Winner's Curse Model

Consider data of the form $\mathcal{D}_{\text{obs}} = \big\{(\hat{\theta}_i, \sigma_i^2)\big\}$, where $\hat{\theta}_i$ is an estimated parameter and σ_i^2 is the sampling variance of $\hat{\theta}_i$, considered known. Suppose the data are generated from the following hierarchical model:

$$
\begin{gathered}
Z_i \sim \text{Bern}(\pi), \\
\theta_i|(Z_i = 1) \sim \delta_0, \\
\theta_i|(Z_i = 0) \sim N(0, \tau^2), \\
\hat{\theta}_i|\theta_i \sim N(\theta_i, \sigma_i^2),
\end{gathered}
$$

where π is the probability of belonging to the null component, δ_0 is a point-mass at zero, and τ^2 is a variance component. The marginal model for $\hat{\theta}_i$ is:

$$
f(\hat{\theta}_i|\pi, \tau^2) = \pi \cdot N(\hat{\theta}_i|0, \sigma_i^2) + (1-\pi) \cdot N(\hat{\theta}_i|0, \sigma_i^2 + \tau^2). \tag{1.0.1}
$$

1.1 Objective

Given $\mathcal{D}_{\text{obs}}$, the objective is to estimate π, τ^2, and $\mathbb{E}(\theta_i|\hat{\theta}_i)$.

1.2 Complete Data

Consider the complete data $\mathcal{D}^* = \big\{(\hat{\theta}_i, \sigma_i^2, Z_i)\big\}$, where Z_i is a latent binary indicator, with $Z_i = 1$ if θ_i originated from the null component. The complete data likelihood is:

$$
\begin{aligned}
L(\vartheta) &= \prod_{i=1}^{n} \mathbb{P}(\hat{\theta}_i, Z_i|\sigma_i^2, \tau^2) \\
&= \prod_{i=1}^{n} \big\{\pi \cdot N(\hat{\theta}_i; 0, \sigma_i^2)\big\}^{Z_i} \big\{(1-\pi) \cdot N(\hat{\theta}_i; 0, \sigma_i^2 + \tau^2)\big\}^{1-Z_i}
\end{aligned}
$$

1.2.1 Log Likelihood

The complete data log likelihood is:

$$
\begin{aligned}
\ell(\vartheta) \propto \sum_{i=1}^{n} & Z_i \left\{\ln(\pi) - \frac{1}{2}\ln(\sigma_i^2) - \frac{\hat{\theta}_i^2}{2\sigma_i^2}\right\} \\
& + (1-Z_i)\left\{\ln(1-\pi) - \frac{1}{2}\ln(\sigma_i^2 + \tau^2) - \frac{\hat{\theta}_i^2}{2(\sigma_i^2 + \tau^2)}\right\}.
\end{aligned} \tag{1.2.2}
$$

1.2.2 Score Equations

The complete data score equation for τ^2 is:

$$\mathcal{U}(\tau^2) = -\frac{1}{2}\sum_{i=1}^{n}(1 - Z_i)\left\{\frac{1}{\sigma_i^2 + \tau^2} - \frac{\hat{\theta}_i^2}{(\sigma_i^2 + \tau^2)^2}\right\}.$$

The complete data score equation for π is:

$$\mathcal{U}(\pi) = \sum_{i=1}^{n}\frac{Z_i}{\pi} - \frac{1 - Z_i}{1 - \pi}.$$

1.3 Expectation Maximization

1.3.1 EM Objective

The EM objective function is:

$$\begin{aligned} Q(\vartheta|\vartheta^{(r)}) = &\sum_{i=1}^{n}\gamma_i^{(r)}\left\{\ln(\pi) - \frac{1}{2}\ln(\sigma_i^2) - \frac{\hat{\theta}_i^2}{2\sigma_i^2}\right\} \\ &+ (1 - \gamma_i^{(r)})\left\{\ln(1 - \pi) - \frac{1}{2}\ln(\sigma_i^2 + \tau^2) - \frac{\hat{\theta}_i^2}{2(\sigma_i^2 + \tau^2)}\right\}, \end{aligned}$$

where:

$$\gamma_i^{(r)} = \mathbb{E}\big(Z_i|\mathcal{D}_{\text{obs}}, \vartheta^{(r)}\big) = \mathbb{P}(Z_i = 1|\hat{\theta}_i; \sigma_i^2, \vartheta^{(r)}).$$

1.3.2 EM Score Equations

For τ^2

The EM score equation for τ^2 is:

$$\begin{aligned} \mathcal{U}(\tau^2|\vartheta^{(r)}) &= \mathbb{E}\big\{\mathcal{U}(\tau^2)|\mathcal{D}_{\text{obs}}, \vartheta^{(r)}\big\} \\ &= -\frac{1}{2}\sum_{i=1}^{n}\big(1 - \gamma_i^{(r)}\big)\left\{\frac{1}{\sigma_i^2 + \tau^2} - \frac{\hat{\theta}_i^2}{(\sigma_i^2 + \tau^2)^2}\right\}. \end{aligned}$$

τ^2 must be obtained by solving $\mathcal{U}(\tau^2|\vartheta^{(r)}) \overset{\text{Set}}{=} 0$ numerically.

For π

The EM score equation for π is:

$$\mathcal{U}(\pi|\vartheta^{(r)}) = \sum_{i=1}^{n}\frac{\gamma_i^{(r)}}{\pi} - \frac{1 - \gamma_i^{(r)}}{1 - \pi}.$$

Define $n^{(r)} = \sum_{i=1}^{n}\gamma_i^{(r)}$, then solving $\mathcal{U}(\pi|\vartheta^{(r)}) \overset{\text{Set}}{=} 0$ gives:

$$\pi^{(r+1)} = \frac{n^{(r)}}{n}.$$

1.4 Responsibility

From Bayes' theorem:

$$\gamma_i^{(r)} = \frac{f(\hat{\theta}_i|Z_i = 1; \sigma_i^2, \vartheta^{(r)})\mathbb{P}(Z_i = 1; \vartheta^{(r)})}{f(\hat{\theta}_i|Z_i = 1; \sigma_i^2, \vartheta^{(r)})\mathbb{P}(Z_i = 1; \vartheta^{(r)}) + f(\hat{\theta}_i|Z_i = 0; \sigma_i^2, \vartheta^{(r)})\mathbb{P}(Z_i = 0; \vartheta^{(r)})},$$

which is expressible as:

$$\gamma_i^{(r)} = \frac{N(\hat{\theta}_i; 0, \sigma_i^2)\pi^{(r)}}{N(\hat{\theta}_i; 0, \sigma_i^2)\pi^{(r)} + N(\hat{\theta}_i; 0, \sigma_i^2 + \tau^{2,(r)})(1 - \pi^{(r)})}.$$

1.5 EM Iteration

Algorithm 1 Winner's Curse EM Iteration

Require: Parameter estimate $\hat{\theta}_i$ and sampling variance σ_i^2 for each observation.
Require: Initialize $\tau^{2,(0)}$ and $\pi^{(0)}$.

1: **repeat**
2: Calculate the responsibilities $(\gamma_i^{(r)})$.
3: Calculate the initial EM objective $Q_0^{(r)} = Q(\vartheta^{(r)}|\vartheta^{(r)})$.
4: Update τ^2 by solving $\mathcal{U}(\tau^2|\vartheta^{(r)}) \overset{\text{Set}}{=} 0$ to obtain $\tau^{2,(r+1)}$.
5: Update π via $\pi^{(r+1)} = n^{(r)}/n$.
6: Calculate the final EM objective $Q_1^{(r)} = Q(\vartheta^{(r+1)}|\vartheta^{(r)})$.
7: **until** $Q_1^{(r)} - Q_0^{(r)} < \epsilon$.

1.6 Posteriors

1.6.1 Distributions

The posterior of Z_i given $\hat{\theta}_i$ is the responsibility:

$$\mathbb{E}(Z_i|\hat{\theta}_i) = \mathbb{P}(Z_i = 1|\hat{\theta}_i) = \gamma_i.$$

The posterior density of the true parameter θ_i given $(\hat{\theta}_i, Z_i = 0)$ is:

$$\begin{aligned} f(\theta_i|\hat{\theta}_i, Z_i = 0) &\propto f(\hat{\theta}_i|\theta_i) \cdot f(\theta_i|Z_i = 0) \\ &\propto N(\hat{\theta}_i|\theta_i, \sigma_i^2) \cdot N(\theta_i|0, \tau^2) \\ &\propto \exp\left\{-\frac{(\hat{\theta}_i - \theta_i)^2}{2\sigma_i^2}\right\} \cdot \exp\left\{-\frac{\theta_i^2}{2\tau^2}\right\} \equiv \exp\left\{-\frac{1}{2}g(\theta_i)\right\}. \end{aligned}$$

Completing the square:

$$
\begin{aligned}
g(\theta_i) &= \frac{(\hat{\theta}_i - \theta_i)^2}{\sigma_i^2} + \frac{\theta_i^2}{\tau^2} \\
&= \frac{\hat{\theta}_i^2}{\sigma_i^2} - 2\frac{\hat{\theta}_i\theta_i}{\sigma_i^2} + \frac{\theta_i^2}{\sigma_i^2} + \frac{\theta_i^2}{\tau^2} \\
&\propto \frac{1}{\sigma_i^2\tau^2}\left(\tau^2\theta_i^2 + \sigma_i^2\theta_i^2 - 2\tau^2\hat{\theta}_i\theta_i\right) \\
&= \frac{\tau^2 + \sigma_i^2}{\sigma_i^2\tau^2}\left(\theta_i^2 - 2\frac{\tau^2}{\tau^2 + \sigma_i^2}\hat{\theta}_i\theta_i\right) \\
&\propto \left(\frac{\sigma_i^2\tau^2}{\tau^2 + \sigma_i^2}\right)^{-1}\left(\theta_i - \frac{\tau^2}{\tau^2 + \sigma_i^2}\hat{\theta}_i\right)^2.
\end{aligned}
$$

Therefore, $f(\theta_i|\hat{\theta}_i, Z_i = 0)$ is normal with mean and variance:

$$
\mathbb{E}(\theta_i|\hat{\theta}_i, Z_i = 0) = \frac{\tau^2}{\tau^2 + \sigma_i^2}\hat{\theta}_i, \qquad \mathbb{V}(\theta_i|\hat{\theta}_i, Z_i = 0) = \frac{\sigma_i^2\tau^2}{\tau^2 + \sigma_i^2}.
$$

Note that since the distribution of θ_i given $Z_i = 1$ is a point-mass at zero, the posterior distribution of θ_i given $(\hat{\theta}_i, Z_i = 1)$ remains a point-mass at zero.

1.6.2 Expectation

The posterior expectation of θ_i given $\hat{\theta}_i$ is:

$$
\begin{aligned}
\mathbb{E}(\theta_i|\hat{\theta}_i) &= \mathbb{E}\big\{\mathbb{E}(\theta_i|\hat{\theta}_i, Z_i)\big\} \\
&= \sum_{z\in\{0,1\}} \mathbb{E}(\theta_i|\hat{\theta}_i, Z_i = z)\cdot\mathbb{P}(Z_i = z|\hat{\theta}_i) \\
&= \mathbb{E}(\theta_i|\hat{\theta}_i, Z_i = 0)\cdot\mathbb{P}(Z_i = 0|\hat{\theta}_i) \\
&= \frac{\tau^2}{\tau^2 + \sigma_i^2}\hat{\theta}_i\cdot(1 - \gamma_i).
\end{aligned}
$$

The posterior expectation may be estimated by plugging in the final value of τ^2 and the responsibility from the EM algorithm:

$$
\hat{\mathbb{E}}(\theta_i|\hat{\theta}_i) = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \sigma_i^2}\cdot\hat{\theta}_i\cdot(1 - \hat{\gamma}_i).
$$