12.3 Degenerate U-Statistics

U 統計量の列の漸近正規性が Theorem 12.3 で示され，その漸近分散が $r^2\zeta_1$ で与えられることが分かった．この漸近分散がゼロに縮退するとき，U 統計量の列（あるいはそのカーネル関数）は *degenerate* であるという．こうなると，当然漸近的な帰無分布を検定に用いるといったことができず困ってしまう．$n^{c/2}$ の c を適切に（大きめに）選んで，収束先に広がりを持たせようというのがこの節の目的である．

まず，Theorem 12.3 の証明の中で U 統計量の分散は

$$\mathrm{var}U = \sum_{d=0}^{r} \frac{r!^2}{d!(r-d)!^2} \frac{(n-r)(n-r-1)\cdots(n-2r+d+1)}{n(n-1)\cdots(n-r+1)}\zeta_d$$

と与えられた．n に依存している部分に着目すると，もし，$\zeta_1 = \cdots = \zeta_{c-1} = 0 < \zeta_c$ とすれば，

$$\begin{aligned}\mathrm{var}U &= \sum_{d=c}^{r} \frac{r!^2}{d!(r-d)!^2} \frac{(n-r)(n-r-1)\cdots(n-2r+d+1)}{n(n-1)\cdots(n-r+1)}\zeta_d \\ &= O(n^{-c})\end{aligned}$$

となることが分かる．$n^{c/2}(U_n - \theta)$ の極限分布を求める．

まず，11.4 節で議論したヘフディング分解を考える（10/13 の資料参照）．観測 $X_1, ..., X_n$ に基づき，r 次のカーネル h で定義される U 統計量 U_n はヘフディング分解により，

$$U_n = \sum_{c=0}^{r} \sum_{|A|=c} \frac{1}{\binom{n}{r}} \sum_{\beta} P_A h(X_{\beta_1}, ..., X_{\beta_r}) = \sum_{c=0}^{r} \binom{r}{c} U_{n,c}$$

と分解できる．ただし，$U_{n,c}$ は $0 \le c \le r$ として，カーネル

$$h_c(X_1, ..., X_c) = P_{\{1,...,c\}} h(X_1, ..., X_r)$$

で定義される c 次元の U 統計量．

ここで，P_A は $A \subset \{1, ..., n\}$ として

$$H_A := \{g_A(X_i : i \in A) : E[g_A^2(X_i : i \in A)] < \infty, E[g_A(X_i : i \in A)|X_j : j \in B] = 0 (\forall B : |B| < |A|)\}$$

と定義される互いに直交する空間への射影であった．

２つ目の等号を説明する．右辺の和の中身は，

$$\binom{r}{c} U_{n,c} = \binom{r}{c} \frac{1}{\binom{n}{c}} \sum_{\beta} P_{\{1,...,c\}} h(X_{\beta_1}, ..., X_{\beta_c})$$

と書ける．これが，$\sum_{|A|=c} \frac{1}{\binom{n}{r}} \sum_{\beta} P_A h(X_{\beta_1}, ..., X_{\beta_r})$ と一致することが言えればよい．c 個の要素を持つ A を fix する．集合 B が A を含まない，つまり $A \not\subset B$ であれば，H_A は $g(X_j : j \in B)$ で表される全ての関数（すなわち，$\sum_{C \subset B} H_C$）に直交する．なぜなら，

$$\begin{aligned}E[g_A(X_i : i \in A)g(X_j : j \in B)] &= E[E[g_A(X_i : i \in A)g(X_j : j \in B)|X_j : j \in A \cap B]] \\ &= E[g(X_j : j \in B)E[g_A(X_i : i \in A)|X_k : k \in A \cap B]] \\ &= E[g(X_j : j \in B)0] \quad (H_A \text{ の定義より}) \\ &= 0.\end{aligned}$$

２つ目の等号は $X_1, ..., X_n$ が独立であることを用いた．少々分かりづらいので具体的に $n = 4$, すなわち全体の添字集合は $\{1, 2, 3, 4\}$ で $A = \{1, 2\}, B = \{1, 3\}$ とした場合，

$$\begin{aligned}E[g_{\{1,2\}}(X_1, X_2)g(X_1, X_3)] &= E[E[g_{\{1,2\}}(X_1, X_2)g(X_1, X_3)|X_1]] \\ &= E[g(X_1, X_3)E[g_{\{1,2\}}(X_1, X_2)|X_1]] \\ &= E[g(X_1, X_3)0] = 0\end{aligned}$$

である．

よって射影 $P_A h(X_{\beta_1},...,X_{\beta_r})$ は $A \subset \beta = \{\beta_1,...\beta_r\}$ でないときゼロとなる．$A \subset \beta$ となるような β については射影 $P_A h(X_{\beta_1},...,X_{\beta_r})$ は β によらない（すなわち $\beta - A$ の $r-c$ 個の要素によらない）で，$(X_j : j \in A)$ の固定された関数 h_c である．fix された集合 A を包む β は $\binom{n-c}{r-c}$ 通りある．$\binom{n-c}{r-c}/\binom{r}{c}\binom{n}{r} = 1/\binom{n}{c}$ を用いれば，

$$\begin{aligned}\sum_{|A|=c} \frac{1}{\binom{n}{r}} \sum_{\beta} P_A h(X_{\beta_1},...,X_{\beta_r}) &= \binom{n}{c} \frac{1}{\binom{n}{r}} \binom{n-c}{r-c} h_c(X_{\beta_1},...,X_{\beta_r}) \\ &= \binom{r}{c} h_c(X_{\beta_1},...,X_{\beta_r})\end{aligned}$$

となり等号が成り立つことが確認できた．

空間 $H_{\{1,...,c\}}$ の性質から，カーネル h_c は $c \geq 2$ に対して degenerate である．これは

$$\begin{aligned}\zeta_1 &= cov(h_c(X_1,X_2,...,X_c), h_c(X_1,X_2',...,X_c')) \\ &= E[h_c(X_1,X_2,...,X_c)h_c(X_1,X_2',...,X_c')] - E[h_c(X_1,X_2,...,X_c)]E[h_c(X_1,X_2',...,X_c')] \\ &= 0\end{aligned}$$

より分かる．事実，空間 $H_{\{1,...,c\}}$ の定義から，任意の $X_1,...,X_c$ の狭義部分集合が与えられた時の $h_c(X_1,...,X_c)$ の条件付き期待値が 0 になるという意味で h_c は *strongly degenerate* である．言い換えると，任意の 1 つの position で積分したとき $\int h(x,X_2,...,X_c)dP(x) = 0$ として消える．同様の理由で，$U_{n,c}$ は $X_1,...,X_n$ の c より少ない要素に依存した任意の可測関数と無相関である．

任意の $c \geq 1$ に対して，$n^{c/2}U_{n,c}$ はその分散が $c!Eh_c^2(X_1,...,X_c)$ であるような分布に分布収束することが示される．これにより，$h_c \not\equiv 0$ であるような最小の c に対して，$n^{c/2}(U_n - \theta)$ は分布収束することが分かる．$c \geq 2$ に対して，極限分布は正規分布ではなく，*Gaussian chaos*（ガウシアンカオス）として知られる．

アイデアはシンプルなのだが，定理の主張は複雑であるのでまず $c = 3$ でカーネルとして

$$h_3(x_1,x_2,x_3) = f_1(x_1)f_2(x_2)f_3(x_3)$$

の形の積カーネルを用いる場合を考える．

積カーネルに対応する U 統計量は観測の和の多項式としてかける．記号の簡単のために

$$\begin{aligned}\mathbb{P}_n f &= \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad \text{(経験測度)} \\ \mathbb{G}_n f &= \sqrt{n}(\mathbb{P}_n - P)f \quad \text{(経験過程)}\end{aligned}$$

とする．P は観測 $X_1,...,X_n$ の分布である．

カーネル h_3 が狭義 degenerate であれば，任意の i で各関数 f_i は平均 0 より，$\mathbb{G}_n f_i = \sqrt{n}\mathbb{P}_n f_i$ である．よって，(i_1,i_2,i_3) を $\{1,...,n\}$ から 3 つ異なるものとしてとると（位置も考慮に入れる，例えば $(1,2,3)$ と $(2,1,3)$ は別物），

$$\begin{aligned}\frac{3!}{n^{3/2}}\binom{n}{3}U_{n,3} &= \frac{3!}{n^{3/2}}\binom{n}{3}\frac{1}{\binom{n}{3}}\sum_{(i_1,i_2,i_3)} f_1(x_{i_1})f_1(x_{i_2})f_1(x_{i_3}) \\ &= \frac{3!}{n^{3/2}}\sum_{(i_1,i_2,i_3)} f_1(x_{i_1})f_1(x_{i_2})f_1(x_{i_3}) \\ &= \mathbb{G}_n f_1 \mathbb{G}_n f_2 \mathbb{G}_n f_3 - \mathbb{P}_n(f_1f_2)\mathbb{G}_n f_3 \\ &\quad -\mathbb{P}_n(f_1f_3)\mathbb{G}_n f_2 - \mathbb{P}_n(f_2f_3)\mathbb{G}_n f_1 + 2\frac{\mathbb{P}_n(f_1f_2f_3)}{\sqrt{n}}\end{aligned}$$

２つ目の等号は右辺第 1 項目で左辺に現れない $i_1 = i_2 \neq i_3$, $i_1 = i_3 \neq i_2$, $i_1 \neq i_2 = i_3$, $i_1 = i_2 = i_3$ の場合を数えていることに注意すれば明らか.

大数の法則より，任意の f で $\mathbb{P}_n f \overset{\text{a.s}}{\to} Pf$．確率過程に対する中心極限定理から，$f \mapsto \mathbb{G}_n f$ はガウス過程に弱収束する．ここで

$$\{\mathbb{G}f : f \in L_2(\mathcal{X}, \mathcal{A}, P)\}$$

で平均 0，共分散関数 $E\mathbb{G}f\mathbb{G}g + Pfg - PfPg$ のガウス過程を表すとすれば，$\mathbb{G}_n \rightsquigarrow \mathbb{G}$ である．よって，

$$n^{3/2}U_{n,3} \rightsquigarrow \mathbb{G}f_1\mathbb{G}f_2\mathbb{G}f_3 - P(f_1f_2)\mathbb{G}f_3 - P(f_1f_3)\mathbb{G}f_2 - P(f_2f_3)\mathbb{G}f_1$$

である．極限はガウスベクトル $(\mathbb{G}f_1, \mathbb{G}f_2, \mathbb{G}f_3)$ の 3 次の多項式である．

一般の degenerate な U 統計量の極限に対する単純な形の公式はない．しかしながら，任意のカーネルは積カーネルの線形な無限個の組み合わせで書ける．U 統計量はそのカーネルの線形和であるので，一般の degenerate カーネル列の極限も同様のカーネルの極限の線形和である．

この問題を見ていく上で，与えられたカーネルを積カーネルの正規直交基底で分解すると便利である．これは常に可能である．ここで，$L_2(\mathcal{X}, \mathcal{A}, P)$ は可分であると仮定する．すなわち加算個の基底を持つ．

Example 12.8. General kernel. もし，$1 = f_0, f_1, f_2, ...$ が $L_2(\mathcal{X}, \mathcal{A}, P)$ の直交基底だとすると，$(k_1, ..., k_c)$ が非負整数上で値をとるとしたときの関数 $f_{k_1} \times \cdots \times f_{k_c}$ は $L_2(\mathcal{X}^c, \mathcal{A}^c, P^c)$ の直交基底となる．任意の二乗可積分カーネルは $a(k_1, ..., k_c) = \langle h_c, f_{k_1} \times ... \times f_{k_c}\rangle$ と定義すると，

$$h_c(x_1, ..., x_c) = \sum a(k_1, ..., k_c) f_{k_1} \times \cdots \times f_{k_c}$$

と書ける（ただのフーリエ展開）．

Proof. もう少し一般的な形で示しておく．ヒルベルト空間 $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ と内積から定義されたノルム $||\cdot||$ を考え，そこでの正規直交基底として $\{u_i\}_{i=1}^{\infty}$ をとる．このとき，任意の $x \in \mathcal{H}$ は

$$x = \sum_{i=1}^{\infty} \langle x, u_i\rangle u_i$$

と書けることを示せば Example の主張も成り立つ．直交性から

$$||x - \sum_{i=1}^{n} \langle x, u_i\rangle u_i||^2 = ||x||^2 - \sum_{i=1}^{n} |\langle x, u_i\rangle|^2$$

となる．よって $\sum_{i=1}^{n} |\langle x, u_i\rangle|^2$ は上に有界であり，$n \to \infty$ のとき収束する．

ここで，$y_n = x - \sum_{i=1}^{n} \langle x, u_i\rangle u_i$ と定めると，$m > n$ なる自然数に対して

$$||y_m - y_n||^2 = ||\sum_{i=n+1}^{m} \langle(x, u_i\rangle u_i||^2 = \sum_{i=n+1}^{m} |\langle x, u_i\rangle|^2$$

より，$\{y_n\}_{n=1}^{\infty}$ はコーシー列である．よってある $y \in \mathcal{H}$ が存在して $y_n \to y$ である．

このとき，$y = 0$ を示そう．任意の $n \in \mathbb{N}$ に対し

$$\langle y, u_n\rangle = \lim_{m\to\infty} \langle y_m, u_n\rangle = \lim_{m\to\infty} \langle x - \sum_{i=1}^{m} \langle x, u_i\rangle u_i, u_n\rangle$$

となる．右辺の内積は $m > n$ のとき 0 である．よって $\langle y, u_n\rangle = 0$ と分かる．$\{u_n\}$ は完全正規直交基底より $y = 0$ である．よって主張の式が得られた．□

Example 12.9. Second-order kernel. $c = 2$ の場合，我々の目的に特別に適した選択がある．カーネル h_2 は対称で 2 乗可積分と仮定していたので，積分作用素 $K : L_2(\mathcal{X}, \mathcal{A}, P) \to L_2(\mathcal{X}, \mathcal{A}, P)$ を，$Kf(x) =$

$\int h_2(x,y)f(y)dP(y)$ で定義すると，これは自己共役なヒルベルトシュミット作用素である．つまり $L_2(\mathcal{X},\mathcal{A},P)$ 上の内積 $\langle\cdot,\cdot\rangle$ とその内積により定義されるノルム $||\cdot||$ について

$$\langle Kf,g\rangle = \langle f,Kg\rangle$$

が成り立ち，$L_2(\mathcal{X},\mathcal{A},P)$ のある正規直交基底 $\{\varphi_i\}_{i=1}^I (I\in\mathbb{N}\cup\{\infty\})$ に対して，

$$\sum_{i=1}^I ||K\varphi_i||^2 < \infty$$

が成り立つ．自己共役性は

$$\begin{aligned}
(\text{左辺}) &= \int \left\{ \int h_2(x,y)f(y)dP(y) \right\} g(x)dP(x) \\
&= \int f(y) \left\{ \int h_2(y,x)g(x)dP(x) \right\} g(x)dP(y) \\
&= (\text{右辺})
\end{aligned}$$

より．ヒルベルトシュミット作用素であることは，

$$\begin{aligned}
\int |h_2(x,y)|^2 dP(y) &= \sum_{i=1}^I |\langle h_2(x,\cdot),\varphi_i\rangle|^2 \quad (\text{Parseval の等式}) \\
&= \sum_{i=1}^I |K\varphi_i(x)|^2 \quad (\text{積分作用素の定義}) \\
&< \infty \quad (h_2 \text{ の 2 乗可積分性})
\end{aligned}$$

よって，

$$\int\int |h_2(x,y)|^2 dP(x)dP(y) = \sum_{i=1}^I ||K\varphi_i||^2 < \infty$$

となり，確かにヒルベルトシュミット作用素である．ヒルベルトシュミット積分作用素の積分核である h_2 については，スペクトル定理から加算個の固有値 $\lambda_0,\lambda_1,...(\sum\lambda_k^2<\infty)$ と，対応する固有関数 $f_0,f_1,...$ が存在して

$$h_2(x,y) = \sum_{k=1}^\infty \lambda_k f_k(x) f_k(y)$$

と書ける．degenerate カーネル h_2 に対して，固有値 0 に対応する固有関数は 1 である．$f_0=1$ とする．

一般の場合における分解によるメリットは $f\times f$ の形の積関数しか必要なくなることである．