

格子点上カイ 2 乗確率場の最大値分布の近似と その連鎖解析への応用

栗木 哲 (統計数理研, 総合研究大学院大)

1. 主要な結果

$T \subset \mathbb{R}^d$ を \mathbb{R} の区間の直積とし, $Z_k(t)$, $t = (t_i)_{1 \leq i \leq d} \in T$ を各 $k = 1, \dots, m$ について独立な, 平均 0, 分散 1, 共分散関数

$$\text{Cov}(Z_k(t), Z_k(\tilde{t})) = 1 - \sum_{i=1}^d \rho_{ki} |t_i - \tilde{t}_i| + o(\|t - \tilde{t}\|) \quad (\rho_{ki} > 0)$$

を持つ実数値正規確率場とする. この共分散構造は局所的に OU 過程の共分散の直積型である. さらに $Q = (q_{kl})$ をランク r の $m \times m$ 直交射影行列とし

$$Y(t) = \sqrt{\sum_{k,l=1}^m q_{kl} Z_k(t) Z_l(t)}$$

とおく. $Y(t)^2$, $t \in T$ は各 t について自由度 r のカイ 2 乗分布に従うカイ 2 乗確率場となる. $j = (j_i)_{1 \leq i \leq d} \in \mathbb{Z}^d$, $\Delta = (\Delta_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ に対し $j\Delta = (j_i \Delta_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ とおく. また $J = \{j \in \mathbb{Z}^d \mid j\Delta \in T\}$ とおく. 本報告は, 格子点で定義されたカイ 2 乗確率場 $Y(j\Delta)^2$, $j \in J$ の最大値の分布に関するものである.

$h = (h_1, \dots, h_m)'$ を部分空間に制限された $r - 1$ 次元単位球面

$$\{h \in \mathbb{R}^m \mid \|h\| = 1, h = Qh\} \simeq S^{r-1}$$

上の一様分布ベクトルとする. T のルベーグ測度を $|T|$ とおく.

定理. $b \rightarrow \infty$, $\Delta_i \rightarrow 0$, s.t. $b\sqrt{\Delta_i} \rightarrow c_i \in (0, \infty)$ ($i = 1, \dots, d$) のとき

$$\begin{aligned} P\left(\max_{j \in J} Y(j\Delta) > b\right) &\sim C \times \frac{2|T|}{2^{r/2} \Gamma(r/2)} b^{r+2d-2} e^{-b^2/2} \\ &\sim C \times |T| \frac{\Gamma(r/2 + d)}{2^d \Gamma(r/2)} P(\chi_{r+2d} > b), \end{aligned} \quad (1)$$

ただし χ_{r+2d} は自由度 $r + 2d$ のカイ変数,

$$C = E\left[\prod_{i=1}^d \bar{\rho}_i(h) \nu(c_i \sqrt{2\bar{\rho}_i(h)})\right], \quad \bar{\rho}_i(h) = \sum_{k=1}^r h_k^2 \rho_{ki},$$

また $\Phi(\cdot)$ を標準正規分布の分布関数とするとき

$$\nu(x) = \begin{cases} 2x^{-2} \exp\left\{-2 \sum_{n=1}^{\infty} n^{-1} \Phi\left(-\frac{1}{2}x\sqrt{n}\right)\right\} & (x > 0) \\ 1 & (x = 0). \end{cases}$$

2. 証明の考え方

簡単のため, $Q = I_m$ の場合について説明する. 最初に添字 t^0 を固定し, その近傍 $t \approx t^0$ における $Y(t)$ の挙動を考察する. 定常性から一般性を失うことなく $t^0 = 0$ とおく. 以下では $Z_k(0) = \xi_k$ ($k = 1, \dots, m$) を与えた下での条件付き分布を考える.

Kim and Siegmund (1989, *Biometrika*) と同じ設定で, 次を示すことができる:
 $\|\xi\| \rightarrow \infty, b \rightarrow \infty, \Delta_i \rightarrow 0$, s.t. $\|\xi\| \sim b, \|\xi\| \sqrt{\Delta_i} \rightarrow c_i \in (0, \infty)$ ($i = 1, \dots, d$) のとき

$$b\{Y(j\Delta) - \|\xi\|\} \Big|_{Z_k(0)=\xi_k, \forall k} \Rightarrow \sum_{i=1}^d S_{ji}^{(i)} \quad (\text{有限個の } j \text{ の同時分布として})$$

ただし各 $i = 1, \dots, d$ について $X_t^{(i)}$ ($t \in \mathbb{Z}$) を $N(-\bar{\rho}_i c_i^2, 2\bar{\rho}_i c_i^2)$, $\bar{\rho}_i = \sum_{k=1}^m \xi_k^2 \rho_{ki} / \|\xi\|^2$ に従う独立な確率変数列とし,

$$S_t^{(i)} = \begin{cases} X_1^{(i)} + \dots + X_t^{(i)} & (t > 0) \\ 0 & (t = 0) \\ X_{-1}^{(i)} + \dots + X_{-|t|}^{(i)} & (t < 0) \end{cases}$$

とおいた. このことから, 最大値の分布 (1) は漸近的には独立な確率変数の和の最大値の分布で記述され, その導出において再生理論で知られた関係式 (Siegmund, 1992, *Probability Theory* など) を用いることができる.

3. 問題の背景

種を分ける遺伝的しくみである生殖的隔離は, 遺伝子座間の相互作用によってもたらされると考えられている. マーカー遺伝子座の総当たりを行い 2 遺伝子座間の「分離のゆがみ」を適当なカイ 2 乗検定統計量によってとらえることによって, 生殖的隔離を引き起こす遺伝子ペアを検出できる. しかしながらマーカー間隔が狭く, マーカー数が大きい場合は, 検定の総数が組合せ的に膨大となるため, 多重比較, すなわち p 値の多重性調整が必要となる. 連鎖が引き起こす確率構造を同定すると, 検定統計量の同時分布はカイ 2 乗確率場ととらえることができ, (1) によって多重性調整 p 値の近似値を得ることが可能となる [春島, 倉田 (国立遺伝学研), 藤澤, 栗木 (統計数理研) による情報・システム研究機構新領域融合プロジェクト研究].