§8. 単調性と不変性

あたえられた統計的モデルから別の統計的モデルへの変換を考えると, Fisher 計量は単調性という性質をもち, その不変性は変換が十分統計量というものであるときになりたつことが分かる.
$(\Omega, \mathcal{F}, P)$ を確率空間, X を $(\Omega, \mathcal{F}, P)$ 上の P 可積分な確率変数とする. Y も $(\Omega, \mathcal{F}, P)$ 上の確率変数とし, Y の生成する σ 加法族を $\sigma(Y)$ と表すことにしよう. すなわち, $\sigma(Y)$ は Y を可測にする最小の σ 加法族で,

$$\sigma(Y) = \{Y^{-1}(B) | B \in \mathcal{B}(\mathbf{R})\}$$

である. このとき, §7 において扱ったことより, $\sigma(Y)$ に関する X の条件付き期待値 $E[X|\sigma(Y)]$ が定まる.
一方, $B \in \mathcal{B}(\mathbf{R})$ に対して

$$\nu(B) = \int_{Y^{-1}(B)} X(\omega)P(d\omega)$$

とおくと, ν は Y の分布 μ_Y に関して絶対連続な $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の有限符号付き測度となる.
よって, Radon-Nikodym の定理より, $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mu_Y)$ 上の μ_Y 可積分な確率変数 f が存在し, 任意の $B \in \mathcal{B}(\mathbf{R})$ に対して

$$\nu(B) = \int_B f(y)\mu_Y(dy)$$

がなりたつ. この f を $E[X|Y = y]$ と表し, $Y = y$ の下での X の条件付き平均値または条件付き期待値という. 部分 σ 加法族に関する条件付き期待値の場合と同様に, 確率 1 で等しい確率変数を同一視すれば, 上の条件付き期待値の存在は一意的である.
更に, $A \in \mathcal{F}$ のとき, $E[\chi_A|Y = y]$ を $P(A|Y = y)$ と表し, $Y = y$ の下での A の条件付き確率という.
次に示すように, 上の 2 つの条件付き期待値は本質的には同じものである.

定理　$(\Omega, \mathcal{F}, P)$ を確率空間, X, Y を $(\Omega, \mathcal{F}, P)$ 上の確率変数とする. X が P 可積分ならば,

$$E[X|\sigma(Y)](\omega) = E[X|Y = Y(\omega)] \quad \text{a.s. } \omega.$$

証明　$B \in \mathcal{B}(\mathbf{R})$ とすると, 変数変換公式と条件付き期待値の定義より,

$$\begin{aligned}
\int_{Y^{-1}(B)} E[X|Y = Y(\omega)]P(d\omega) &= \int_\Omega \chi_{Y^{-1}(B)}(\omega)E[X|Y = Y(\omega)]P(d\omega) \\
&= \int_\Omega \chi_B(Y(\omega))E[X|Y = Y(\omega)]P(d\omega) \\
&= \int_{\mathbf{R}} \chi_B(y)E[X|Y = y]\mu_Y(dy) \\
&= \int_B E[X|Y = y]\mu_Y(dy) \\
&= \int_{Y^{-1}(B)} X(\omega)P(d\omega) \\
&= \int_{Y^{-1}(B)} E[X|\sigma(Y)](\omega)P(d\omega).
\end{aligned}$$

よって,

$$E[X|\sigma(Y)](\omega) = E[X|Y = Y(\omega)] \quad \text{a.s. } \omega.$$

□

さて, Ω を高々可算集合または $\mathbf{R}^k$ とし, $\mathcal{F}$ を Ω が高々可算集合のときは Ω の部分集合全体で, Ω が $\mathbf{R}^k$ のときは $\mathbf{R}^k$ の Borel 集合族であるとする. $\Omega', \mathcal{F}'$ についても同様に定める.
まず,

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

を Ω 上の n 次元統計的モデルとし, $p(x;\xi)$ が可測空間 $(\Omega, \mathcal{F})$ に値をとる確率変数 X_ξ に対する確率関数または密度関数であるとする.
次に, F を $(\Omega, \mathcal{F})$ からもう 1 つの可測空間 $(\Omega', \mathcal{F}')$ への可測写像とする. このとき, $(\Omega', \mathcal{F}')$ に値をとる確率変数 $F \circ X_\xi$ に対する確率関数または密度関数 $q(y;\xi)$ が定まるから, Ω' 上の n 次元統計的モデル S_F を

$$S_F = \{q(y;\xi) | \xi \in \Xi\}$$

により定めることができる. 更に, 関数 $r(x;\xi)$ を

$$r(x;\xi) = \frac{p(x;\xi)}{q(F(x);\xi)}$$

により定める. $r(x;\xi)$ が ξ に依存しないとき, F を S に関する十分統計量という.
確率空間 $(\Omega, \mathcal{F}, \mu_{X_\xi})$ を考え, $A \in \mathcal{F}$, $y \in \Omega'$ とする. このとき, $F = y$ の下での A の条件付き確率 $P(A|F=y;\xi)$ は

$$P(A|F=y;\xi) = \int_A r(x;\xi)\delta(y - F(x))dx \quad \text{a.s. } y$$

と表すことができる. ただし, δ は Dirac の δ 関数である. 実際, $B \in \mathcal{F}'$ とすると, Fubini の定理と δ の定義より,

$$\begin{aligned}
\int_B \left(\int_A r(x;\xi)\delta(y - F(x))dx \right) \mu_{F\circ X_\xi}(dy) &= \int_B \int_A r(x;\xi)\delta(y - F(x))q(y;\xi)dxdy \\
&= \int_A r(x;\xi) \int_B \delta(y - F(x))q(y;\xi)dydx \\
&= \int_{A\cap F^{-1}(B)} r(x;\xi)q(F(x);\xi)dx \\
&= \int_{A\cap F^{-1}(B)} p(x;\xi)dx \\
&= \int_{F^{-1}(B)} \chi_A p(x;\xi)dx \\
&= \int_{F^{-1}(B)} \chi_A \mu_{X_\xi}(dx)
\end{aligned}$$

となるからである. よって, $P(A|F=y;\xi)$ が ξ に依存しないときに F が十分統計量であるということもできる.
更に, $G(\xi) = (g_{ij}(\xi))$ および $G_F(\xi) = (g^F_{ij}(\xi))$ をそれぞれ S, S_F の Fisher 情報行列とし,

$$\Delta G(\xi) = (\Delta g_{ij}(\xi)) = G(\xi) - G_F(\xi)$$

とおく. Fisher 計量の単調性と不変性は次のように述べることができる.

定理　$\Delta g_{ij}(\xi)$ は

$$\Delta g_{ij}(\xi) = \int_\Omega \partial_i \log r(x;\xi) \partial_j \log r(x;\xi) p(x;\xi)dx$$

をみたす. 特に, ΔG は半正定値で, ΔG が零行列となるのは F が S に関する十分統計量のとき.

証明 Fisher 情報行列および $r(x;\xi)$ の定義と変数変換公式より,

$$
\begin{aligned}
\Delta g_{ij}(\xi) &= \int_\Omega \partial_i \log p(x;\xi)\partial_j \log p(x;\xi)p(x;\xi)dx - \int_{\Omega'} \partial_i \log q(y;\xi)\partial_j \log q(y;\xi)q(y;\xi)dy \\
&= \int_\Omega \left(\partial_i \log r(x;\xi) + \partial_i \log q(F(x);\xi)\right)\left(\partial_j \log r(x;\xi) + \partial_j \log q(F(x);\xi)\right)p(x;\xi)dx \\
&\quad - \int_\Omega \partial_i \log q(F(x);\xi)\partial_j \log q(F(x);\xi)p(x;\xi)dx \\
&= \int_\Omega \partial_i \log r(x;\xi)\partial_j \log r(x;\xi)p(x;\xi)dx + \int_\Omega \partial_i \log r(x;\xi)\partial_j \log q(F(x);\xi)p(x;\xi)dx \\
&\quad + \int_\Omega \partial_j \log r(x;\xi)\partial_i \log q(F(x);\xi)p(x;\xi)dx.
\end{aligned}
$$

ここで,

$$\int_{\Omega'} q(y;\xi)dy = 1$$

に注意すると,

$$
\begin{aligned}
\int_\Omega \partial_i \log r(x;\xi)\partial_j \log q(F(x);\xi)p(x;\xi)dx &= \int_\Omega \partial_i r(x;\xi)\partial_j q(F(x);\xi)dx \\
&= \partial_i \int_\Omega \frac{p(x;\xi)}{q(F(x);\xi)}\partial_j q(F(x);\xi)dx \\
&\quad - \int_\Omega \frac{p(x;\xi)}{q(F(x);\xi)}\partial_i\partial_j q(F(x);\xi)dx \\
&= \partial_i \int_{\Omega'} \partial_j q(y;\xi)dy - \int_{\Omega'} \partial_i\partial_j q(y;\xi)dy \\
&= 0.
\end{aligned}
$$

同様に,

$$\int_\Omega \partial_j \log r(x;\xi)\partial_i \log q(F(x);\xi)p(x;\xi)dx = 0.$$

よって,

$$\Delta g_{ij}(\xi) = \int_\Omega \partial_i \log r(x;\xi)\partial_j \log r(x;\xi)p(x;\xi)dx.$$

ここで,

$$(\partial_i \log r(x;\xi)\partial_j \log r(x;\xi))$$

は半正定値だから, $(\Delta g_{ij}(\xi))$ も半正定値.
また, $\Delta g_{ij}(\xi) = 0$ となるのは, 任意の $i = 1, 2, \ldots, n$ に対して

$$\partial_i \log r(x;\xi) = 0$$

となるとき, すなわち $r(x;\xi)$ が ξ に依存しないときで, このとき F は十分統計量. □

関連事項 8. マルチンゲール

確率論における基本的な概念として, マルチンゲールというものが挙げられる.
確率変数が時間とともに変化していくと, 確率過程というものが得られる. 簡単のため, 時間 t は連続的に変化し, $t \geq 0$ としておき, 1 次元, すなわち $\mathbf{R}$ に値をとる確率過程を考えよう.
$\{X_t\}_{t\geq 0}$ を確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率過程とする. 更に, $\mathcal{F}$ の部分 σ 加法族 $\mathcal{F}_t$ で, $s < t$ をみたす任意の $s, t \geq 0$ に対して

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$$

となるものがあたえられているとしよう. $\{\mathcal{F}_t\}_{t\geq 0}$ をフィルトレーションまたは増加情報系という.
$\{X_t\}_{t\geq 0}$ は次の (1)〜(3) をみたすとき, $\{\mathcal{F}_t\}_{t\geq 0}$ に関するマルチンゲールという.

(1) 任意の $t \geq 0$ に対して, $E[|X_t|] < +\infty$.
(2) 任意の $t \geq 0$ に対して, X_t は $\mathcal{F}_t$ 可測.
(3) $s \leq t$ をみたす任意の $s, t \geq 0$ に対して, $E[X_t|\mathcal{F}_s] = X_s$.

上の定義では (3) が本質的である. これは時刻 s までの情報 $\mathcal{F}_s$ があたえられたとして, それより先の時刻 t における X_t の期待値を求めると, それは時刻 s における値 X_s に等しいということを意味する. よって, マルチンゲールは公平な賭けを確率過程を用いて表現したものということができる. なお, ここでは詳しくは述べないが, マルチンゲールの定義においては $\{X_t\}_{t\geq 0}$ や $\{\mathcal{F}_t\}_{t\geq 0}$ の右連続性を仮定することが多い.
上の (3) の条件を

$$E[X_t|\mathcal{F}_s] \geq X_s$$

と置き替えて定められる $\{X_t\}_{t\geq 0}$ を劣マルチンゲールという. $\{X_t\}_{t\geq 0}$ がマルチンゲールならば, $\{|X_t|\}_{t\geq 0}$ は劣マルチンゲールである. 実際, §7 において扱った Jensen の不等式より,

$$\begin{aligned} |X_s| &= |E[X_t|\mathcal{F}_s]| \\ &\leq E[|X_t||\mathcal{F}_s] \end{aligned}$$

となるからである.
マルチンゲールの重要な例として, Brown 運動が挙げられる. 簡単のため, 1 次元の場合について述べよう. 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率過程 $\{B_t\}_{t\geq 0}$ は次の (1)〜(3) をみたすとき, Brown 運動という.

(1) $B_0 = 0$ a.s.
(2) $\{B_t\}_{t\geq 0}$ は連続, すなわち任意の $\omega \in \Omega$ に対して, $B_t(\omega)$ は t について連続.
(3) $0 = t_0 < t_1 < \cdots < t_n$ とすると, $\{B_{t_i} - B_{t_{i-1}}\}_{1\leq i\leq n}$ は独立で, それぞれ平均値 0, 分散 $t_i - t_{i-1}$ の正規分布に従う, すなわち

$$p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \quad (x \in \mathbf{R},\ t > 0)$$

とおくと, 任意の $A_1, A_2, \ldots, A_n \in \mathcal{B}(\mathbf{R})$ に対して

$$P(\{\omega \in \Omega | B_{t_i}(\omega) - B_{t_{i-1}}(\omega) \in A_i,\ 1 \leq i \leq n\}) = \prod_{i=1}^{n} \int_{A_i} p(x_i, t_i - t_{i-1}) dx_i.$$