A Frequentist Model Averaging Approach

Proposed Method

Consider a single-arm basket trial evaluating a binary endpoint in J baskets. Suppose there are n subjects in total. Denote X_i the basket membership, such that $X_i = j$ if subject i is in basket j, for $i = 1, \ldots, n$, and $j = 1, \ldots, J$. Then the observed response for subject i will be obtained as

$$Y_i \sim Bernoulli(\pi(X_i)),$$

for $i = 1, \ldots, n$, where $\pi(j)$ denotes the response rate in basket j. Let $\mathcal{D} = \{(Y_i, X_i), i = 1, \ldots, n\}$ denote the observed data. The proposed model averaging approach is based on the K-fold cross-validation:

1. Randomly split $\mathcal{D}$ into K folds (e.g., $K = 10$), denoted as $\mathcal{D}_{(1)}, \ldots, \mathcal{D}_{(K)}$.

2. For $k = 1, \ldots, K$, do:

   i. Let $\mathcal{D}^{(k)} = \mathcal{D} - \mathcal{D}_{(k)}$. The first model estimates the response rates as

   $$\widehat{\pi}^1_{(k)}(X = x) = \sum_{j=1}^{J} I(x = j) \frac{\sum_i^n Y_i I(X_i = j) I(i \in \mathcal{D}^{(k)})}{\sum_i^n I(X_i = j) I(i \in \mathcal{D}^{(k)})}.$$

   ii. Conduct all pairwise Fisher's exact test for all baskets using $\mathcal{D}^{(k)}$. Treat two baskets with the largest p-value among all pairwise tests as homogeneous, say baskets p and q, where $p, q \in \{1, \ldots, J\}$ and $p \neq q$. The second model estimates the response rates of baskets p and q jointly, such that

   $$\begin{aligned} \widehat{\pi}^2_{(k)}(X = x) &= I(x \neq p, q) \widehat{\pi}^1_{(k)}(x) \\ &\quad + I(x = p \text{ or } x = q) \frac{\sum_i^n Y_i \{I(X_i = p) + I(X_i = q)\} I(i \in \mathcal{D}^{(k)})}{\sum_i^n \{I(X_i = p) + I(X_i = q)\} I(i \in \mathcal{D}^{(k)})}. \end{aligned}$$

   iii. Combine baskets p and q. Continue the process that treats two baskets with the largest p-value among all pairwise tests as homogeneous, and then estimates their response rates jointly. This process gives a sequence of models, denoted as $\widehat{\pi}^1_{(k)}(X = x), \widehat{\pi}^2_{(k)}(X = x), \ldots, \widehat{\pi}^J_{(k)}(X = x)$. Note that there are J such models, and the last model assumes all baskets are homogeneous, such that

   $$\widehat{\pi}^J_{(k)}(X = x) = \frac{\sum_i^n Y_i I(i \in \mathcal{D}^{(k)})}{\sum_i^n I(i \in \mathcal{D}^{(k)})}.$$

   iv. Calculate the out-of-fold predicted response rate for $\mathcal{D}_{(k)}$ based on the sequence of models.

3. Denote the out-of-fold predicted response rate based on the sequence of models as $\widetilde{\pi}^1(X = x), \ldots, \widetilde{\pi}^J(X = x)$. For each $j = 1, \ldots, J$, calculate the optimal model

averaging weights by minimizing the cross-validated Sigmoid cross entropy loss:

$$\{\widehat{w}_{j1},\ldots,\widehat{w}_{jJ}\} = \text{argmin}_{w_1,\ldots,w_J} \sum_{i=1}^{n} I(X_i = j)\Bigg\{Y_i \log\Bigg(\sum_{l=1}^{J} w_l \widetilde{\pi}^l(X_i)\Bigg) + (1-Y_i)\log\Bigg(1-\sum_{l=1}^{J} w_l \widetilde{\pi}^l(X_i)\Bigg)\Bigg\}.$$

4. The final estimated response rate is

$$\widehat{\pi}(X = x) = \sum_{j=1}^{J} I(x = j) \sum_{l=1}^{J} \widehat{w}_{jl}\widehat{\pi}^l(x),$$

where $\widehat{\pi}^l(X = x)$ are calculated based on the sequence of models formed and estimated based on all observations in $\mathcal{D}$, where $l = 1,\ldots,J$.

Decision Criteria

For testing the following hypothesis:

$$H_{0j} : \pi(j) \leq \pi_0$$

$$H_{1j} : \pi(j) > \pi_0$$

for each $j = 1,\ldots,J$, the decision criteria targeting at a basket-wise type I error rate of α can be obtained as the following: simulate data from B (e.g., $B = 1000$) basket trials in which all J baskets has a response rate of π_0. For each $b = 1,\ldots,B$, calculate the estimated response rates, $\{\widehat{\pi}(1),\ldots,\widehat{\pi}(J)\}_b$ using the proposed method. For each $j = 1,\ldots,J$, the null hypothesis H_{0j} is rejected if the estimated response rate based on the observed data is greater than ψ_j, where ψ_j is the $1-\alpha$ quantile of $\{\widehat{\pi}(j)\}_{b=1,\ldots,B}$.

If an interim analysis for futility is conducted, and we target at a basket-wise type I error rate of α and a basket-wise type II error rate of β, then the following decision criteria are used:

1. Simulate data from B basket trials in which all J baskets has a response rate of π_0. At the interim analysis, the basket j will be stopped for futility if the estimated response rate of basket j is lower than a cutoff of ψ_0. At the final analysis, the null hypothesis will be rejected for basket j if the estimated response rate of basket j is higher than a cutoff of ψ_1.

2. Simulate data from B basket trials in which all J baskets has a response rate of π_0 except the basket with the lowest accrual rate, where the response rate is simulated under the alternative hypothesis of π_1. Use the same decision rule as in Step 1.

3. Choose ψ_0 and ψ_1 such that the empirical basket-wise rejection rate in Step 1 is lower than α and the empirical basket-wise rejection rate for basket with response rate π_1 in Step 2 is lower than β. Among the set of $\{\psi_0,\psi_1\}$ that satisfy the above conditions, choose the one that minimizes the total sample size.