1 Approximated Quadratic Exponential Model

Let L be a K-dimensional binary random variable denoting the true state of the lung. The general form of the log-linear model is:

$$P(L = l; \Theta) = \exp\{\Theta_1^T l + \Theta_2^T u_2 + \ldots + \Theta_K^T u_K\}/A(\Theta)$$

where w_k is a $\binom{K}{k} \times 1$ vector of the k-way cross-products of l, $k = 1, \ldots, K$, and $\Theta = (\Theta_1, \ldots, \Theta_K)$ contains the the canonical parameters, which is a $(2^K - 1) \times 1$ vector. Θ_1 contains the k conditional log odds' and the rest contains the conditional log odds ratios, regarded as the association parameters. Moreover, let $l^* = (l, w_2, \ldots, w_K)^T$, the normalizing term is defined as

$$A(\Theta) = \sum_{l^* \in \{0,1\}^K} \exp\{\Theta^T l^*\}$$

We propose an Approximated Quadratic Exponential Model (AQE) for large K with competing binary variables by defining

$$P(L = l; \Theta) = \exp\{\Theta_1^T l + \Theta_2^T u_2\}/\hat{A}(\Theta) \tag{1}$$

$$\hat{A}(\Theta) = \sum_{l : l^T \mathbf{1} \leq S_{\max}} \exp\{\Theta^T l\} \tag{2}$$

Let $M_i^{GS} \in \{0,1\}^K$ be the observed GS measurement, $M_i^{SS} \in \{0,1\}^K$ be the observed SS measurement, $M_i^{BS} \in \{0,1\}^K$ be the observed BS measurement and $L_i \in \{0,1\}^K$ be the latent status for subject i. Let $\gamma \in [0,1]^K$ and $\delta \in [0,1]^K$ represent the True Positive Rate (TPR) and False Positive Rate (FPR) for BS measurements respectively, and let $\eta \in [0,1]^K$ be the TPR for SS measurements. Also, let $\mathbb{L}$ be the set of all allowed values of L, such that $|\mathbb{L}| = J^*$ and l_j be the jth element in $\mathbb{L}$.

1.1 The Likelihood for Cases

For cases without GS measurements, and under the conditional independence assumption for measurement given latent class, the likelihood function is

$$\begin{aligned}
P(M_i^{SS}, M_i^{BS}|\mu,\pi,\eta,\gamma,\delta) =& \sum_{j=1}^{J^*} P(M_i^{SS}, M_i^{BS}, l_j|\mu,\pi,\eta,\gamma,\delta) \\
=& \sum_{j=1}^{J^*} \left[P(M_i^{SS}|l_j,\eta)P(M_i^{BS}|l_j,\gamma,\delta)P(l_j|\mu,\pi)\right]
\end{aligned}$$

where $P(l_j|\mu,\pi)$ is defined using (10), and

$$\begin{aligned}
P(M_i^{SS}|l_j,\eta) &= \prod_{k=1}^{K} P(M_{ik}|l_{jk},\eta_k) \\
&= \prod_{k=1}^{K} (\eta_k^{l_{jk}} l_{jk})^{M_{ik}} (1-\eta_k)^{l_{jk}(1-M_{ik})} \quad (3) \\
P(M_i^{BS}|l_j,\gamma,\delta) &= \prod_{k=1}^{K} P(M_{ik}|l_{jk},\gamma_k,\delta_k) \\
&= \prod_{k=1}^{K} (\gamma_k^{l_{jk}} \delta_k^{1-l_{jk}})^{M_{ik}} [(1-\gamma_k)^{l_{jk}}(1-\delta_k)^{1-l_{jk}}]^{1-M_{ik}} \quad (4)
\end{aligned}$$

For cases with GS measurements, we have $L_i = M_i^{GS}$, then the likelihood is

$$\begin{aligned}
P(M_i^{GS}, M_i^{SS}, M_i^{BS}|\mu,\pi,\eta,\gamma,\delta) =& P(M_i^{SS}, M_i^{BS}|M_i^{GS}\eta,\gamma,\delta)P(M_i^{GS}|\mu,\pi) \\
=& P(M_i^{SS}|M_i^{GS},\eta)P(M_i^{BS}|M_i^{GS},\gamma,\delta)P(M_i^{GS}|\mu,\pi)
\end{aligned}$$

where $P(M_i^{SS}|M_i^{GS},\eta)$ is defined using (12), $P(M_i^{BS}|M_i^{GS},\gamma,\delta)$ is defined using (13), and $P(M_i^{GS}|\mu,\pi)$ is defined using (10).

1.2 The Likelihood for Controls

For controls, we only have BS measurements and we know that their lungs were not infected. Since (μ,π) are defined for case only, they are not involved in the likelihood for controls, thus the likelihood function is:

$$P(M_i^{BS}|\gamma, \delta) = \prod_{k=1}^{K} \delta_k^{M_{ik}} (1-\delta_k)^{(1-M_{ik})}$$

1.3 The Joint Density