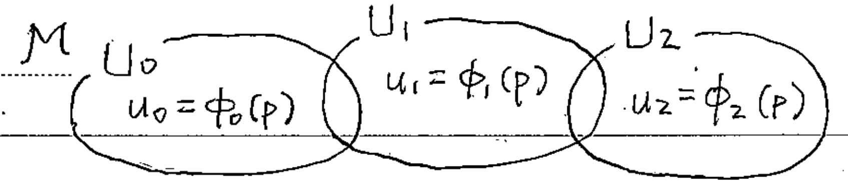


(U_0,ϕ_0) チャート0
(U_1,ϕ_1) チャート1
(U_2,ϕ_2) チャート2

定理6(特異点解消定理のベイズ一般理論向け版)

$K(w)$：開集合 $W\subset\mathbb{R}^d$ 上の非負値をとる解析関数。
$K(w)=0$ を満たす $w\in W$ が存在するとする。

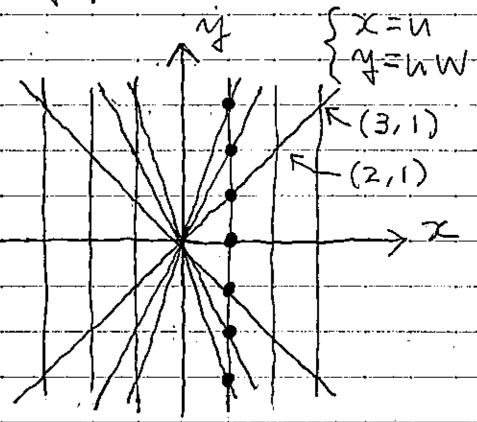
このような任意の $K(w)$ に対し、ある d 次元多様体 M と
M 上の局所座標から W への解析写像 g が存在して
M の局所座標系ごとに式(4.1),(4.2)が成立するようにできる。
但し $k=(k_1,k_2,\cdots,k_d)$, $h=(h_1,h_2,\cdots,h_d)$ は非負整数の集合
であり、局所座標系ごとに異なる。$(k_1,k_2,\cdots,k_d)$ のうち少なくとも
一つは0でない。

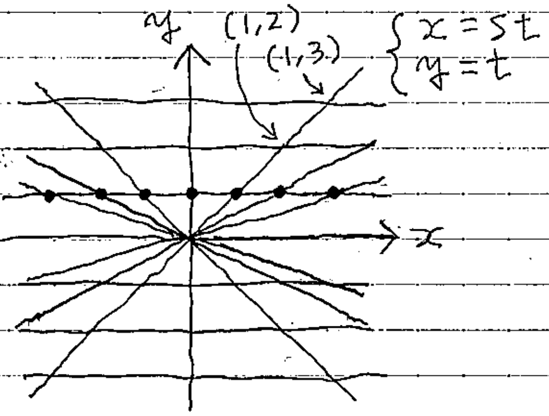
$$K(g(u))=u_1^{2k_1}u_2^{2k_2}\cdots u_d^{2k_d}=u^{2k} \qquad (4.1)$$

$$|g'(u)|=b(u)\,|u_1^{h_1}u_2^{h_2}\cdots u_d^{h_d}|=b(u)\,|u^h| \qquad (4.2)$$

但し $|g'(u)|$ は g のヤコビアン(ヤコビ行列の行列式)の絶対値で、
$b(u)$ は正の値をとる解析関数(おそらく、局所座標系ごとに異なる)。

例14. 例14をやる前に、例13で多様体を定義しているのでそれをみる。
xy 平面の上の各点に、新しい座標 (u,w) と (s,t) をわりあてるとする。





新しい座標 (u,w) は、
u はその点の x 座標と同じ。
w はその点の y 座標を
その点の x 座標で割ったもの。
(xy 平面上に (u,w) の格子を
かくと上図のようになる。)
しかし、xy 平面上の y 軸上
(原点除く)には (u,w) をわりあてられない。

新しい座標 (s,t) は、
t はその点の y 座標と同じ。
s はその点の x 座標を
その点の y 座標で割ったもの。
(xy 平面上に (s,t) の格子を
かくと上図のようになる。)
しかし、xy 平面上の x 軸上
(原点除く)には (s,t) をわりあて
られない。

・位相空間 M のアトラス；位相空間 M を覆いつくすチャートの集合。
・位相空間 M のチャート；M 上の開集合 U から $\mathbb{R}^d$ 上の開集合 V への
同相写像 $\phi: U\to V$ があったら、(U,ϕ) を M のチャートという。

$\{(u,w)\}$ は y 軸上(原点除く)を覆わないし、
$\{(s,t)\}$ は x 軸上(原点除く)を覆わないので、
$\{(u,w)\}$ と $\{(s,t)\}$ はそれぞれ単独では $\mathbb{R}^2$ のアトラスにならないが、
これらを組み合わせれば $\mathbb{R}^2$ を覆い尽くすので $\mathbb{R}^2$ のアトラスになる。
$\mathbb{R}^2$ にこのようなアトラスをわりあてた多様体を M とする(これが例13の多様体)。

いま $\mathbb{R}^2$ 上の関数 $K(x,y)=x^2+y^2$ を考える。
この関数は $K(x,y)=f_x(x)f_y(y)$ の形をしていない。
座標を取りかえてこのような因数分解ができるようにしたい。

そこで、先の多様体 M 上の局所座標から $\mathbb{R}^2$ への写像 g_0, g_1

$$(x,y)=g_0(u,w)=(u,uw)$$
$$(x,y)=g_1(s,t)=(st,t)$$

を考えると、以下のように因数分解できる(ただ定理6の形とはちょっと違う)。

$$K(g_0(u,w))=u^2+u^2w^2=u^2(1+w^2)$$
$$K(g_1(s,t))=s^2t^2+t^2=t^2(s^2+1)$$

また、$|g_0'(u,w)|=\left|\det\begin{pmatrix}1 & w\\ 0 & u\end{pmatrix}\right|=|u|$ より、g_0, g_1 のヤコビアンも
因数分解できている。

$$|g_1'(s,t)|=\left|\det\begin{pmatrix}t & 0\\ s & 1\end{pmatrix}\right|=|t|$$

定理6の形と合うようにやり直す。写像 g_0, g_1 を、

$$(x,y)=g_0(u,w)=\left(\frac{u}{\sqrt{1+w^2}},\frac{uw}{\sqrt{1+w^2}}\right)$$
$$(x,y)=g_1(s,t)=\left(\frac{st}{\sqrt{1+s^2}},\frac{t}{\sqrt{1+s^2}}\right)$$

とすると、

$$K(g_0(u,w))=\left(\frac{u}{\sqrt{1+w^2}}\right)^2+\left(\frac{uw}{\sqrt{1+w^2}}\right)^2=u^2$$
$$K(g_1(s,t))=\left(\frac{st}{\sqrt{1+s^2}}\right)^2+\left(\frac{t}{\sqrt{1+s^2}}\right)^2=t^2$$

$$|g_0'(u,w)|=\left|\det\begin{pmatrix}(1+w^2)^{-\frac12} & w(1+w^2)^{-\frac12}\\ -uw(1+w^2)^{-\frac32} & u(1+w^2)^{-\frac12}-uw^2(1+w^2)^{-\frac32}\end{pmatrix}\right|=\frac{1}{1+w^2}|u|$$

$$|g_1'(s,t)|=\left|\det\begin{pmatrix}t(1+s^2)^{-\frac12}-s^2t(1+s^2)^{-\frac32} & -st(1+s^2)^{-\frac12}\\ s(1+s^2)^{-\frac12} & (1+s^2)^{-\frac12}\end{pmatrix}\right|=\frac{1}{1+s^2}|t|$$

$k^{(0)}=(1,0)$
$k^{(1)}=(0,1)$
$h^{(0)}=(1,0)$
$h^{(1)}=(0,1)$
$b^{(0)}(u)=\dfrac{1}{1+u_2^2}$
$b^{(1)}(u)=\dfrac{1}{1+u_1^2}$
定理6の形と合う！

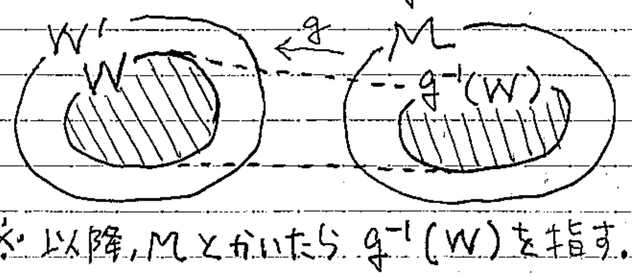
No.
Date

定義域の各点においてその点の周りでのテイラー展開と一致

$\mathbb{R}^d$ の開集合であって、じゅうぶん大きな半径 $r>0$ の球で覆うことができる。

☆注意34

- 本書ではパラメータ集合 W は $\mathbb{R}^d$ のコンパクトな部分集合であり、かつ、開集合 $W' \supset W$ が存在して平均誤差 $K(w)$ は W' 上の解析関数であると仮定する（ので定理6がつかえる）。
- 定理6を満たす g はコンパクト集合への引き戻しがコンパクトである。すなわち、$C \subset W'$ がコンパクトであれば $g^{-1}(C)=\{u \mid g(u) \in C\}$ もコンパクトである。よって $g^{-1}(W)$ もコンパクトである。



位相空間の部分集合がコンパクト；任意の開被覆が有限部分開被覆をもつ。

※以降、M とかいたら $g^{-1}(W)$ を指す。

※1の分割（後述）が存在するためにはコンパクトが要る。

☆注意36

- ところで、事後分布は $p(w|X^n) \propto \exp(-n\beta K_n(w))\varphi(w)$ だった。

これを、W 上の分布から M 上の分布にすると、以下となる（※仮に、チャート1枚で覆われているとする。）

$$\underbrace{\exp(-n\beta K_n(g(u)))}\ \underbrace{\varphi(g(u))}\ \underbrace{|g'(u)|}$$

定理6から $K(g(u))$ は標準形 u^{2k} になる。それに近づく。

?

定理6から標準形 $b(u)|u^h|$ になる。$b(u)$ は正の値をとる解析関数。

（確率変数の変換 $u=g^{-1}(w)$ のとき、$p_u(u)=|g'(u)|p_w(g(u))$）

$\varphi(g(u))$ をどうすればいいのか。

① $\varphi(w)$ が $w \in W$ でゼロとならないとき、$\varphi(g(u))$ は $b(u)$ に含めてしまえばよい。

定理8のため

（但し、$b(u)$ は無限回微分できないといけないので、そういう都合のよい $\varphi(w)$ をつかうことにする。）

② $\varphi(w)$ が $w \in W$ でゼロとなるとき、$\varphi(w)=\underbrace{\varphi_1(w)}_{\text{ゼロになる解析関数。}}\underbrace{\varphi_2(w)}_{\text{ゼロにならない\&都合がいい。}}$ として、$\varphi_2(w)$ は $b(u)$ に含めてしまう。（但し、$b(u)$ は無限回微分〜〜。）

No.
Date

そういう都合のいい $\varphi(w)$ をつかう。

$\varphi_1(w)$ が解析関数ならば

$\varphi_1(w)$ については、$K(w)$ に定理6を適用するとき $\varphi_1(g(u))=|u^{h'}|$ となるように g を選ぶことができる。

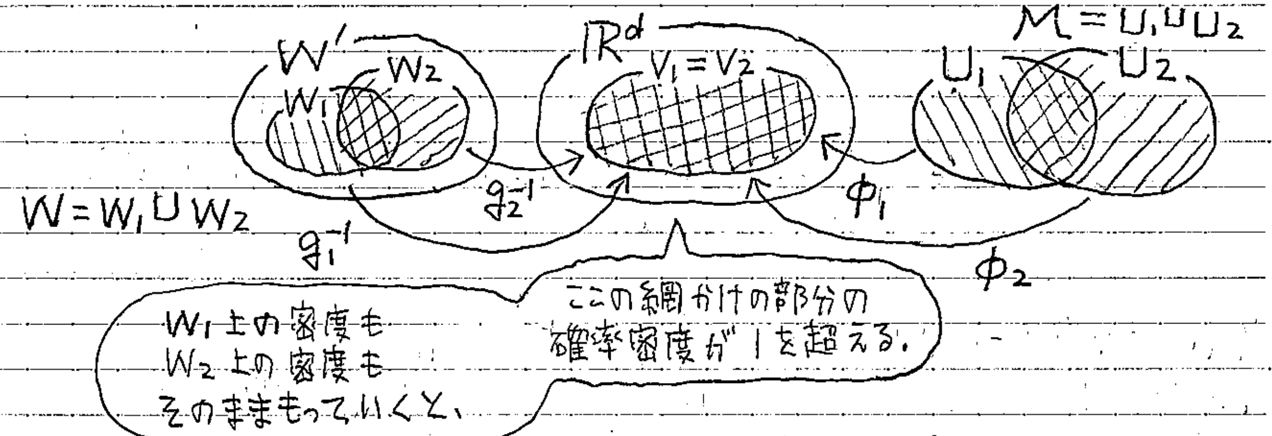
（「$K(w)$ と $\varphi_1(w)$ は同時特異点解消が可能である。」）

なので結局事後分布は $\exp(-n\beta K_n(g(u)))\,b(u)\,|u^h|$ の形になる。

（$b(u)$ は正の値をとり、無限回微分できる関数。）

☆注意37（定義15）

- しかし、M はチャート1枚で覆われているとは限らない。どのチャートについても注意36の考え方をすると、確率密度がくるってくる。



そこで、M がコンパクトなとき「1の分割」ができることを利用する。
つまり、M がコンパクトなとき、M 上の任意の点 p に対して
$f_1(p)+f_2(p)=1$ であり、かつ、$f_1(p) \geq 0$, $f_2(p) \geq 0$ であり
$f_1(p)$ は U_1 の外では0、$f_2(p)$ は U_2 の外では0 となるような
写像 f_1, f_2 が存在する（U_1 と U_2 が重なる部分では「1を分け合う」）。
実際には、1を分け合うのではなく、事前分布 $\varphi(w)$ を分け合う。
$\varphi(w)=\sum_\alpha \varphi_\alpha(w)$ のように各チャートが担当する確率密度を分けておけば、確率密度がくるうことはない。

- また、上図でいう V_1, V_2 は $\mathbb{R}^d$ の部分集合としてどんな形なのか気になるが、（M がコンパクトなので）適宜チャートの覆い方を細分化なりすれば各 V_α は $[0,1]^d$（超立方体）の形をしてるとして一般性を失わない。

ここから実際に定理6を用いて事後分布を標準形にしていく。
まず平均誤差$K(w)$に定理6を適用すると以下となる。

$$K(g(u)) = u^{2k}$$

また、この章では相対的に有限な分散を仮定しているので、任意のwで

$$\mathbb{E}_X[f(X,w)] \geq c_1 \mathbb{E}_X[f(X,w)^2] \quad (c_1>0\text{は定数})$$

$$\Leftrightarrow u^{2k} \geq c_1 \int q(x) f(x,g(u))^2 dx$$

なので、この右辺はu^{2k}で割り切れる。

$\Rightarrow f(x,g(u))^2$ は u^{2k} で割り切れる。

$\Rightarrow f(x,g(u))$ は u^k で割り切れる。

補題20

対数尤度比 $f(x,g(u))$ が u の解析関数であるとする。
このとき、ある解析関数 $a(x,u)$ が存在して、以下が成り立つ。

$$f(x,g(u)) = u^k a(x,u)$$

もし成り立たないとすると $f(x,g(u)) = u^k a(x,u) + b(x,u)$
(但し $b(x,u)$ は u^k で割り切れず、常にゼロでもない) となるが、
そうすると、以下が成り立たなければならない(任意のuで)。

$$u^{2k} \geq c_1 \int q(x)\left(u^k a(x,u) + b(x,u)\right)^2 dx$$

$$1 \geq c_1 \int q(x)\left(a(x,u)^2 + \frac{2}{u^k} a(x,u) b(x,u) + \frac{b(x,u)^2}{u^{2k}}\right) dx$$

ここで、$K(w_0)=0$ より、$g(u_0)=w_0$ となる u_0 では、

$K(g(u_0)) = u_0^{2k} = 0 \Rightarrow u_0^k = 0$ である。

このu_0では $\left|\frac{b(x,u_0)}{u_0^k}\right|$ は正の無限大になってしまう。これは c_1が正の定数であることに反する。

$$\left.\begin{aligned} f(x,g(u)) &= u^k a(x,u) \\ \underset{u^k}{\underset{\parallel}{K(g(u))}} &= \int f(x,g(u)) q(x) dx \end{aligned}\right\} u^k = \int a(x,u) q(x) dx$$

uを添字とした↓

$\xi_n(u)$を以下のように定義する(サンプルの選び方に依存する確率過程)。

$$\xi_n(u) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(u^k - a(X_i,u)\right) = \sqrt{n}\, u^k - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(X_i,u)$$

$$\begin{aligned}\mathbb{E}[\xi_n(u)] &= \sqrt{n}\, u^k - \frac{1}{\sqrt{n}} \mathbb{E}\left[\sum_{i=1}^{n} a(X_i,u)\right] \\ &= \sqrt{n}\, u^k - \frac{1}{\sqrt{n}} \int \left(\sum_{i=1}^{n} a(x_i,u)\right) q(x_1)\cdots q(x_n)\, dx_1 \cdots dx_n \\ &= \sqrt{n}\, u^k - \frac{1}{\sqrt{n}} n u^k = 0\end{aligned}$$

サンプルの選び方に対する期待値

$$\begin{aligned}\mathbb{E}[\xi_n(u)\xi_n(v)] &= \mathbb{E}\left[\left(\sqrt{n}\, u^k - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(X_i,u)\right)\left(\sqrt{n}\, v^k - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a(X_j,v)\right)\right] \\ &= n u^k v^k - n u^k v^k - n u^k v^k + \frac{1}{n} \mathbb{E}\left[\left(\sum_{i=1}^{n} a(X_i,u)\right)\left(\sum_{j=1}^{n} a(X_j,v)\right)\right] \\ &= -n u^k v^k + \frac{1}{n} \int \left(\sum_{i=1}^{n}\sum_{j=1}^{n} a(x_i,u) a(x_j,v)\right) q(x_1)\cdots dx_1 \cdots \\ &= -n u^k v^k + \frac{1}{n} \cdot n \int a(x,u) a(x,v) q(x)\, d(x) \\ &\quad + \frac{1}{n}(n^2-n) \left(\int a(x,u) q(x) dx\right)\left(\int a(x,v) q(x) dx\right) \\ &= -n u^k v^k + \mathbb{E}_X[a(X,u) a(X,v)] + (n-1) u^k v^k \\ &= \mathbb{E}_X[a(X,u) a(X,v)] - u^k v^k\end{aligned}$$

なので、確率過程 $\xi_n(u)$ は平均 0、自己共分散 $\mathbb{E}_X[a(X,u)a(X,v)] - u^k v^k$。

また、$\xi_n(u)$ は n を大きくしたとき、同じ平均と自己共分散をもつガウス過程に法則収束する($\because$ 8.5.3の(2))。

これを $\xi(u)$ とかく。

$\xi(u)$ に関する平均を $\mathbb{E}_\xi[\ ]$ とかく。

補題2

真の分布が確率モデルで実現可能ならば、$K(g(u))=0$ なる u で
$\mathbb{E}[\xi_n(u)^2] = \mathbb{E}_\xi[\xi(u)^2] = \mathbb{E}_X[a(X,u)^2] = 2$ である。

$$\mathbb{E}[\xi_n(u)^2] = \mathbb{E}_X[a(X,u)^2] - u^{2k} \fallingdotseq \mathbb{E}[a(X,u)^2]$$
$$\mathbb{E}_\xi[\xi(u)^2] = \mathbb{E}_X[a(X,u)^2] - u^{2k} \fallingdotseq \mathbb{E}[a(X,u)^2]$$
$$\mathbb{E}_X[a(X,u)^2] = \int a(x,u)^2 q(x)\,dx = \underbrace{\frac{1}{u^{2k}}\int f(x,g(u))^2 q(x)\,dx}$$

補題4,5の帰結より、ここが $K(g(u))$ なる u に近づくとき $2u^{2k}$ に近づく。

定理7

- 経験誤差は $nK_n(g(u)) = nu^{2k} - \sqrt{n}\,u^k \xi_n(u)$.
- 事前分布は $|u^h|\,b(u)$.
- $\xi_n(u)$ は n を大きくすると $\xi(u)$ に法則収束する。

$$nK_n(g(u)) = \sum_{i=1}^n f(X_i, g(u)) = \sum_{i=1}^n u^k a(X_i, u) = u^k\left(nu^k - \sqrt{n}\,\xi_n(u)\right).$$

$$p(w|X^n) \propto \exp(-n\beta K_n(w))\,\varphi(w)$$

は、M の世界では以下のようになる。

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

ここは n が増えると事後分布がとがっていくようなイメージ。　ここは $K(w)$ と $K_n(w)$ のずれ。

後の都合により、u^{2k} の密度を考えていきたい。

・デルタ関数 —— 必ず他の関数との積分の形で用いて、$x=0$ の値を抜き出すもの。

$$\int \delta(x)\varphi(x)\,dx = \varphi(0)$$

- $\int \delta(-x)\varphi(x)\,dx = \int \delta(x')\varphi(-x')\,dx' = \varphi(0)$ より、
  $\int [\delta(x) - \delta(-x)]\varphi(x)\,dx = 0$ なので $\delta(x) - \delta(-x) = 0$
- $\int \delta(t-x)\varphi(x)\,dx = \int \delta(x')\varphi(t-x')\,dx' = \varphi(t)$
- $\int \delta(ax)\varphi(x)\,dx = \begin{cases} \int \delta(x')\varphi\left(\frac{x'}{a}\right)\frac{dx'}{a} = \frac{1}{a}\varphi(0) & (a>0) \\ -\int \delta(x')\varphi\left(\frac{x'}{a}\right)\frac{dx'}{a} = -\frac{1}{a}\varphi(0) & (a<0) \end{cases}$
- $\Theta(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x<0) \end{cases}$ と定義すると $\int \Theta(t-x)\varphi(x)\,dx = \int_{-\infty}^{t}\varphi(x)\,dx$

なので、$\frac{d}{dt}\int \Theta(t-x)\varphi(x)\,dx = \varphi(t)$ より、$\frac{d}{dt}\Theta(t-x) = \delta(t-x)$

- $\int \Theta(t-x^2-y^2-z^2)\varphi(x,y,z)\,dx\,dy\,dz = \int_{x^2+y^2+z^2<t}\varphi(x,y,z)\,dx\,dy\,dz$

なので、

$$\int \delta(t-x^2-y^2-z^2)\varphi(x,y,z)\,dx\,dy\,dz = \underbrace{\frac{d}{dt}\left(\int_{x^2+y^2+z^2<t}\varphi(x,y,z)\,dx\,dy\,dz\right)}_{(*)}$$

・特に $t \to +0$ のとき、原点の近傍で

$\varphi(x,y,z) = \varphi(0) + \underline{O\left(\sqrt{x^2+y^2+z^2}\right)}$? なので、

$$(*) = \frac{d}{dt}\left(\varphi(0)\frac{4\pi}{3}t^{3/2} + O(\sqrt{t})\frac{4\pi}{3}t^{3/2}\right) = 2\pi\varphi(0)\sqrt{t} + O(t)$$

となることから $t \to 0$ で $\delta(t-x^2-y^2-z^2)$ は以下に近づく。

$$2\pi\delta(x)\delta(y)\delta(z)\sqrt{t} + \underline{O(\sqrt{t})}\ ?$$

No.

Date

定義17

多重指数 $k=(k_j)$, $h=(h_j)$ から定まる以下の値を考える。

$$\left(\frac{h_j+1}{2k_j}\right) \quad (j=1,\cdots,d)$$

$k_j=0$ のときは $+\infty$ というこことにする。

これの最小値を λ, 最小値をとる j の個数を m とするとき

λ のことを k, h から定まる実対数閾値, m をその多重度とよぶ。

例. $k=(2,3,4)$, $h=(1,2,6)$ なら

$$\left(\frac{h_j+1}{2k_j}\right)=\left(\frac{1+1}{4},\frac{2+1}{6},\frac{6+1}{8}\right)=\left(\frac{1}{2},\frac{1}{2},\frac{7}{8}\right)$$

なので $\lambda=\frac{1}{2}$, $m=2$ である。

以降, 座標 u の成分を並べかえて, 上の最小値をとる j を最小に集め、

そうでない j を後ろにまわし, $u=(u_a, u_b)$ とかく。

メリン変換とメリン逆変換を以下のように定義する。

$$f(t) \underset{\text{メリン逆変換}}{\overset{\text{メリン変換}}{\rightleftarrows}} F(z) \qquad (z\in\mathbb{C})$$

下の a,b とは違う。

$$F(z)=(Mf)(z)=\int_0^{\infty}t^z f(t)\,dt \qquad \left(\begin{aligned}t^z&=e^{z\log t}\\&=e^{a\log t}(\cos b+i\sin b)^{\log t}\end{aligned}\right)$$

$$f(t)=(M^{-1}F)(t)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}F(z)\,t^{-z}\,dz$$ (但し複素平面上の縦軸に沿って積分)

一般に $F(z)$ は $a<\mathrm{Re}(z)<b$ で解析関数となるような定数 a,b をもつ

(ので, 逆変換の定数 c は $a<c<b$ であるものをとる)。

メリン変換は、次の①②が等価であるという性質をもつ。

① $t\to 0$ のとき $f(t)\cong t^{\lambda-1}(-\log t)^{m-1}$

② $(Mf)(z)\cong\dfrac{1}{(z+\lambda)^m}$

No.

Date

補題22

$\lambda>0$ を実数, $m>0$ を自然数とする。

$$f_m(t)=\begin{cases}t^{\lambda-1}(-\log t)^{m-1} & (0<t<1)\\ 0 & (\text{それ以外})\end{cases}$$ のメリン変換は

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

$$\begin{aligned}
(Mf_m)(z)&=\int_0^1 t^z t^{\lambda-1}(-\log t)^{m-1}\,dt\\
&=\left[\frac{t^{z+\lambda}}{z+\lambda}(-\log t)^{m-1}\right]_0^1+\frac{m-1}{z+\lambda}\int_0^1 t^{z+\lambda}\frac{(-\log t)^{m-2}}{t}\,dt\\
&=\frac{(m-1)}{z+\lambda}(Mf_{m-1})(z)\\
&=\frac{(m-1)}{(z+\lambda)}\frac{(m-2)}{(z+\lambda)}(Mf_{m-2})(z)
\end{aligned}$$

のようにどんどん m を減らせる。

$$(Mf_1)(z)=\int_0^1 t^{z+\lambda-1}\,dt=\frac{1}{(z+\lambda)}$$ なので,

$$(Mf_m)(z)=\frac{(m-1)}{(z+\lambda)}\cdot\frac{(m-2)}{(z+\lambda)}\cdots\cdot\frac{2}{(z+\lambda)}\cdot\frac{1}{(z+\lambda)}\cdot\frac{1}{(z+\lambda)} \qquad \square$$