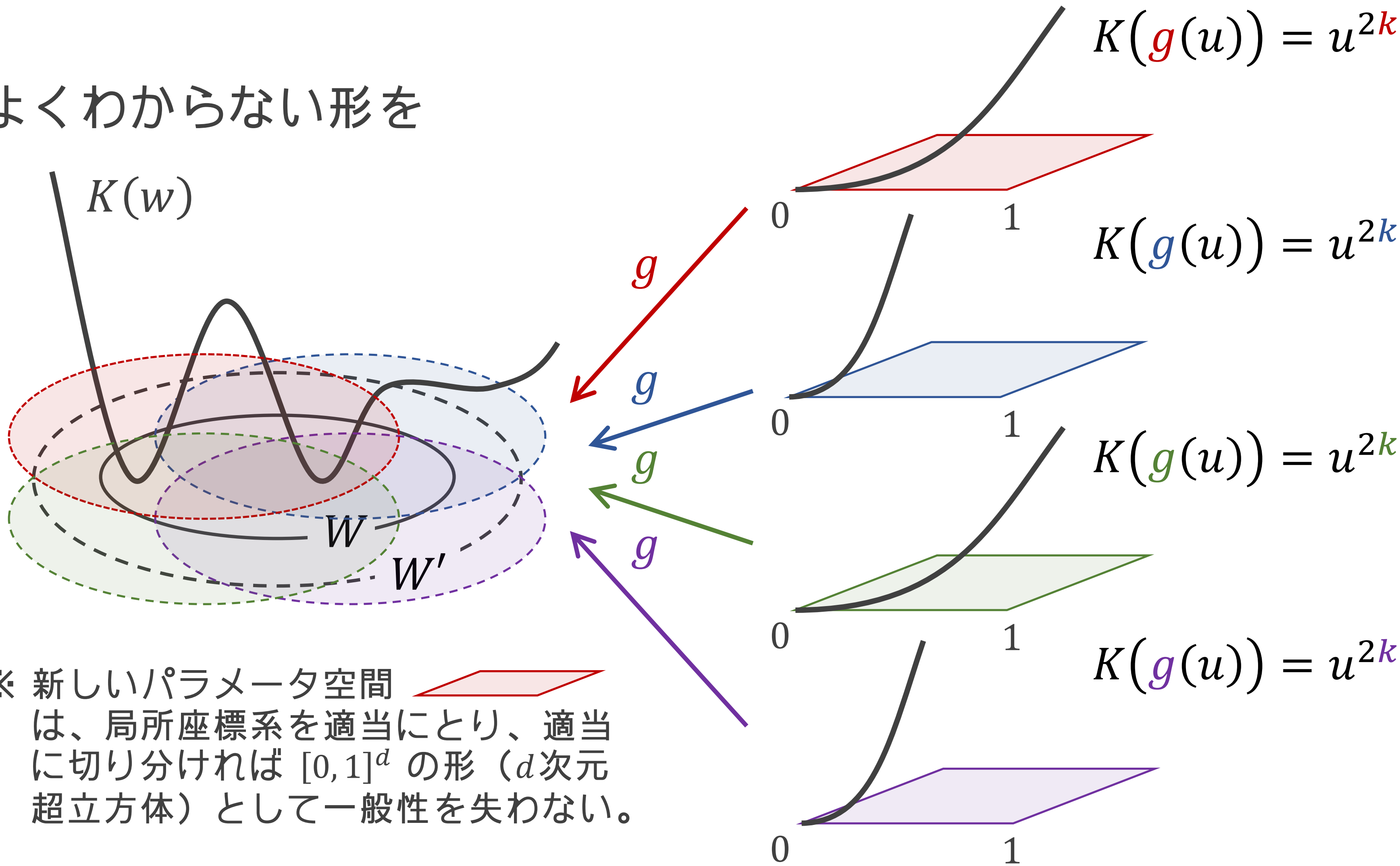
$$k = (k_1, k_2, \cdots, k_d), \quad h = (h_1, h_2, \cdots, h_d)$$

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

よくわからない形を

有限個の統一的な形にできる
（ g, k はそれぞれ異なる）

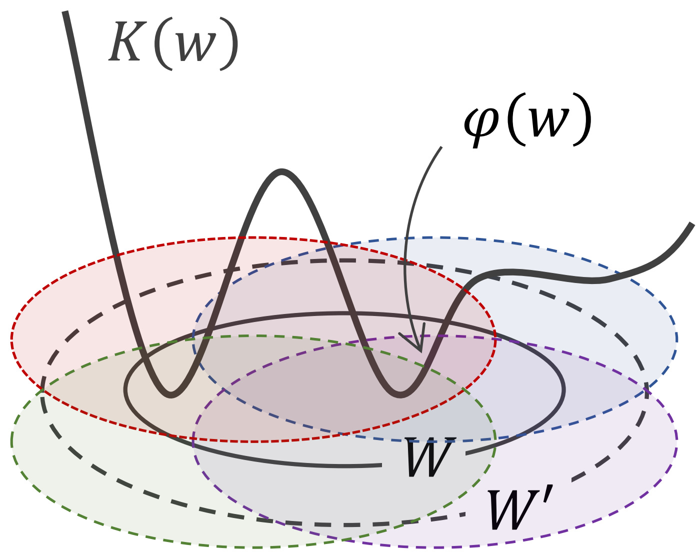


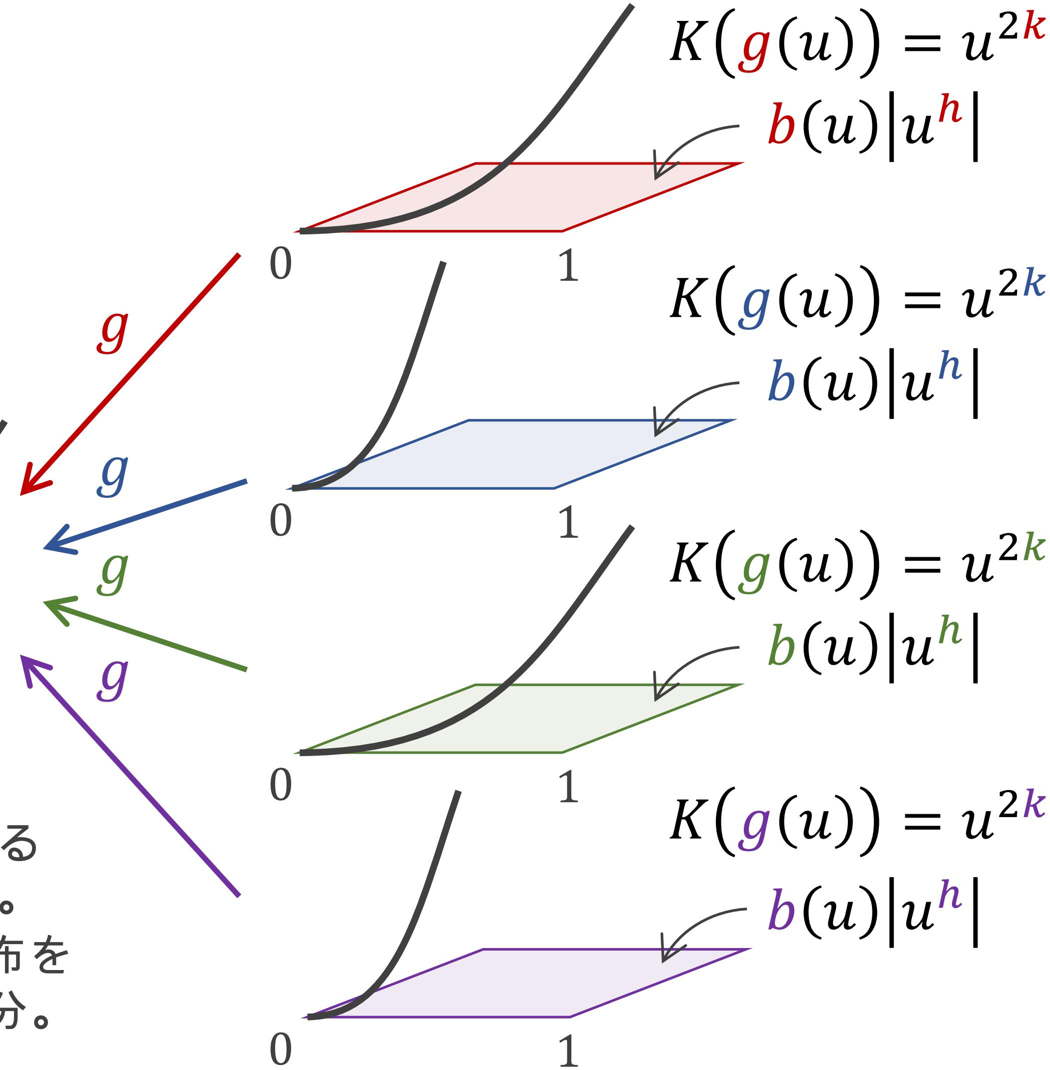
※ 新しいパラメータ空間は、局所座標系を適当にとり、適当に切り分ければ $[0,1]^d$ の形（d次元超立方体）として一般性を失わない。

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



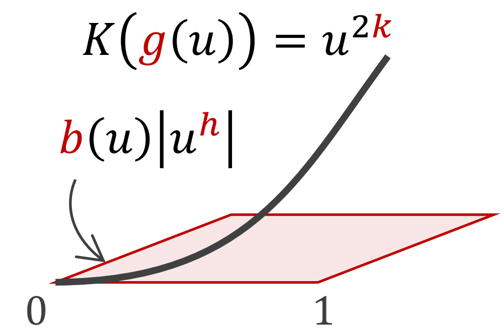


※ W がコンパクトなら適当に分ける
　ことができる（参考. 1の分割）。
※ 新しい密度 $b(u)|u^h|$ は、事前分布を
　分けた分 × 空間を g で歪めた分。

事後分布の標準形

事後分布 $p(w|X^n) \propto \exp\big(-n\beta K_n(w)\big)\,\varphi(w)$ 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。



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

この形を考えればよい。

→ $K_n\big(g(u)\big)$ は $K\big(g(u)\big) = u^{2k}$ 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、

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

とかける。

→ つまり、事後分布の標準形はこうなる。

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

確率的にゆらがない　確率的にゆらぐ

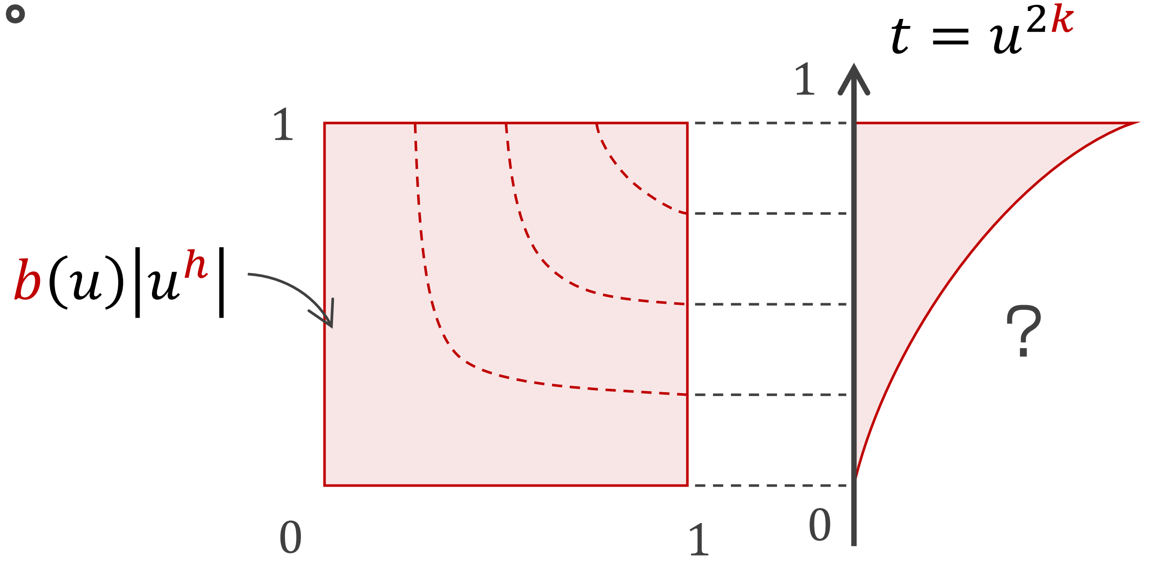
事後分布の標準形

じゃあ、$p_{赤}(u|X^n) \propto \exp\left(-n\beta u^{2k} + \sqrt{n}\beta u^k \xi_n(u)\right) b(u)|u^h|$ 上で $K(g(u)) = u^{2k}$ の平均や $f(x, g(u)) = u^k a(x, u)$ 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 $b(u)|u^h|$ 上での $t = u^{2k}$ の密度を知りたい（こっちの密度を状態密度という）。

→ t の関数に $\delta(t - u^{2k}) b(u)|u^h|$ をかけて積分するとどうなるか調べればよい。



状態密度

$\delta(t - u^{2k}) b(u) |u^h|$ 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。

実はメリン変換という変換をすると、 $t \to 0$ で支配的な成分が取り出せる。

まとめ（仮）

事後分布上での期待値や分散をとる準備ができた（予定）。

＿人人人人人人人人人人人人人人人＿
＞　つづきは4章後半を読んでね！　＜
￣Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y￣