Pooling incomplete samples における Kullback 情報量の直和分解について

関東学院大学経済学部　布能 英一郎

Kullback 情報量の直和分解　多項分布の 2 標本問題 $(X_1^{[i]},\cdots,X_k^{[i]}) \sim M(N^{[i]}: p_1^{[i]},\cdots,p_k^{[i]})$, $i=1,2$, $H_1: p_j^{[1]} \neq p_j^{[2]}$, $H_2: p_j^{[1]} = p_j^{[2]} = p_j$, $(j=1,\cdots,k)$ を考える。このとき、Kullback 情報量 $\hat{I}(p^*: p) = \sum_{i=1}^2 N_1^{[i]} \sum_{j=1}^k x_j^{[i]} \log(x_j^{[i]}/N_1^{[i]} p_j)$ は、between $\hat{I}(\hat{p}: p) = \sum_{j=1}^k (x_j^{[1]} + x_j^{[2]}) \log((x_j^{[1]} + x_j^{[2]})/(N^{[1]} + N^{[2]}) p_j)$ と within $\hat{I}(p^*: \hat{p}) = \sum_{i=1}^2 \sum_{j=1}^k x_j^{[i]} \log((N^{[1]} + N^{[2]}) x_j^{[i]}/N^{[i]}(x_j^{[1]} + x_j^{[2]}))$ に直和分解できる。

Pooling incomplete samples の場合　離散データ解析において、オリジナルな観測と、カテゴリー減少が生じて、その各セルの生起確率がオリジナルな観測のセル確率を比例配分とした観測との同時分布にもとづいて統計的推測を行うのが Pooling incomplete samples である。この場合の 2 標本問題において、Kullback 情報量の直和分解が成り立つこともあれば、成り立たないこともある。

定理 1.　$\boldsymbol{X}^{[1]}, \boldsymbol{X}^{[2]}, \boldsymbol{Y}^{[1]}, \boldsymbol{Y}^{[2]}$ は互いに独立で $\boldsymbol{X}^{[i]} = (X_1^{[i]},\cdots,X_m^{[i]},\cdots,X_k^{[i]})$, $\boldsymbol{Y}^{[i]} = (Y_1^{[i]},\cdots,Y_m^{[i]})$, $i=1,2$、$\boldsymbol{X}^{[i]} \sim M(N_1^{[i]}; p_1^{[i]},\cdots,p_m^{[i]},\cdots,p_k^{[i]})$,　$\boldsymbol{Y}^{[i]} \sim M(N_2^{[i]}; p_1^{[i]}/\sum_{l=1}^m p_l^{[i]},\cdots,p_m^{[i]}/\sum_{l=1}^m p_l^{[i]})$ $i=1,2$,　$H_1: p_j^{[1]} \neq p_j^{[2]}$, $H_2: p_j^{[1]} = p_j^{[2]}$ とする。このとき、Kullback 情報量の直和分解が成り立つ。

さて、負の多項分布の場合、通常の 2 標本問題では Kullback 情報量の直和分解が成り立つものの、Pooling incomplete samples の場合、すなわち、定理 1. にて、分布の仮定を $\boldsymbol{X}^{[i]} \sim NM(r_1^{[i]}; p_1^{[i]},\cdots,p_m^{[i]},\cdots,p_k^{[i]})$,　$\boldsymbol{Y}^{[i]} \sim NM(r_2^{[i]}; p_1^{[i]}/\sum_{l=0}^m p_l^{[i]},\cdots,p_m^{[i]}/\sum_{l=0}^m p_l^{[i]})$ $i=1,2$,　とした場合、Kullback 情報量の直和分解が成り立たない。[**注** 負の多項分布の場合、$p_0^{[i]} = 1 - \sum_{l=1}^k p_l^{[i]} > 0.$]　他方、次の定理が得られた。

定理 2.　$\boldsymbol{X}^{[1]}, \boldsymbol{X}^{[2]}, \boldsymbol{Y}^{[1]}, \boldsymbol{Y}^{[2]}$ は互いに独立で $\boldsymbol{X}^{[i]} \sim NM(r_1^{[i]}; p_1^{[i]},\cdots,p_m^{[i]},\cdots,p_k^{[i]})$, $\boldsymbol{Y}^{[i]} \sim M(N_2^{[i]}; p_1^{[i]}/\sum_{l=1}^m p_l^{[i]},\cdots,p_m^{[i]}/\sum_{l=1}^m p_l^{[i]})$, $i=1,2$,　ならば、Kullback 情報量の直和分解が成り立つ。

同様に、Poisson 分布の場合にも、次の定理が得られた。

定理 3.　各 $i=1,2$ に対して $X_j^{[i]} \sim Poisson(\lambda_j^{[i]})$,　$j=1,\cdots,k$,　$\boldsymbol{Y}^{[i]} = (Y_1^{[i]},\cdots,Y_m^{[i]}) \sim M(N_2^{[i]}, \lambda_1^{[i]}/\sum_{l=1}^m \lambda_l^{[i]},\cdots,\lambda_m^{[i]}/\sum_{l=1}^m \lambda_l^{[i]})$ であって $X_1^{[i]},\cdots,X_k^{[i]}, \boldsymbol{Y}^{[i]}$, $i=1,2$ はすべて独立。このとき、Kullback 情報量の直和分解が成り立つ。

定理 2, 3 が成り立つ背景の考察については、当日報告する。

文献　[**1**] Asano, C. (1965). On estimating multinomial probabilities by pooling incomplete samples, *Ann. Inst. Stat. Math.* **17**, 1-13.　[**2**] Kullback, S. (1968). *Information Theory and Statistics.* Revised version　Dover.