＿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 $p^*(x) \equiv \int_W p(x|w)p(x|X^n)dw$ 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

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

（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。

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

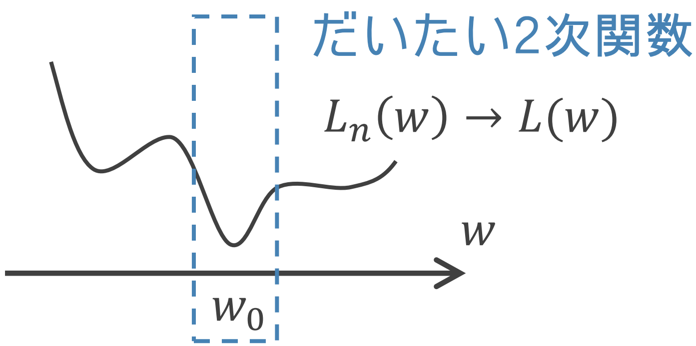
事後分布上の平均

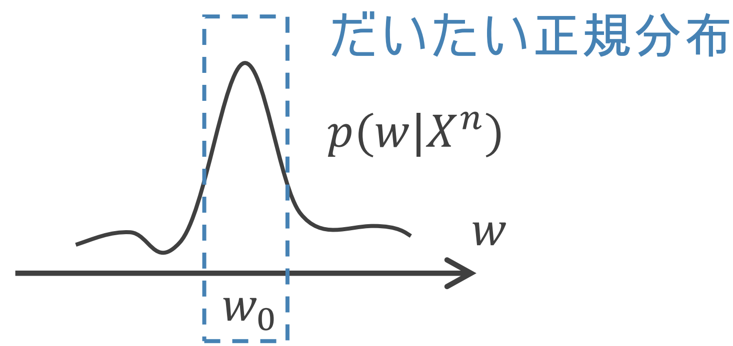
事後分布上の分散

→ 事後分布 $p(w|X^n) \propto \exp\big(-n\beta L_n(w)\big)\varphi(w)$ 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)

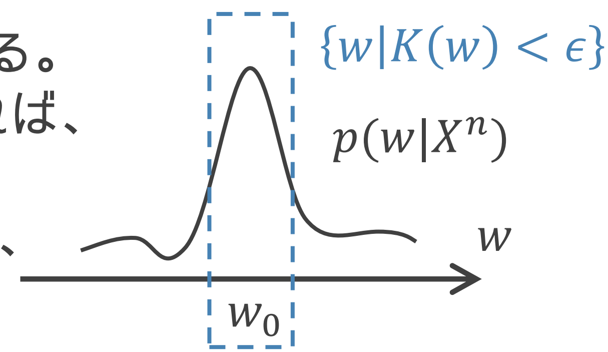
→ 平均対数損失 $L(w)$ は「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものと仮定してみると、G_n は以下のような確率変数であると示せる。ゴール達成！

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

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

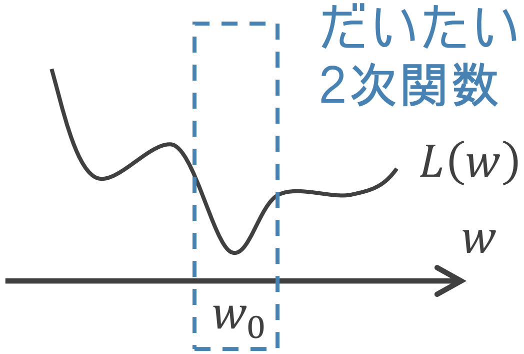
（サンプルの選び方に依存する確率変数）

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。

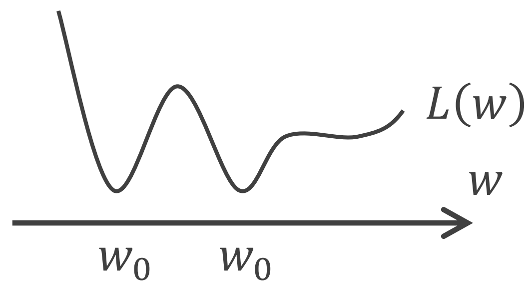


3章を終えて普通に気になること

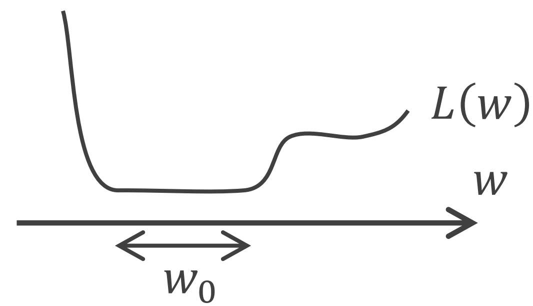
平均対数損失 $L(w)$ が「ただ1つの最小点 w_0 をもち、w_0 でのヘッセ行列が正定値である」ものではない場合は、ベイズ推測は「推測としてこのようにいい」といえないの？？



こういうときはいえた。



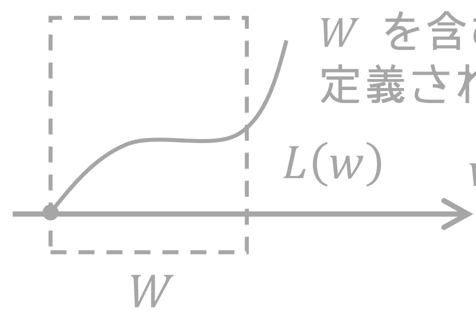
こういうときは？



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。

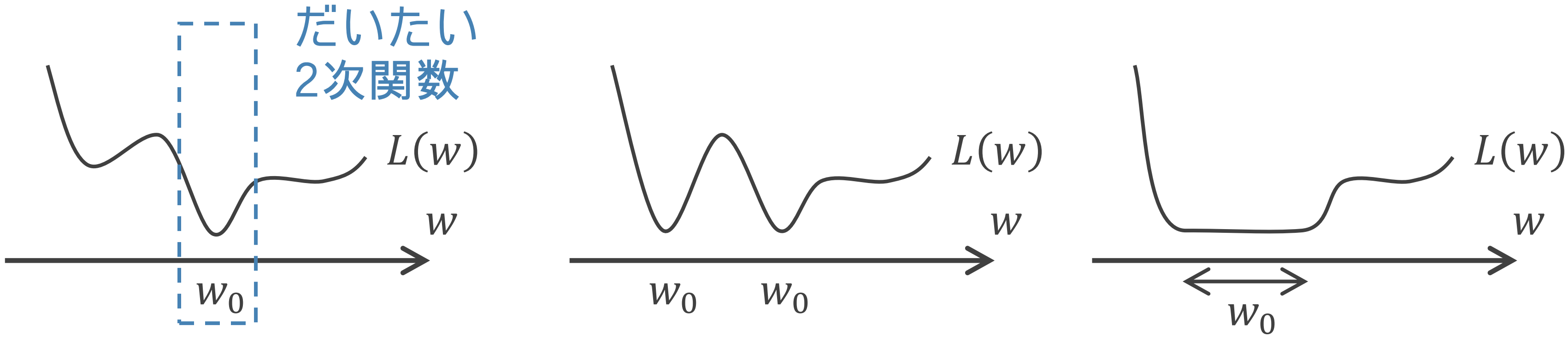


W を含む開集合で定義されていない。

- 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

ここまでもこれからも最低限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とする。
2. 対数尤度比 $f(x, w_0, w)$ が相対的に有限な分散をもつ。

いきなり立ちはだかる壁

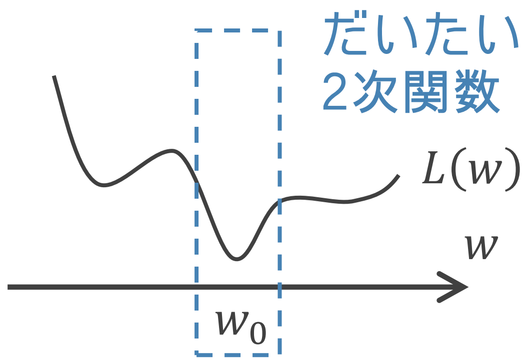
汎化損失 $G_n = -\mathbb{E}_X[\log \mathbb{E}_w[p(X|w)]]$ のしたがう分布を知りたい。

→ G_n は以下のようにキュムラント展開できる。

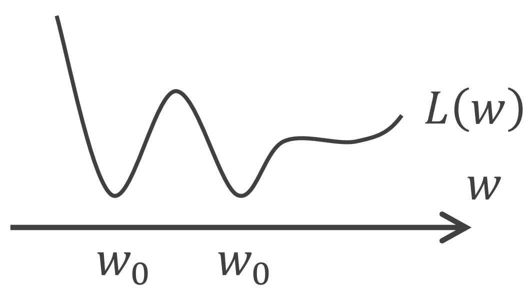
$$G_n = L(w_0) + \mathbb{E}_w[K(w)] - \frac{1}{2}\mathbb{E}_X[\mathbb{E}_w[f(X,w)^2] - \mathbb{E}_w[f(X,w)]^2] - \frac{1}{3!}\mathcal{G}^{(3)}(0) - \cdots$$

→ 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。

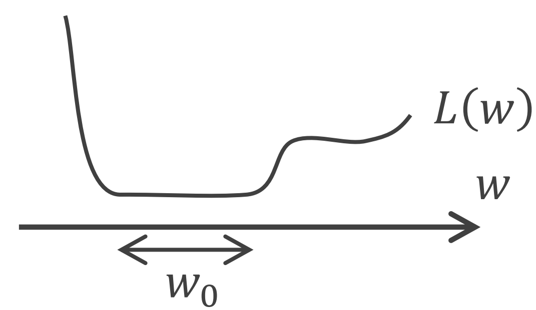
→ $L(w)$ に何も仮定をおかないとわからない！



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

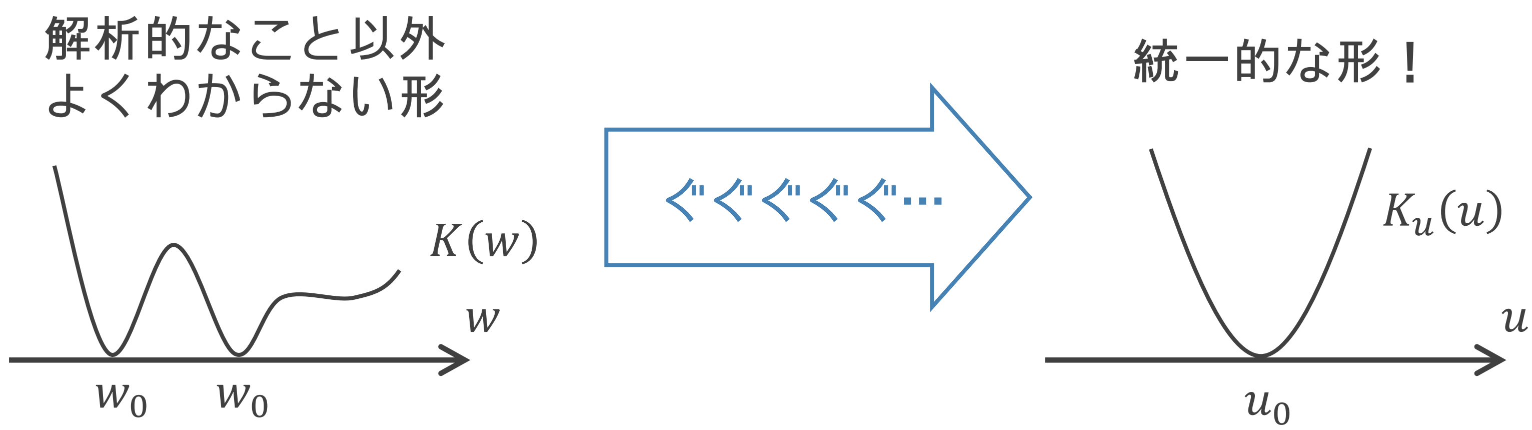
事後分布 $p(w|X^n) \propto \exp\left(-n\beta L_n(w)\right)\varphi(w)$ の形を知りたい。

→ $p(w|X^n) \propto \exp\left(-n\beta K_n(w)\right)\varphi(w)$ 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）（広中の定理）

$K(w) \geqq 0$ 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。

待って

定理6（特異点解消定理）（広中の定理）

（前略）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、

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

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

が成立するようにできる。（後略）

＿人人人人人人人人人人人人人人人人＿

＞　多様体とその上の局所座標って何　＜

￣Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y^Y￣

工事中…

多様体ってこういうものである条件下で1の分割ができるよ。

誤差関数と事前分布を同時に特異点解消できるよ。そういう特異点解消で誤差関数を標準形にするよ。経験誤差関数も標準形になるよ。つまり事後分布を標準形にできたよ。

ただこのままだと、肝心の事後分布上での「平均誤差の平均」とか「対数尤度比の分散」がとりづらいから $t = u^{2k}$ と変換したいよ。そうすると t の世界での密度を求めないといけないけどそれは定理8だよ。

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￣