ARTICLE

Supplementary Materials

1 | PROPOSITIONS AND THEOREMS

We have used some propositions and theorems in Ghosal and van der Vaart (2017). Here we listed those theorems using the same numbering in the book.

**Definition 4.1** A random measure \( P \) on \((X, \mathcal{X})\) is said to possess a Dirichlet process distribution \( DP(\alpha) \) with base measure \( \nu \), if for every finite measurable partition \( A_1, \ldots, A_k \) of \( X \), \( (P(A_1), \ldots, P(A_k)) \sim \text{Dir}(k; \alpha(A_1), \ldots, \alpha(A_k)) \), where \( \text{Dir}(k; \alpha_1, \ldots, \alpha_k) \) denotes a Dirichlet distribution with \( k \) categories and concentration parameters \( \alpha(A_1), \ldots, \alpha(A_k) \).

**Theorem 4.6** The posterior distribution given an i.i.d sample \( X_1, \ldots, X_n \) from the \( DP(\alpha) \)-process is \( DP(\alpha + \sum_{i=1}^n \delta_X) \)-process.

**Proposition G.10** If \( X \sim \text{Dir}(k; \alpha), Y \sim \text{Dir}(k; \beta) \), and \( V \sim \text{Be}(||\alpha||, ||\beta||) \) are independent random vectors, then \( VX + (1-V)Y \sim \text{Dir}(k; \alpha + \beta) \). In particular, if \( X \sim \text{Dir}(k; \alpha) \) and \( V \sim \text{Be}(||\alpha||, ||\beta||) \), then \( VX + (1-V)e_i \sim \text{Dir}(k; \alpha + \beta e_i) \), where \( e_i \) is the \( i \)-th unit vector in \( \mathbb{R}^k \), and \( i \in \{1, \ldots, k\} \).

**Theorem 4.19 (ii)** Let \( P_N \) be a Dirichlet-multinomial process of order \( N \to \infty \) with parameters \( (M/N, \ldots, M/N) \) and \( G \), then \( \int \phi dP_N \to \int \phi dP \), where \( P \sim DP(MG) \), for any \( \phi \in L_1(G) \).

2 | ALGORITHMS

3 | SIMULATION RESULTS WITHOUT VERIFICATION BIAS

We set \( n_0 = n_1 = n_2 = n \) and \( n \) is taken to be 50 and 100. Still we consider the 6 different underlying true model as we done in the simulations under verification bias.

The results for the BB method will be based on 1000 MCMC resamples. We compare the BB method to two empirical methods and two kernel methods. The first empirical method is proposed by Li and Zhou by integrating the empirical estimate of the ROC surface, which will be denoted as EP in the table given below. The second empirical method is the unbiased nonparametric Mann-Whitney U-statistic of the probability \( P(X < Y < Z) \) and is extended to the circumstances with ties by Nakas and Yiannoutsos. This method will be denoted as MW in the table. Notice that the empirical estimate of ROC surface is not smooth. The two kernel estimator are proposed by Kang and Tian, which smooth out the surface using Gaussian kernel. The results are given in Table.

The methods are all very comparable. BB has relatively low bias and MSE in most cases although the difference is insignificant. BRL, as expected, has the best performance for the trinormal setting. MW has the best performance among other estimators and should also be considered as a good choice when estimating the VUS. The problem with MW is that it cannot give a smooth estimate for the ROC surface, which is often desirable.
Algorithm 1 Dirichlet Process (DP) method for estimating ROC surface under verification bias.

Input: $S, L$, initial value $D$, $(\lambda_0, \lambda_1, \lambda_2)$, niter. $M, a, \beta, N^*$
Output: $(F_0, F_1, F_2)$

1: $\mu = S$
2: $\sigma^2 \sim IG(0.1, 0.1)$ # Generate initial values for $\mu$ and $\sigma$
3: for $m \leftarrow 1$ to niter do
4:     for $i \leftarrow 1$ to $N$ do
5:         # Step 1: Impute labels for $D_i$ if the true disease status are unknown
6:             if $L_i = 3$ then
7:                 for $d \leftarrow 0$ to 2 do
8:                     $n_{d}^i = \sum_{j \neq d} 1\{D_j = d\}$
9:                     $V \sim Be(M_f, n_{d}^i)$
10:                    $U \sim U(0, 1)$
11:                    if $U < V$ then
12:                        $(q_1, \ldots, q_{N^*}) \sim Dir(N^*; M_{d}/N^*, \ldots, M_{d}/N^*)$
13:                        for $k \leftarrow 0$ to $N^*$ do
14:                            $\theta_{d,k,i} \sim G_d$
15:                        end for
16:                        $f_d(S_i) = \sum_{k=1}^{N^*} q_k \phi(S_i - \mu_{d,k})$
17:                    else
18:                        $(w_1, \ldots, w_{n_d}) \sim Dir(n_d^i; 1, 1, \ldots, 1)$
19:                        $f_d(S_i) = \sum_{k=1}^{n_d^i} w_k \phi(S_i - \mu_{j_k})$ where $j_k$ satisfies that $D_{j_k} = d$ and $j_k \neq i$
20:                     end if
21:                 end for
22:                 $f(S_i) = \lambda_0 f_0(S_i) + \lambda_1 f_1(S_i) + \lambda_2 f_2(S_i)$
23:                 $(1\{D_i = 0\}, 1\{D_i = 1\}, 1\{D_i = 2\}) \sim Mult\left(1, \left(\frac{\lambda_0 f_0(S_i)}{f(S_i)}, \frac{\lambda_1 f_1(S_i)}{f(S_i)}, \frac{\lambda_2 f_2(S_i)}{f(S_i)}\right)\right)$
24:             end if
25:         end for
26:     # Step 2: Update $\theta_i$ according to Dirichlet Process
27:     for $j \leftarrow 0$ to $N$ do
28:         if $j = 0$ then
29:             $q_{i,j} = \frac{M_{D_i} \sqrt{\alpha_{D_i}} \Gamma(s_{D_i} + 1/2) \rho_{D_i}^{r_{D_i}}}{\sqrt{1 + \alpha_{D_i} \Gamma(s_{D_i})(\beta_{D_i} + a_{D_i}(s_D - m_{D_i}))^{r_D}}}^{r_{D_i}+1/2}$
30:         else if $D_j = D_i$ and $j \neq i$
31:             $q_{i,j} = \sigma_j^{-1} \exp\{-(S_i - \mu_j)^2/(2\sigma_j^2)\}$
32:         else
33:             $q_{i,j} = 0$
34:         end if
35:     end for
36:     normalize $q_i = (q_{i,0}, \ldots, q_{i,N})$
37:     $\sigma_{0,j}^{-2} \sim Ga(s_{D_j} + 1/2, \beta_{D_j} + d(S_j - m_{D_j})^2/(2(1 + a_{D_j}))$
38:     $\mu_{0,j} \sim N((S_j + a_{D_j}m_{D_j})/(1 + a_{D_j}), \sigma_{0,j}^{-2}/(1 + a_{D_j}))$
39:     $\theta_{i,j} = (\mu_{0,j}, \sigma_{0,j}^{-2})^T$
40:     $r_i = Mult(1, q_i)$
41:     $\theta_i = (\theta_{0,i}, \theta_1, \ldots, \theta_N)^T$
42: end for
Trinormality assumption checking

To do a quick visual check on the trinormality assumption, we can estimate the transformation function $H(x) = \Phi^{-1} \cdot F_1(x)$ where $F_1(x)$ can be estimated using a empirical distribution function of $\{S_i : L_i = 1\}$. We can then apply this transformation on $S_i$ for $i = 1, \ldots, N$. Denote the transformed observations by $\hat{Q}$. The transformed observations $\{\hat{Q}_i : L_i = 1\}$ should follow a standard normal distribution so if we plot the quantiles of $\{\hat{Q}_i : L_i = 1\}$ versus the quantiles of a standard normal distribution, it should follow the $y = x$ line. If the trinormality assumption holds, the transformed observations $\{\hat{Q}_i : L_i = 0\}$ should follow a standard normal distribution.

### 4 | REAL DATA

#### 4.1 | Trinormality assumption checking

First, recall the trinormality assumption is that, under some strictly monotone increasing transformation $H$, the transformed observations $Q_i = H(S_i)$, $i = 1, \ldots, N$, satisfy

$$Q_i | \{D_i = k\} \overset{i.i.d.}{\sim} N(\mu_k, \sigma^2_k), \quad k = 0, 1, 2,$$

for some $\mu_0 < \mu_1 < \mu_2$ and $\sigma_0, \sigma_1, \sigma_2 > 0$. To ensure the identifiability of the model, the distribution of the middle group (without loss of generality) has been set to be the standard normal (i.e. $\mu_1 = 0, \sigma_1 = 1$).

To do a quick visual check on the trinormality assumption, we can estimate the transformation function $H(x) = \Phi^{-1} \cdot F_1(x)$ where $F_1(x)$ can be estimated using a empirical distribution function of $\{S_i : L_i = 1\}$. We can then apply this transformation on $S_i$ for $i = 1, \ldots, N$. Denote the transformed observations by $\hat{Q}$. The transformed observations $\{\hat{Q}_i : L_i = 1\}$ should follow a standard normal distribution so if we plot the quantiles of $\{\hat{Q}_i : L_i = 1\}$ versus the quantiles of a standard normal distribution, it should follow the $y = x$ line. If the trinormality assumption holds, the transformed observations $\{\hat{Q}_i : L_i = 0\}$ should follow a standard normal distribution.
Before estimating ROC surface, we did a quick check on the data by plotting out the boxplot of the labeled observations of CA125 and HE4 within each class. Judging from the plots, the trinormality assumption is quite satisfied for both CA125 and HE4. So we can apply BRL method to estimate the ROC surface.

**4.2  Data exploratory analysis**

Before estimating ROC surface, we did a quick check on the data by plotting out the boxplot of the labeled observations of CA125 and HE4 within each class. Judging from the plots, the trinormality assumption is quite satisfied for both CA125 and HE4. So we can apply BRL method to estimate the ROC surface.

Judging from the plots, the trinormality assumption is quite satisfied for both CA125 and HE4. So we can apply BRL method to estimate the ROC surface.
FIGURE 1 Trinormality check for CA125

FIGURE 2 Trinormality check for HE4


FIGURE 3 Boxplot for CA125

FIGURE 4 Boxplot for HE4

FIGURE 5 Boxplot for serum albumin