11 Projection

11.3 Projection onto Sums

$X_1, ..., X_n$ は独立な確率ベクトルとする．$\mathcal{S}$ は

$$\sum_{i=1}^{n} g_i(X_i),$$

(g_i は $Eg_i^2(X_i) < \infty$ を満たす任意の可測関数) で表される確率変数の集合であるとする（可測関数の和は可測関数なので当然線形空間になっている）．この元の分布収束は中心極限定理により得られる．このクラスへの確率変数の射影は Hajek projection と呼ばれる．

Lemma 11.10. $X_1, ..., X_n$ は独立な確率ベクトルとする．有限な 2 次モーメントを持つ任意の確率変数 T のクラス $\mathcal{S}$ への射影は

$$\hat{S} = \sum_{i=1}^{n} E(T|X_i) - (n-1)ET$$

で与えられる．

Proof. 右辺の確率変数は明らかに $\mathcal{S}$ の元になっている．よって後は，直交性を示せばよい．

$X_i (i = 1, ..., n)$ は独立なので，任意の i, j について，条件付き期待値 $E(E(T|X_i)|X_j)$ は

$$E(E(T|X_i)|X_j) = EE(T|X_i) = ET$$

となる．よって，任意の j で

$$\begin{aligned} E(\hat{S}|X_j) &= E\left(\sum_{i=1}^{n} E(T|X_i) - (n-1)ET|X_j\right) \\ &= (n-1)ET + E(T|X_j) - (n-1)ET \\ &= E(T|X_j) \end{aligned}$$

となる．ゆえに，

$$\begin{aligned} E(T-\hat{S})g_j(X_j) &= EE[(T-\hat{S})g_j(X_j)|X_j] \\ &= E\{g_i(X_j)E[T-\hat{S}|X_j]\} \\ &= E\{g_i(X_j)(E[T|X_j] - E[\hat{S}|X_j])\} \\ &= Eg_j(X_j)0 = 0. \end{aligned}$$

つまり，$T - \hat{S}$ は $\mathcal{S}$ に直交している．前節の議論により，示せた． □

$X_1, ..., X_n$ が独立で，さらに同一分布のケースについて考える．$T = T(X_1, ..., X_n)$ は置換対称な X_i の可測関数とする．このとき，

$$E(T|X_i = x) = ET(x, X_2, ..., X_n)$$

である．これは i に依らないので，射影 $\hat{S}$ はまた，$\sum_{i=1}^{n} g_i(X_i)$ の形で表されるより少ない確率変数の集合へ，T を射影したものとなっている．

11.4 Hoeffding Decomposition

Hajek projection は各 X_i の関数の和による最も良い近似を与えた．この近似は 2 つ，あるいはそれ以上の変数の関数の和を使うことにより改良することができる．これは Hoeffding docomposition(ヘフディング分解) を導く．

直交空間の和への射影は，一つ一つの空間への射影の和であるので（射影定理），提案された射影空間を直交分解すると便利である．独立な確率変数 $X_1, ..., X_n$ と部分集合 $A \subset \{1, ..., n\}$ が与えられたとき，H_A は

$$g_A(X_i : i \in A)$$

の形のすべての 2 乗可積分確率変数の集合を表すとする．ただし，g_A は

$$E(g_A(X_i : i \in A)|X_j : j \in B) = 0, \quad \text{every } B : |B| < |A|$$

なる $|A|$ 個の引数を持つ可測関数（$E(T|\phi) := ET$）．すなわち，

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

である．

A を $\{1, ..., n\}$ の部分集合としてとるとき，空間 H_A は互いに直交している．このことを用いて次のような次第に複雑になりながら，互いに直交する空間の列を作る．

$$[1], \quad \left[\sum_i g_{[i]}(X_i)\right], \quad \left[\sum \sum_{i<j} g_{[i,j]}(X_i, X_j)\right], \quad \cdots,$$

ただし，$g_{[i]} \in H_{[i]}, g_{[i,j]} \in H_{[i,j]}, \cdots$ である．これらの互いに直交する空間に確率変数を射影して近似していくことを考える．

$P_A T$ は T の H_A への射影を表すとする．このとき，H_A の直交性により，最初の r 個の空間の和に対する射影は，各空間への射影の和 $\sum_{|A| \leq r} P_A T$ である．最初の 2 つの空間の和への射影が Hajek projection である．より一般に，

$$\begin{aligned}
P_\phi T &= ET, \\
P_{[i]} T &= E(T|X_i) - ET, \\
P_{[i,j]} T &= E(T|X_i, X_j) - E(T|X_i) - E(T|X_j) + ET, .
\end{aligned}$$

となることが示せる．

以上の議論をより一般にまとめたのが次の補題である．

Lemma 11.11. $X_1, ..., X_n$ は独立な確率ベクトルとする．T は $ET^2 < \infty$ なる任意の確率ベクトルとする．このとき，T の H_A への射影は

$$P_A T = \sum_{B \subset A} (-1)^{|A|-|B|} E(T|X_i : i \in B)$$

で与えられる．

もし，与えられた集合 A について，任意の $B \subset A$ で $T \perp H_B$ であれば，

$$E(T|X_i : i \in A) = 0$$

となる．従って，任意の $B \subset A$ なる空間 H_B の和は，$(X_i : i \in A)$ を引数に持つすべての二乗可積分関数を含む．

Proof. 記号の簡略のために，$E(T|A)$ で $E(T|X_i : i \in A)$, g_A で $g_A(X_i : i \in A)$ を表すことにする．

まず．lemma で定義された P_AT が H_A に含まれていることを示す．$X_1, ..., X_n$ は独立より，任意の $\{1, ..., n\}$ の部分集合 A と B で $E(E(T|A)|B) = E(T|A \cap B)$ が成り立つ．よって，lemma で定義された P_AT と $C \subsetneq A$ について，

$$\begin{aligned}
E(P_AT|C) &= \sum_{B \subset A} (-1)^{|A|-|B|} E(T|B \cap C) \\
&= \sum_{D \subset C} \sum_{j=0}^{|A|-|C|} (-1)^{|A|-|D|-j} \binom{|A|-|C|}{j} E(T|D) \\
&= \sum_{D \subset C} E(T|D) \sum_{j=0}^{|A|-|C|} \binom{|A|-|C|}{j} 1^j (-1)^{|A|-|C|-j} (-1)^{|C|-|D|} \\
&= \sum_{D \subset C} E(T|D) 0 \\
&= 0.
\end{aligned}$$

つまり，$E(P_AT|C) = 0$ である．

２つ目の等号を導く．まず，$B = (B \cap C) \cup (B \cap C^c)$ と排反に分解する．ここで，$D := B \cap C,\ D' := B \cap C^c$ と定義する．すると，

$$\begin{aligned}
B \cap C &= \{(B \cap C) \cup (B \cap C^c)\} \cap C \\
&= B \cap C = D
\end{aligned}$$

で，また $|B| = |D| + |D'|$ より，

$$\begin{aligned}
\sum_{B \subset A} (-1)^{|A|-|B|} E(T|B \cap C) &= \sum_{D \subset C} \sum_{D' \subset C^c} (-1)^{|A|-|D|-|D'|} E(T|D) \\
&= \sum_{D \subset C} \sum_{j=0}^{|A|-|C|} \binom{|A|-|C|}{j} (-1)^{|A|-|D|-j} E(T|D)
\end{aligned}$$

となり，確かに成り立つ．

$C \subset A$ としても $C = A$ のときは２式目を見ると和が取られないことが分かるので，一般性を失わない．よって任意の $C \subset A$ で $E(P_AT|C) = 0$ である．$|C| < |A|$（$C \subset A$ とは限らない）としても，独立性の仮定から $E(P_AT|C) = 0$ となる．P_AT が $g_A(X_i : i \in A)$ の形で書けるのは明らかなので，以上により $P_AT \in H_A$ である．

次に直交性，すなわち任意の $g_A \in H_A$ に対して，

$$E(T - P_AT)g_A = 0$$

となることを示す．任意の $g_A \in H_A$ について，

$$\begin{aligned}
E(T - P_AT)g_A &= E\left[(T - \sum_{B \subset A} (-1)^{|A|-|B|} E(T|B))g_A\right] \\
&= E(T - E(T|A))g_A - \sum_{B \subsetneq A} (-1)^{|A|-|B|} E[E(T|B)g_A] \\
&= E(T - E(T|A))g_A - \sum_{B \subsetneq A} (-1)^{|A|-|B|} E[E(T|B)E(g_A|B)].
\end{aligned}$$

２式目は $\sum_{B \subset A} (-1)^{|A|-|B|} E(T|B) = E(T|A) + \sum_{B \subsetneq A} (-1)^{|A|-|B|} E(T|B)$ と分解した．任意の $g_A \in H_A$ に対して，まず，前章の議論から一項目は $\{X_i : i \in A\}$ で貼られる空間での直交性から 0 となる．また，

$$E(g_A|B) = 0$$

から二項目も 0 である．よって $E(T-P_AT)g_A=0$ であり直交性も示せた．lemma で定義された P_AT は確かに T の H_A への射影になっている．

((こっから)) 次に，２つ目の主張，もし，与えられた集合 A について，任意の $B\subset A$ で $T\perp H_B$ であれば，

$$E(T|X_i:i\in A)=0$$

となることを示す．$r=|A|$ とおいて帰納法で示す．$T\perp H_\phi$ ならば $E(T|\phi)=ET=0$ である．よって $r=0$ のとき，主張は成り立つ．

次に，$|A|=0,...,r-1$ で成り立つと仮定して，$|A|=r$ で成り立つことを示す．任意の $B\subset A$ で $T\perp H_B$ で成立であれば，任意の $C\subset B$ で，$T\perp H_C$ である．ゆえに仮定から，任意の $r-1$ 個以下の要素を持つ $B\subset A$ で $E(T|B)=0$ である．ゆえに，

$$\begin{aligned}P_AT&=E(T|A)+\sum_{B\subsetneq A}(-1)^{|A|-|B|}E(T|B)\\&=E(T|A)=0\end{aligned}$$

となり示せた．

最後の主張，『$B\subset A$ なる空間 H_B の和は，$(X_i:i\in A)$ を引数に持つすべての二乗可積分関数を含む』を示す．

これは，$T_A:=T-\sum_{B\subset A}P_BT$ が $(X_i:i\in A)$ にのみ依存した任意の T について 0 となれば成り立つ．すなわち，$T=T(X_i:i\in A)$ とすると，

$$T_A:=T-\sum_{B\subset A}P_BT=0.$$

$T=T(X_i:i\in A)$ のとき，T_A は当然 $\{X_i:\in A\}$ のみに依存する．よって，任意の $B\subset A$ について $T_A\perp H_B$ であるので，

$$T_A=E(T_A|A)=0$$

である．２つ目の等号は２つ目の主張より．

以上により全て示せた． □

$T=T(X_1,...,X_n)$ は置換対称で，$X_1,...,X_n$ は i.i.d とする．T のヘフディング分解（$\sum_{r=0}^n\sum_{|A|=r}P_AT$ で定義される）は

$$\begin{aligned}T&=\sum_{r=0}^n\sum_{|A|=r}g_r(A)\\&=\sum_{r=0}^n\sum_{|A|=r}\sum_{B\subset A}(-1)^{|A|-|B|}E(T|B)\\&=\sum_{r=0}^n\sum_{|A|=r}P_AT\\\text{for }g_r(x_1,...,x_r)&=\sum_{B\subset\{1,...,r\}}(-1)^{r-|B|}ET(x_i\in B,X_i\notin B)\end{aligned}$$

となる．T とそのヘフディング分解が等しくなるのは，先の lemma の３つ目の主張より，$H_A(\forall A\subset\{1,...,n\})$ を足し合わせた空間が $\{X_i:i\in A\}$ を引数に持つ任意の二乗可積分関数を含むことより分かる．

実は T の表現の内側の sum は r 次の degenerate kernel により定義される U 統計量となっている（１２章）．P_AT はそれぞれ直交する空間に射影しているので，T の分散は

$$\mathrm{var}T=\sum_{r=1}^n\binom{n}{r}Eg_r^2(X_1,...,X_r)$$

として得られる．