Theory Behind the Iterative Method for Estimating Parameters for the PEB Algorithm

Jason Thorpe

April 21, 2015

At the center for the Parametric Empirical Bayes (PEB) algorithm are the assumptions that (1) after an appropriate transformation and in the absence of a condition of interest, expected marker levels $\{\mu_i\}$ for individuals in a population follow a normal distribution with mean μ and variance τ^2, (2) that markers are measured at a discrete series of events and (3) at each event, marker levels $y_{i,j}$ follow a normal distribution with mean μ_i and variance σ^2. In particular, differences between realized marker levels and expected marker levels ($\mu_i - y_{i,j} = e_{i,j}$) are assumed to be independent and identically distributed ($i.i.d.$) across all individuals in the population with mean 0 and variance σ^2 on the transformed scale.

The PEB algorithm is implemented by screening individuals at successive events by comparing the current result to a threshold which, in the absence of the condition of interest has a small fixed probability of being exceeded. Having estimates $\hat{\mu}$, $\hat{\tau}^2$, and $\hat{\sigma}^2$, the threshold for individual i's n^{th} screen is set to:

$$\hat{\mu_i} + Z_p \cdot \sqrt{\hat{V}_n}$$

where:

$$\begin{aligned}
\hat{\mu_i} &= \beta_n \cdot \overline{Y}_n + (1 - \beta_n) \cdot \hat{\mu} \\
\hat{V}_n &= \hat{\sigma} + \hat{\tau} \cdot (1 - \beta_n) \\
\beta_n &= \frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2}
\end{aligned}$$

and where Z_p is the p^{th} quantile from the standard normal distribution.

1 Estimating μ, τ^2 and σ^2

1.1 Preliminaries

For a set of independent unbiased estimators $\{s_i\}$ of a parameter S with error variances $EV(s_i)$, if:

$$w_i \propto 1/EV(s_i) \tag{1}$$

then ,

$$\hat{S} = \frac{\sum_i w_i \cdot s_i}{\sum_i w_i} \tag{2}$$

is an minimum variance unbiased estimator of S conditional on $\{s_i\}$. In the case that $w_i = 1/EV(s_i)$, then:

$$EV(\hat{S}) = \frac{1}{\sum_i w_i} \tag{3}$$

The assumption that realized marker levels depend only on the expected marker level of the individual from whom the marker was measured in the absence of disease allows for the use of standard MVU estimators for the mean and variance parameters for that individual and aggregate them in such a way that MVU estimates for the population parameters can be obtained.

1.2 Estimation of σ^2

Given a set of marker levels $\{y_{i,j}\}$ collected from a population with m individuals each of which has contributed n_i marker levels, there are a total of $N = \sum_{i=1}^{m} n_i$ marker levels to estimate the parameters μ, τ^2 and σ^2. The expected marker level μ_i can be estimated for individual i in the usual way with:

$$\overline{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{i,j} \tag{4}$$

We use the indices 'i' for individuals and 'j' for results within individuals throughout this work. For the i^{th} individual we define,

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (\overline{Y}_i - y_{i,j})}{n_i - 1} \tag{5}$$

which is an MVU estimator of σ for the subset $\{y_{i,1} \ldots y_{i,n_i}\}$ and which has an EV of $\frac{2\sigma^4}{n_i - 1}$. We then define,

$$S^2 = \frac{\sum_{i=1}^{m} (w_i \cdot S_i^2)}{\sum_{i=1}^{m} w_i} \tag{6}$$

where,

$$w_i = \frac{n_i - 1}{2} \tag{7}$$

which is an MVU estimator of σ^2 in all of $\{y_{i,j}\}$ and which has an EV of $\frac{2\sigma^4}{N - m}$

1.3 Estimation of τ^2

For each unique number of samples per individual, the population can be divided into groups of individuals who contribute the same number of results $D_k = \{i|n_i = k\}$ which has size m_k. For $i \in D_k$, it follows from the definition of the mean that $var(\overline{Y}_i - \mu_i) = \sigma^2/k$, and since $var(\mu - \mu_i) = \tau^2$, it follows that $\hat{T}_k^2 = var(\{\mu_i|i \in D_k\})$ using the standard variance estimator is an MVU estimate of $\tau^2 + \sigma^2/k$ for $i \in D_k$ which has an error variance of:

$$EV(\hat{T_k}^2) = \frac{2 \cdot \tau^4}{\beta_k^2 \cdot (m_k - 1)} \tag{8}$$

Since $\hat{T}_k^2$ is independent of S^2 , it follows that $T_k^2 = \hat{T_k}^2 - S^2/k$ is an unbiased estimate τ^2 with error variance equal to $EV(\hat{T_k}^2) + EV(S^2)/(k^2)$. Since the estimators $\{\hat{T}_k^2\}$ are independent, an unbiased estimate for τ^2 variance can be obtained using equation (2).

1.4 Estimation of μ

Since $\overline{Y}_i$ is an MVU estimator of μ_i with $EV(\overline{Y}) = \sigma^2/n_i$, it follows that $\overline{Y}_i$ is an unbiased estimator for μ with $Var(\overline{y} - \mu) = \tau^2 + (\sigma^2/n_i)$. We apply equation (2) with $w_i = \tau^2/\beta_n$ to arrive at an MVU estimate $\overline{Y}^*$ with error variance $\tau^2/\sum_i \beta_{n_i}$

In the case where n_i is identical for all individuals, expression $\overline{Y}^*$ reduces to the grand mean $\overline{Y} = \sum y_{i,j}/N$.