§12. コントラスト関数

ここでは, コントラスト関数というものを用いて, 統計多様体を構成できることについて述べよう.
M を C^∞ 級多様体とし, $\rho \in C^\infty(M \times M)$ とする.
q を固定しておくと, $\rho(p,q)$ は p を変数とする M 上の C^∞ 級関数とみなすことができる. よって, $X \in \mathfrak{X}(M)$ に対して p の関数 $\rho(p,q)$ を X で微分することができる. これを

$$X_p\rho(p,q)$$

と表す. 更に, $p=q$ とすると, $X_p\rho(p,q)$ は p を変数とする M 上の C^∞ 級関数となる. これを $\rho(X|\,\cdot\,)$ と表す. 同様に, $\rho(p,q)$ を q を変数とする M 上の C^∞ 級関数とみなすことにより, M 上の C^∞ 級関数 $\rho(\,\cdot\,|X)$ を定めることができる.
更に, $X_1,\dots,X_l,Y_1,\dots,Y_m \in \mathfrak{X}(M)$ とすると, M 上の C^∞ 級関数 $\rho(X_1\dots X_l|Y_1\dots Y_m)$ を

$$\rho(X_1\cdots X_l|Y_1\cdots Y_m)(r) = (X_1\cdots X_l)_p(Y_1\cdots Y_m)_q\rho(p,q)|_{p=r,q=r} \quad (r \in M)$$

により定めることができる.
これらの記号を用いて, コントラスト関数を次のように定める.

定義 ρ は次の (1)〜(3) をみたすとき, M 上のコントラスト関数という.

(1) 任意の $p \in M$ に対して

$$\rho(p,p) = 0.$$

(2) 任意の $X \in \mathfrak{X}(M)$ に対して

$$\rho(X|\,\cdot\,) = \rho(\,\cdot\,|X) = 0.$$

(3) $X, Y \in \mathfrak{X}(M)$ に対して

$$g(X,Y) = -\rho(X|Y)$$

とおくと, g は M の非退化計量となる.

例 M を C^∞ 級多様体とし, $\rho \in C^\infty(M\times M)$ とする.
ここで, 任意の $p, q \in M$ に対して

$$\rho(p,q) \geq 0$$

がなりたち, $\rho(p,q)=0$ となるのは $p=q$ のときに限ると仮定しよう.
まず, 仮定より, ρ が上の定義の (1) をみたすことは明らかである.
また, q を固定しておき, $\rho(p,q)$ を p を変数とする関数とみなすと, 仮定より, $\rho(p,q)$ は q において最小値 0 をとる.
よって, 任意の $X \in \mathfrak{X}(M)$ に対して

$$\rho(X|\,\cdot\,) = 0 \qquad (*)$$

がなりたつ.
同様に,

$$\rho(\,\cdot\,|X) = 0$$

もなりたつから, ρ は上の定義の (2) をみたす.
次に, $Y \in \mathfrak{X}(M)$ とすると, $(*)$ より,

$$\begin{aligned} 0 &= Y\rho(X|\,\cdot\,) \\ &= \rho(YX|\,\cdot\,) + \rho(X|Y). \end{aligned}$$

すなわち,

$$g(X,Y)=\rho(YX|\,\cdot\,).$$

同様に,

$$g(Y,X)=\rho(XY|\,\cdot\,).$$

$[X,Y]\in\mathfrak{X}(M)$ に注意すると, 再び $(*)$ より,

$$\begin{aligned}g(Y,X)-g(X,Y)&=\rho(XY-YX|\,\cdot\,)\\&=\rho([X,Y]|\,\cdot\,)\\&=0.\end{aligned}$$

よって, g は M 上の $(0,2)$ 型対称テンソル場となり,

$$g(X,Y)=\rho(XY|\,\cdot\,)=\rho(\,\cdot\,|XY)$$

がなりたつ. 特に, ρ に対する仮定を再び用いると, g は半正定値である.
したがって, g が正定値ならば, 上の定義の (3) がなりたち, ρ は M 上のコントラスト関数となる.

コントラスト関数を用いて, アファイン接続を定めよう.
M を C^∞ 級多様体, ρ を M 上のコントラスト関数とし, $X,Y\in\mathfrak{X}(M)$ とする.
上の定義の (3) により定まる M の非退化計量 g を用いて, 任意の $Z\in\mathfrak{X}(M)$ に対して

$$g(\nabla_XY,Z)=-\rho(XY|Z),\ g(Z,\nabla^*_XY)=-\rho(Z|XY)$$

がなりたつように, $\nabla_XY,\nabla^*_XY\in\mathfrak{X}(M)$ を定めよう. このとき, ∇,∇^* は M のアファイン接続となる.

定理　∇^* は ∇ の双対接続で, (M,∇,g) は統計多様体.

証明　$X,Y,Z\in\mathfrak{X}(M)$ とする.
まず, g および ∇,∇^* の定義より,

$$\begin{aligned}Xg(Y,Z)&=-X\rho(Y|Z)\\&=-\rho(XY|Z)-\rho(Y|XZ)\\&=g(\nabla_XY,Z)+g(Y,\nabla^*_XZ).\end{aligned}$$

よって, ∇^* は ∇ の双対接続.
次に, T を ∇ の捩率とすると,

$$\begin{aligned}g(T(X,Y),Z)&=g(\nabla_XY-\nabla_YX-[X,Y],Z)\\&=g(\nabla_XY,Z)-g(\nabla_YX,Z)-g([X,Y],Z)\\&=-\rho(XY|Z)+\rho(YX|Z)+\rho(XY-YX|Z)\\&=0.\end{aligned}$$

よって,

$$T=0.$$

同様に, ∇^* も捩れをもたない.
したがって, (M,∇,g) は統計多様体. □

統計的モデルに対する Fisher 計量と α 接続は α ダイバージェンスというコントラスト関数を用いて表すことができる.

例 (α ダイバージェンス)
Ω を高々可算集合または $\mathbf{R}^k$ とし,

$$S = \{p(x;\xi) | \xi \in \Xi\}$$

を Ω 上の n 次元統計的モデルとする.
$\alpha \in \mathbf{R}$ に対して, $t > 0$ の範囲で定義された関数 $\varphi^{(\alpha)} = \varphi^{(\alpha)}(t)$ を

$$\varphi^{(\alpha)}(t) = \begin{cases} \dfrac{4}{1-\alpha^2}\left(1 - t^{\frac{1+\alpha}{2}}\right) & (\alpha \neq \pm 1), \\ t \log t & (\alpha = 1), \\ -\log t & (\alpha = -1) \end{cases}$$

により定める.
このとき, $p, q \in S$ に対して

$$\rho(p,q) = \int_\Omega p \varphi^{(\alpha)}\left(\frac{q}{p}\right) dx$$

とおく.
$\varphi^{(\alpha)}$ は凸関数であることが分かるから, §7 においても扱った Jensen の不等式より,

$$\begin{aligned} \rho(p,q) &\geq \varphi^{(\alpha)}\left(\int_\Omega p\frac{q}{p} dx\right) \\ &= \varphi^{(\alpha)}(1) \\ &= 0. \end{aligned}$$

また, $\rho(p,q) = 0$ となるのは $p = q$ のときに限ることも分かる.
よって, ρ は上の例の仮定をみたす. この $\rho(p,q)$ を $D^{(\alpha)}(p\|q)$ と表し, $D^{(\alpha)}$ を α ダイバージェンスという. $\alpha = 0$ のときは Hellinger 距離ともいう. また, $\alpha = \pm 1$ のときは相対エントロピーまたは Kullback-Leibler ダイバージェンスともいう.

定理 S の Fisher 計量, α 接続はそれぞれ $D^{(\alpha)}$ の定める非退化計量, アファイン接続.

証明 $\alpha \neq \pm 1$ の場合に Fisher 計量についてのみ示す.
$i, j = 1, 2, \ldots, n$ とし, $p, q, r \in S$ とすると,

$$D^{(\alpha)}(p\|q) = \frac{4}{1-\alpha^2}\int_\Omega \left(p - p^{\frac{1-\alpha}{2}} q^{\frac{1+\alpha}{2}}\right) dx$$

となるから,

$$\begin{aligned} D^{(\alpha)}(\partial_i \| \partial_j)(r) &= \frac{4}{1-\alpha^2}\int_\Omega \left(-\frac{1-\alpha}{2}p^{-\frac{1+\alpha}{2}}\partial_i p \cdot \frac{1+\alpha}{2} q^{-\frac{1-\alpha}{2}}\partial_j q\right) dx \bigg|_{p=r,q=r} \\ &= -\int_\Omega (\partial_i \log r)(\partial_j \log r) r dx. \end{aligned}$$

□

なお, 逆に任意の統計多様体に対して, それに対応するコントラスト関数が存在することが知られている.

関連事項 12. Karamata の不等式

φ を $\mathbf{R}$ 上の凸関数とし, $x_1, x_2, \ldots, x_n \in \mathbf{R}$ とすると, 不等式

$$\frac{\varphi(x_1)+\varphi(x_2)+\cdots+\varphi(x_n)}{n} \geq \varphi\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)$$

がなりたつ. これは Jensen の不等式の 1 つの例である. 実際, 確率空間 $(\Omega, \mathcal{F}, P)$ を

$$\Omega=\{1,2,\ldots,n\},\ \mathcal{F}=2^{\Omega},\ P(\{i\})=\frac{1}{n}\ (i=1,2,\ldots,n)$$

により定め,

$$X(i)=x_i \quad (i=1,2,\ldots,n)$$

により定まる $(\Omega, \mathcal{F}, P)$ 上の確率変数 X を考えればよい.
上の不等式を一般化したものとして, Karamata の不等式というものが知られている.
φ を $\mathbf{R}$ 上の凸関数とし, $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in \mathbf{R}$ とする. 更に,

$$x_1+x_2+\cdots+x_n=y_1+y_2+\cdots+y_n$$

がなりたち, 必要ならば番号を付け替えることにより,

$$x_1 \geq x_2 \geq \cdots \geq x_n,\ y_1 \geq y_2 \geq \cdots \geq y_n$$

としたとき, 任意の $i=1,2,\ldots,n-1$ に対して

$$x_1+x_2+\cdots+x_i \geq y_1+y_2+\cdots+y_i$$

がなりたつと仮定する.
このとき, 不等式

$$\varphi(x_1)+\varphi(x_2)+\cdots+\varphi(x_n) \geq \varphi(y_1)+\varphi(y_2)+\cdots+\varphi(y_n)$$

がなりたつ. これが Karamata の不等式である. 特に, $y_1, y_2, \ldots, y_n$ がすべて $x_1, x_2, \ldots, x_n$ の平均の場合が最初に述べた形の Jensen の不等式である.
Karamata の不等式を証明するには, 任意の $i=1,2,\ldots,n$ に対して $x_i \neq y_i$ としてよい. このとき,

$$c_i=\frac{\varphi(x_i)-\varphi(y_i)}{x_i-y_i},\ S_k=\sum_{i=1}^{k} x_i,\ T_k=\sum_{i=1}^{k} y_i,\ S_0=T_0=0$$

とおくと,

$$\begin{aligned}
\sum_{i=1}^{n}\varphi(x_i)-\sum_{i=1}^{n}\varphi(y_i) &= \sum_{i=1}^{n} c_i(x_i-y_i) \\
&= \sum_{i=1}^{n} c_i(S_i-S_{i-1}-T_i+T_{i-1}) \\
&= c_n(S_n-T_n)+\sum_{i=0}^{n-1}(c_i-c_{i+1})(S_i-T_i)
\end{aligned}$$

となる. 仮定より, 最後の式の第 1 項は 0 で, 第 2 項は φ の凸性と仮定より, 0 以上である.