Empirical Bayes approach to Truth Discovery problems - Supplementary material

Tsviel Ben Shabat[1]   **Reshef Meir**[1]   **David Azriel**[1]

[1]Dept. of Industrial Engineering and Management, Technion—Israel Institute of Technology

1 DERIVING THE EMPIRICAL BAYES ESTIMATOR (EBE)

In Appendix A we present known results as a background for the Empirical Bayes approach.

On our path to deriving an Empirical Bayes Estimator (EBE) for $\vec{\mu}$ we first need to derive the appropriate Bayes estimator. Recall $\mathcal{L}(A(X), \vec{\mu})$ is a loss square euclidean loss function, i.e, $\mathcal{L}(A(X), \vec{\mu}) = \|\vec{\mu} - A(X)\|_{l2}^2$.

Bayesian Estimator The Bayesian Estimator minimizes the Bayes risk which we will soon present, but first, we introduce the notion of the Prior distribution.

Prior probability distribution The Bayesian framework assumes a known prior, this will later we will relax the known-prior-assumption and estimate its parameters. First, we assume $\vec{\mu} = (\mu_1, .., \mu_m)$ is normally distributed, that is, we assume

$$\mu_j \sim \mathcal{N}_0 = \mathcal{N}(\mu_0, \sigma_0^2)$$

For example we can think of a group of workers answering "what are the weights of the people is these images?", thus, the underlying ground truth are weights of people which are close to a normal distribution.

Observations $\vec{X} = (X_1, \ldots, X_m)$ Assuming a prior distribution over $\vec{\mu}$, (notice the difference from the preliminaries where $X_j \sim \mathcal{N}(\mu_j, \sigma^2)$), we now denote the conditional distribution $X_j | \mu_j \sim \mathcal{N}(\mu_j, \sigma^2)$

Posterior probability distribution $\mu_j | X_j$ can be viewed as an update for the assumed prior distribution after viewing the observations. It is well known that the Bayesian Estimator (which we will later show that it is better than the BLUE estimator shown in Theorem 2.1) for the square loss function is the posterior's mean, that is, $E_{\mu \sim N_0, \sigma}(\mu_j | X_j) = \min_{A(X_j)} E_{\mu \sim N_0, \sigma}(A(X_j) - \mu_j)^2$ where the latter expectation is under the Bayesian framework, that is $E_{\mu \sim N_0, \sigma}(A(X_j) - \mu_j)^2 = \int_\mu E_{\vec{\mu}, \vec{\sigma}}(A(X_j) - \mu)^2 d\mu$, hence, to estimate μ_j via a Bayesian Estimator we first need to calculate the posterior mean.

Theorem 1.1. *The posterior distribution is* $\mu_j | X_j \sim \mathcal{N}(\tilde{\mu}_j, \tilde{\sigma}^2)$,

$$\tilde{\mu}_j = X_j * \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} + \mu_0 * \frac{\sigma^2}{\sigma_0^2 + \sigma^2}, \; \tilde{\sigma}^2 = \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2}$$

Proof.

$$\mathcal{P}(\mu_j | X_j) = \frac{\mathcal{P}(X_j | \mu_j)\mathcal{P}(\mu_j)}{\mathcal{P}(X_j)} \propto \mathcal{P}(X_j | \mu_j)\mathcal{P}(\mu_j)$$

$$\propto e^{\frac{\frac{(X_j - \mu_j)^2}{\sigma^2} + \frac{(\mu_j - \mu_0)^2}{\sigma_0^2}}{-2}} \propto e^{\frac{\mu_j^2(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}) - 2\mu_j(\frac{X_j}{\sigma^2} + \frac{\mu_0}{\sigma_0^2})}{-2}} \propto e^{\frac{(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2})(\mu_j - \tilde{\mu})^2}{-2}}$$

Where the last proportional-to follows from completing-the-square technique and

$$\tilde{\mu}_j = \frac{X_j \sigma_0^2 + \mu_0 \sigma^2}{\sigma_0^2 + \sigma^2}$$

$\square$

1.1 THE EMPIRICAL BAYES ESTIMATOR (EBE)

We now relax the previous known-prior-assumption, this subsection is the empirical part in Empirical Bayes. We will show how to estimate the posterior mean using the observed data.

First we will derive the marginal distribution of X_j, that is, we evaluate the following expression:

$$\mathcal{P}_{\mu_j \sim \mathcal{N}_0,\sigma}(X_j) = \int_{\mu_j} \mathcal{P}_{\mu_j,\sigma}(X_j|\mu_j)\mathcal{P}(\mu_j)d\mu_j$$

But we already saw on (1.1) that $\mathcal{P}_{\mu_j,\sigma}(X_j|\mu_j)\mathcal{P}(\mu_j)$ is normally distributed, thus what is left is to calculate its mean and variance.

By law of total expectation

$$E_{\mu_j \sim N_0,\sigma}(X_j) = E_{\mu_j \sim N_0,\sigma}(E_{\mu_j,\sigma}(X_j|\mu_j)) = E_{\mu_j \sim N_0,\sigma}(\mu_j) = \mu_0$$

By law of total variance

$$Var_{\mu_j \sim N_0,\sigma}(X_j) = E_{\mu_j \sim N_0,\sigma}(Var_{\mu_j,\sigma}(X_j|\mu_j)) + Var_{\mu_j \sim N_0,\sigma}(E_{\mu_j,\sigma}(X_j|\mu_j)) = \sigma^2 + \sigma_0^2$$

Empirical Bayes Estimator   Similarly to the work *An introduction to Empirical Bayes Data Analysis* [George Casella, 1985], we can construct an estimate for $\tilde{\mu}_j$ at (1.1), denote $\bar{X} = \frac{1}{m}\sum_{j=1}^{m} X_j$

Proposition 1.2. $E_{\mu \sim N_0,\sigma}\left[\frac{\sigma^2 + \sigma_0^2}{\sum_{j=1}^{m}(X_j - \bar{X})^2}\right] = \frac{1}{m-3}$

Proof. Notice that the expectation is over the marginal of X_j and thus, the X_j's are i.i.d which means the expectation is of an inverse-chi-squared distribution with $m-1$ degrees of freedom. $\square$

An immediate corollaries of 1.2 are:

Corollary 1.2.1. $E_{\mu \sim N_0,\sigma}\left[\frac{(m-3)\sigma^2}{\sum_{j=1}^{m}(X_j - \bar{X})^2}\right] = \frac{\sigma^2}{\sigma^2 + \sigma_0^2}$

Corollary 1.2.2. *An Empirical Bayes Estimator is*

$$\phi_{\text{EB}}(\vec{X},\sigma) = \frac{(m-3)\sigma^2}{\sum_{j=1}^{m}(X_j - \bar{X})^2}\bar{X}\mathbf{1} + \left[1 - \frac{(m-3)\sigma^2}{\sum_{j=1}^{m}(X_j - \bar{X})^2}\right]\vec{X}$$

2 PROOFS

2.1 THEOREM 2.2 (CASELLA, 1985'S THEOREM)

In the AWG model with a single worker and $m > 3$ questions,

$$\mathcal{R}_{\vec{\mu},\sigma}(\phi_{\text{EB}}^{\sigma}) < \mathcal{R}_{\vec{\mu},\sigma}(\phi_I) \text{ for all } \vec{\mu} \in \mathbb{R}^m, \sigma \in \mathbb{R}^+. \tag{1}$$

Proof. Denote $\hat{B} = \frac{(m-3)\sigma^2}{S^2(\vec{X})}$

$$\begin{aligned}
\mathcal{R}_{\vec{\mu},\sigma}(\phi_{\text{EB}}) &= E_{\vec{\mu},\sigma}[\|\bar{X} + (1-\hat{B})(X - \bar{X}) - \mu\|^2] \\
&= \Sigma_{j=1}^{m} E_{\vec{\mu},\sigma}[(\bar{X} + (1-\hat{B})(X_j - \bar{X}) - \mu_j)^2] \\
&= \Sigma_{j=1}^{m} E_{\vec{\mu},\sigma}[(X_j - \mu_j - \hat{B}(X_j - \bar{X}))^2] \\
&= \Sigma_{j=1}^{m} E_{\vec{\mu},\sigma}[(X_j - \mu_j)^2] - 2E_{\vec{\mu},\sigma}[\hat{B}(X_j - \mu_j)(X_j - \bar{X})] + E_{\vec{\mu},\sigma}[(\hat{B}(X_j - \bar{X}))^2] \\
&= \mathcal{R}_{\vec{\mu},\sigma}(\phi_I) - [\Sigma_{j=1}^{m} 2E_{\vec{\mu},\sigma}[\hat{B}(X_j - \mu_j)(X_j - \bar{X})]] + \sigma^4(m-3)^2 E_{\vec{\mu},\sigma}\left[\frac{1}{S^2(\vec{X})}\right]
\end{aligned}$$

Lemma 2.1. *Stein's Lemma:*
Let $X \sim N(\mu, \sigma^2)$ and Let $g(X)$ be a function for which $E_{\vec{\mu},\sigma}[g(X)(X - \mu)]$ and $E_{\vec{\mu},\sigma}[\frac{d}{dx}g(X)]$ both exist, then
$$E_{\vec{\mu},\sigma}[g(X)(X - \mu)] = \sigma^2 E_{\vec{\mu},\sigma}[\frac{d}{dx}g(X)]$$

We now focus on the mixed term:
Denote $g(X_j) = \frac{X_j - \bar{X}}{S^2(\vec{X})}$

$$E_{\vec{\mu},\sigma}[\hat{B}(X_j - \mu_j)(X_j - \bar{X})] = \sigma^2(m-3)E_{\vec{\mu},\sigma}[\frac{X_j - \bar{X}}{S^2(\vec{X})}(X_j - \mu_j)]$$

$$= \sigma^2(m-3)E_{\vec{\mu},\sigma}[E_{\vec{\mu},\sigma}[\frac{X_j - \bar{X}}{S^2(\vec{X})}(X_j - \mu_j)|X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_m]]$$

$$= \sigma^2(m-3)E_{\vec{\mu},\sigma}[E_{\vec{\mu},\sigma}[g(X_j)(X_j - \mu_j)|X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_m]]$$

$$=_{stein} \sigma^4(m-3)E_{\vec{\mu},\sigma}[E_{\vec{\mu},\sigma}[\frac{d}{dX_j}g(X_j)|X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_m]]$$

$$= \sigma^4(m-3)E_{\mu,\sigma}[\frac{d}{dX_j}g(X_j)]$$

$$\frac{d}{dX_j}g(X_j) = \frac{(1 - \frac{1}{m})S^2(\vec{X}) - 2(X_j - \bar{X})^2}{\left(S^2(\vec{X})\right)^2}$$

The mixed term can now be rewritten as follows:

$$-\Sigma_{j=1}^m 2E_{\vec{\mu},\sigma}[\hat{B}(X_j - \mu_j)(X_j - \bar{X})] = -2\sigma^4(m-3)E_{\vec{\mu},\sigma}[\frac{(1 - \frac{1}{m})mS^2(\vec{X}) - 2S^2(\vec{X})}{\left(S^2(\vec{X})\right)^2}]$$

$$= -2\sigma^4(m-3)^2 E_{\vec{\mu},\sigma}[\frac{1}{S^2(\vec{X})}]$$

Summing up everything:

$$\mathcal{R}_{\vec{\mu},\sigma}(\phi_{\text{EB}}) =$$

$$= \mathcal{R}_{\vec{\mu},\sigma}(\phi_I) - 2\sigma^4(m-3)^2 E_{\vec{\mu},\sigma}[\frac{1}{S^2(\vec{X})}] + \sigma^4(m-3)^2 E_{\vec{\mu},\sigma}[\frac{1}{S^2(\vec{X})}]$$

$$= \mathcal{R}_{\vec{\mu},\sigma}(\phi_I) - \sigma^4(m-3)^2 E_{\vec{\mu},\sigma}[\frac{1}{S^2(\vec{X})}] \, \forall m > 3$$

$\square$

2.2 PROPOSITION 3.1

$A_B^\sigma(\boldsymbol{X})$ is a sufficient statistic for $\vec{\mu} = (\mu_1, .., \mu_m)$ under $\mathcal{P}_{\vec{\mu},\vec{\sigma}}$.

Proof. Since the regularity conditions hold for a multiplication of independent normal distributions, to prove sufficiency by the Fisher-Neyman factorization theorem, we only need to show that the model can be represented as a multiplication of a function $\tilde{h}$ of the observations and a function $\tilde{g}$ of a sufficient statistic and the unknown parameter as follows:

$$\mathcal{P}_{(\mu_1,..,\mu_m,\sigma_1,...,\sigma_n)}(X_{11}, .., X_{nm}) = \Pi_{j=1}^m \Pi_{i=1}^n \mathcal{P}_{\mu_j,\sigma_i}(X_{ij})$$

$$= \Pi_{j=1}^m \Pi_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2\sigma_i^2}(X_{ij} - \mu_j)^2}$$

$$= \Pi_{j=1}^m [C \cdot e^{-\frac{1}{2}\sum_{i=1}^n \frac{X_{ij}^2}{\sigma_i^2}}] * [e^{-\frac{1}{2}\mu_j^2 \sum_{i=1}^n \frac{1}{\sigma_i^2} + \mu_j \sum_{i=1}^n \frac{X_{ij}}{\sigma_i^2}}]$$

$$= \Pi_{j=1}^m h(X_{1j}, .., X_{nj})g(\delta_j(X), \mu_j)$$

$$= \tilde{h}(X_{11}, .., X_{nm})\tilde{g}(\delta(X), \mu)$$

where C is a constant.

$\square$

2.3 THEOREM 4.1

For any unbiased algorithm A, and $m > 3$,[1]

$$\mathcal{R}_{\vec{\mu}}(\text{EB}_{\text{A}}^{\psi}) < \mathcal{R}_{\vec{\mu}}(\text{A}) \text{ for all } \vec{\mu} \in \mathbb{R}^m$$

if and only if

$$2(m-3)\Sigma_{j=1}^{m}Cov\left(X_j^A, \frac{\psi(\boldsymbol{X})(X_j^A - \bar{X}^A)}{\|\vec{X}^A - \bar{X}^A\vec{1}\|^2}\right) - (m-3)^2 E_{\vec{\mu}}\left(\frac{(\psi(\boldsymbol{X}))^2}{\|\vec{X}^A - \bar{X}^A\vec{1}\|^2}\right) > 0. \tag{2}$$

For convenience we denote $\vec{X} = \vec{X}^A$

Proof.

$$\mathcal{R}_{\vec{\mu}}(\text{EB}_{\text{A}}^{\psi}) = E_{\vec{\mu}}[\|\bar{X} + (1 - \frac{\alpha\hat{\sigma}^2}{S^2(\vec{X})})(X - \bar{X}) - \mu\|^2]$$

$$= E_{\vec{\mu}}[\|(X - \mu)\|^2] - 2E_{\vec{\mu}}[\frac{\alpha\hat{\sigma}^2}{S^2(\vec{X})}(X - \mu)^T(X - \bar{X})] + E_{\vec{\mu}}[(\|\frac{\alpha\hat{\sigma}^2}{S^2(\vec{X})}(X - \bar{X})\|^2]$$

$$= \mathcal{R}_{\vec{\mu}}(\text{A}) - 2\alpha E_{\vec{\mu}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}(X - \mu)^T(X - \bar{X})] + \alpha^2 E_{\vec{\mu}}[\frac{\hat{\sigma}^4}{S^2(\vec{X})}]$$

Focusing on the mixed term:

$$E_{\vec{\mu}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}(X - \mu)^T(X - \bar{X})] = \sum_{i=1}^{m} E_{\vec{\mu}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}(X_j - \mu_j)(X_j - \bar{X})]$$

$$=_{E_{\vec{\mu}}(X_j)=\mu_j} \sum_{i=1}^{m} Cov\left(X_j - \mu_j, \frac{\hat{\sigma}^2(X_j - \bar{X})}{S^2(\vec{X})}\right)$$

$$= \sum_{i=1}^{m} Cov\left(X_j, \frac{\hat{\sigma}^2(X_j - \bar{X})}{S^2(\vec{X})}\right)$$

Therefore we get:

$$\mathcal{R}_{\vec{\mu}}(\text{A}) - \mathcal{R}_{\vec{\mu}}(\text{EB}_{\text{A}}^{\psi}) = 2\alpha \sum_{i=1}^{m} Cov\left(X_j, \frac{\hat{\sigma}^2(X_j - \bar{X})}{S^2(\vec{X})}\right) - \alpha^2 E_{\vec{\mu}}[\frac{\hat{\sigma}^4}{S^2(\vec{X})}]$$

$\square$

2.4 PROPOSITION 4.2

Denote $\psi(\boldsymbol{X}) = \hat{\sigma}^2$

Choosing $\alpha^* = \dfrac{\Sigma_{j=1}^{m}Cov\left(X_j^A, \frac{\hat{\sigma}^2(X_j - \bar{X})}{S^2(\bar{X}^A)}\right)}{E_{\vec{\mu}}\left(\frac{\hat{\sigma}^4}{S^2(\bar{X}^A)}\right)}$ minimizes $\mathcal{R}_{\vec{\mu}}(\text{EB}_{\text{A}}^{\psi})$.

Proof. Denote $C := Cov\left(X_j^A, \frac{\hat{\sigma}^2(X_j^A - \bar{X}^A)}{S^2(\vec{X}^A)}\right)$ and $E := E_{\vec{\mu}}\left(\frac{\hat{\sigma}^4}{S^2(\vec{X}^A)}\right)$. Then from Theorem 4.1 We wish to maximize the parabola $2\alpha C - \alpha^2 E$ which can be easily shown to maximized at $\alpha^* = \frac{C}{E}$. $\square$

---

[1]Since in this subsection we do not assume that the distribution of $\boldsymbol{X}$ follows the AWG model, we do not need a parameter for the individual competence. Other than that, all definitions remain the same.

2.5   THEOREM 4.3

Denote $\psi(\vec{X}) = \hat{\sigma}^2$ as an estimator of σ^2.

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\mathrm{EB}_A^\psi) = \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) + \frac{(m-3)^2}{m-1}\left(E_{\vec{\mu},\vec{\sigma}}[\frac{(\psi(\vec{X}^A))^2}{S^2(\vec{X}^A)}] - 2\sigma^2\left[E_{\vec{\mu},\vec{\sigma}}[\frac{\psi(\vec{X}^A)}{S^2(\vec{X}^A)}] + E_{\vec{\mu},\vec{\sigma}}\left[\frac{\Sigma_{j=1}^m \frac{d\psi(\vec{X}^A)}{dX_j^A}(X_j^A - \bar{X}^A)}{(m-3)S^2(\vec{X}^A)}\right]\right]\right)$$

(3)

For convenience we denote $\vec{X} = \vec{X}^A$

Proof. Denote $g(X) = \frac{\hat{\sigma}^2(X_j - \bar{X})}{S^2(\vec{X})}$

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\mathrm{EB}_A^\psi) = E_{\vec{\mu},\vec{\sigma}}[\|\bar{X} + (1-\hat{B})(X - \bar{X}) - \mu\|^2]$$

$$= \Sigma_{j=1}^m E_{\vec{\mu},\vec{\sigma}}[(X_j - \mu_j)^2] - 2E_{\vec{\mu},\vec{\sigma}}[\hat{B}(X_j - \mu_j)(X_j - \bar{X})] + E_{\vec{\mu},\vec{\sigma}}[(\hat{B}(X_j - \bar{X}))^2]$$

$$= \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) - 2(m-3)[\Sigma_{j=1}^m E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2(X_j - \mu_j)(X_j - \bar{X})}{S^2(\vec{X})}]] + (m-3)^2 E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^4}{S^2(\vec{X})}]$$

$$E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2(X_j - \mu_j)(X_j - \bar{X})}{S^2(\vec{X})}] = E_{\vec{\mu},\vec{\sigma}}[g(X)(X_j - \mu_j)]$$

$$=_{stein} \sigma^2 E_{\vec{\mu},\vec{\sigma}}[\frac{d}{dX_j}g(X)]$$

$$\frac{d}{dX_j}g(X) = \frac{d}{dX_j}\frac{\hat{\sigma}^2(X_j - \bar{X})}{S^2(\vec{X})}$$

$$= \frac{[(1 - \frac{1}{m})\hat{\sigma}^2 + \frac{d\hat{\sigma}^2}{dX_j}(X_j - \bar{X})]S^2(\vec{X}) - 2\hat{\sigma}^2(X_j - \bar{X})^2}{S^4}$$

$$= \frac{\hat{\sigma}^2[(1 - \frac{1}{m})S^2(\vec{X}) - 2(X_j - \bar{X})^2] + \frac{d\hat{\sigma}^2}{dX_j}(X_j - \bar{X})S^2(\vec{X})}{\left(S^2(\vec{X})\right)^2}$$

$$\Sigma_{j=1}^m E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2(X_j - \mu_j)(X_j - \bar{X})}{S^2(\vec{X})}] = \sigma^2(E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2\Sigma_{j=1}^m \frac{(1 - \frac{1}{m})S^2(\vec{X}) - 2(X_j - \bar{X})^2}{\left(S^2(\vec{X})\right)^2}]$$

$$+ E_{\vec{\mu},\vec{\sigma}}[\frac{S^2(\vec{X})\Sigma_{j=1}^m \frac{d\hat{\sigma}^2}{dX_j}(X_j - \bar{X})}{\left(S^2(\vec{X})\right)^2}])$$

$$= \sigma^2[(m-3)E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}] + E_{\vec{\mu},\vec{\sigma}}[\frac{\Sigma_{j=1}^m \frac{d\hat{\sigma}^2}{dX_j}(X_j - \bar{X})}{S^2(\vec{X})}]]$$

Also mind that:

$$E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2(X - \mu)^T(X - \bar{X})}{S^2(\vec{X})}] = cov(\hat{\sigma}^2, \frac{(X - \mu)^T(X - \bar{X})}{S^2(\vec{X})}) + E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]E_{\vec{\mu},\vec{\sigma}}(\frac{(X - \mu)^T(X - \bar{X})}{S^2(\vec{X})}]$$

$$=_{stein} cov(\hat{\sigma}^2, \frac{(X - \mu)^T(X - \bar{X})}{S^2(\vec{X})}) + (m-3)\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]E_{\vec{\mu},\vec{\sigma}}[\frac{1}{S^2(\vec{X})}]$$

$$E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}] = cov(\hat{\sigma}^2, \frac{1}{S^2(\vec{X})}) + E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]E_{\vec{\mu},\vec{\sigma}}[\frac{1}{S^2(\vec{X})}]$$

And therefore we get that:

$$\sigma^2 E_{\vec{\mu},\vec{\sigma}}\big[\frac{\sum_{j=1}^m \frac{d\hat{\sigma}^2}{dX_j}(X_j - \bar{X})}{S^2(\vec{X})}]] = cov(\hat{\sigma}^2, \frac{(X-\mu)^T(X-\bar{X})}{S^2(\vec{X})})$$

$$+ (m-3)\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]E_{\vec{\mu},\vec{\sigma}}[\frac{1}{S^2(\vec{X})}] - \sigma^2(m-3)E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}]$$

$$= cov(\hat{\sigma}^2, \frac{(X-\mu)^T(X-\bar{X})}{S^2(\vec{X})})$$

$$+ (m-3)\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]E_{\vec{\mu},\vec{\sigma}}[\frac{1}{S^2(\vec{X})}] - \sigma^2(m-3)(cov(\hat{\sigma}^2, \frac{1}{S^2(\vec{X})}) + E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]E_{\vec{\mu},\vec{\sigma}}[\frac{1}{S^2(\vec{X})}])$$

$$= cov(\hat{\sigma}^2, \frac{(X-\mu)^T(X-\bar{X})}{S^2(\vec{X})}) - \sigma^2(m-3)(cov(\hat{\sigma}^2, \frac{1}{S^2(\vec{X})}))$$

$$= cov(\hat{\sigma}^2, \frac{(X-\mu)^T(X-\bar{X}) - \sigma^2(m-3)}{S^2(\vec{X})})$$

Plugging everything:

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\hat{\mu},\mu) = \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) - 2\sigma^2(m-3)\Big((m-3)E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}] + E_{\vec{\mu},\vec{\sigma}}\big[\frac{\sum_{j=1}^m \frac{d\hat{\sigma}^2}{dX_j}(X_j-\bar{X})}{S^2(\vec{X})}\big]\Big)$$

$$+ (m-3)^2 E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^4}{S^2(\vec{X})}]$$

$$= \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) + (m-3)^2\Big(E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^4}{S^2(\vec{X})}] - 2\sigma^2\Big[E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}] + E_{\vec{\mu},\vec{\sigma}}\big[\frac{\sum_{j=1}^m \frac{d\hat{\sigma}^2}{dX_j}(X_j-\bar{X})}{(m-3)S^2(\vec{X})}\big]\Big]\Big)$$

$$= \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) + (m-3)^2(E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^4}{S^2(\vec{X})}] - 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\frac{\hat{\sigma}^2}{S^2(\vec{X})}])$$

$$- 2(m-3)cov\Big(\hat{\sigma}^2, \frac{(X-\mu)^T(X-\bar{X}) - \sigma^2(m-3)}{S^2(\vec{X})}\Big)$$

$$\square$$

Denote by $\bar{Y} := \frac{1}{m}\sum_j Y_j$ and $S^2(Y) := \frac{1}{m-1}\sum_j (Y_j - \bar{Y})^2$ its *mean* and its *sample variance*, respectively which yields the result.

2.6   COROLLARY 4.3.2

Under the Normal model, if ψ is mean-adjusted then

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(EB_A^\psi) < \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A)$$

$$+ \frac{(m-3)^2}{m-1}(E_{\vec{\mu},\vec{\sigma}}[\frac{(\psi(\vec{X}^A))^2}{S^2(\vec{X}^A)}] - 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\frac{\psi(\vec{X}^A)}{S^2(\vec{X}^A)}])$$

Proof. From theorem 4.3:

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(EB_A^\psi) = \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) + \frac{(m-3)^2}{m-1}\Big(E_{\vec{\mu},\vec{\sigma}}[\frac{(\psi(\vec{X}^A))^2}{S^2(\vec{X}^A)}] - 2\sigma^2\Big[E_{\vec{\mu},\vec{\sigma}}[\frac{\psi(\vec{X}^A)}{S^2(\vec{X}^A)}] + E_{\vec{\mu},\vec{\sigma}}\big[\frac{\sum_{j=1}^m \frac{d\psi(\vec{X}^A)}{dX_j^A}(X_j^A - \bar{X}^A)}{(m-3)S^2(\vec{X}^A)}\big]\Big]\Big)$$

Under the assumptions made it is easy to see that:

$$E_{\vec{\mu},\vec{\sigma}}\Big[\frac{\sum_{j=1}^{m}\frac{d\psi(\vec{X}^A)}{dX_j^A}(X_j^A-\bar{X}^A)}{(m-3)S^2(\vec{X}^A)}\Big]\Big] = E_{\vec{\mu},\vec{\sigma}}\Big[\frac{\sum_{j=1}^{m}\frac{d\psi(\vec{X}^A)}{dX_j^A}(X_j^A-\bar{X}^A)}{(m-3)S^2(\vec{X}^A)}\Big]1_{X_j^A\leq\bar{X}^A}\Big]$$

$$+ E_{\vec{\mu},\vec{\sigma}}\Big[\frac{\sum_{j=1}^{m}\frac{d\psi(\vec{X}^A)}{dX_j^A}(X_j^A-\bar{X}^A)}{(m-3)S^2(\vec{X}^A)}\Big]1_{X_j^A>\bar{X}^A}\Big]$$

$$> 0$$

And the result immediately follows. $\qquad\square$

2.7 ADDITIONAL COROLLARIES

Corollary 2.1.1. *If $\hat{\sigma}^2$ is independent of X_j^A ($\frac{d\hat{\sigma}^2}{dX_j^A}=0$) $\forall j$ then directly from Theorem 4.3 we get that:*

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_A^{\psi}) = \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A) + E_{\vec{\mu},\vec{\sigma}}\Big[\frac{1}{S^2(\vec{X}^A)}\Big](m-3)^2(E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4]-2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2])$$

and $\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_A^{\psi}) < \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A)$ if

$$\frac{E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4]}{E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]} < 2\sigma^2 \tag{4}$$

Corollary 2.1.2. *Assume $\hat{\sigma}^2$ is independent of X_j $\forall j$ and that $\exists\epsilon>0$ $|\hat{\sigma}^2-\sigma^2|<\epsilon$ w.p 1 then $\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_A^{\psi}) < \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A)$ if*

$$\epsilon \in (0,\sigma^2) \tag{5}$$

Proof. From corollary 2.1.1, we found that $E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4] < 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]$ and therefore:

$$E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4] < (\sigma^2+\epsilon)E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2] < 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]$$
$$(\sigma^2+\epsilon) < 2\sigma^2$$
$$0 < \epsilon < \sigma^2$$

$\qquad\square$

Corollary 2.1.3. *Assume $\hat{\sigma}^2$ is independent of X_j $\forall j$ and that $\exists\epsilon>0$ $|\hat{\sigma}^2-\sigma^2|<\epsilon$ w.p δ and $\exists B$ s.t $\hat{\sigma}^2 < B$ w.p 1 then $\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_A^{\psi}) < \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A)$ $\forall\mu\in\mathbb{R}^m, m>3$ if*

$$\begin{cases} B & < \frac{\delta}{1-\delta}\sigma^4 \\ \epsilon & \in \big(0, -2\sigma^2 + \sqrt{5\sigma^4 + B(1-\frac{1}{\delta})}\big) \end{cases} \tag{6}$$

We know from corollary 2.1.1, under the norml model, when the competence estimator is independent of the observations X, then EBE dominates BLUE if $E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4] < 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2]$.

Denote the event $G = \{|\sigma^2-\hat{\sigma}^2|<\epsilon\}$

$$E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4] = E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4|G]\delta + E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^4|G^c](1-\delta)$$
$$\leq (\sigma^2+\epsilon)^2\delta + B(1-\delta)$$
$$2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2] = 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2|G]\delta + 2\sigma^2 E_{\vec{\mu},\vec{\sigma}}[\hat{\sigma}^2|G^c]$$
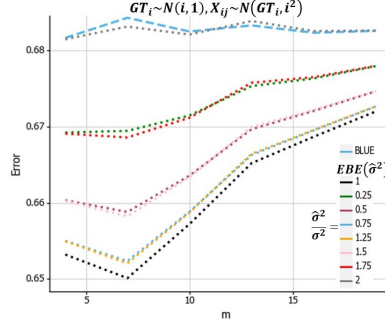$$> 2\sigma^2(\sigma^2-\epsilon)\delta$$

Figure 1: EBE for 5 aggregated workers with biased $\hat{\sigma}^2$ vs BLUE; each data point is a 100,000 iteration average, each iteration includes new GT and new workers' responses

Therefore it is sufficient to require that:

$$(\sigma^2 + \epsilon)^2 \delta + B(1 - \delta) < 2\sigma^2(\sigma^2 - \epsilon)\delta$$

$$0 < -\epsilon^2 - 4\epsilon\sigma^2 + \sigma^4 + B(1 - \frac{1}{\delta})$$

Which is a parabola of ϵ with roots $\epsilon = -2\sigma^2 \pm \sqrt{5\sigma^4 + B(1 - \frac{1}{\delta})}$. Notice that $B(1 - \frac{1}{\delta}) < 0$ and thus, we require that $2\sigma^2 < \sqrt{5\sigma^4 + B(1 - \frac{1}{\delta})}$ with simple algebra we derive the following conditions:

$$\begin{cases} B & < \frac{\delta}{1-\delta}\sigma^4 \\ \epsilon & \in \left(0, -2\sigma^2 + \sqrt{5\sigma^4 + B(1 - \frac{1}{\delta})}\right) \end{cases}$$

Notice when $\delta \to 1$ we get that $\epsilon \in (0, \sigma^2(\sqrt{5} - 2))$, this upper bound is less than one forth of the case when $\delta = 1$, this stricter result is due to the bounds we had to use to derive it.

3 DETERMINISTIC ESTIMATORS FOR MULTIPLE WORKERS

On this model we assume some oracle guessed and told us all of the different $\hat{\sigma}_i^2$ and thus, we treat them as constants (i.e. independent of the data X), we would like to know how close the oracle has to be to the actual competences such that EBE would still have lower risk than estimated BLUE by some algorithm A, i.e $\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_A^\psi) < \mathcal{R}_{\vec{\mu},\vec{\sigma}}(A)$.

Let $X_{ij} \sim \mathcal{N}(\mu_j, \sigma_i^2)$ $i = 1, .., n$ $j = 1, .., m$ and denote $\hat{\sigma}_i^2$ as an estimator which was somehow estimated for $\sigma_i^2, \forall i$. Notice that the BLUE estimator (BLUE-aggregated worker) for μ_j is:

$$\hat{X}_j^A = (\Sigma_{i=1}^n \frac{1}{\hat{\sigma}_i^2})^{-1} \Sigma_{i=1}^n \frac{X_{ij}}{\hat{\sigma}_i^2}$$

Then, notice that $\hat{X}_j^A$ is a linear combination of independent normal random variables therefore normal, i.e, $\hat{X}_j \sim N(\mu_j, \sigma^2)$ where:

$$E_{\vec{\mu},\vec{\sigma}}(\hat{X}_j^A) = E_{\vec{\mu},\vec{\sigma}}((\Sigma_{i=1}^n \frac{1}{\hat{\sigma}_i^2})^{-1} \Sigma_{i=1}^n \frac{X_{ij}}{\hat{\sigma}_i^2}) = \mu_j$$

$$Var_{\vec{\mu},\vec{\sigma}}(\hat{X}_j^A) = (\Sigma_{i=1}^n \frac{1}{\hat{\sigma}_i^2})^{-2} \Sigma_{i=1}^n \frac{\sigma^2}{\hat{\sigma}_i^4}$$

Therefore, we can reduce this case to the case of single worker where:

$$\hat{\sigma}^2 = (\Sigma_{i=1}^n \frac{1}{\hat{\sigma}_i^2})^{-2} \Sigma_{i=1}^n \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^4} = (\Sigma_{i=1}^n \frac{1}{\hat{\sigma}_i^2})^{-1}$$

$$\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_A^\psi)_j = \hat{\bar{X}}^A + [1 - \frac{(m-3)\hat{\sigma}^2}{S^2(\vec{X}^A)}](\hat{X}_j^A - \hat{\bar{X}}^A)$$

$$S^2(\vec{X}^A) = \sum_{j=1}^m (\hat{X}_j^A - \hat{\bar{X}}^A)^2$$

Proposition 3.1. *Under the Oracle model* $\mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{EB}_{\text{A}}^{\psi}) < \mathcal{R}_{\vec{\mu},\vec{\sigma}}(\text{A}) \ \forall \mu \in \mathbb{R}^m, m > 3$ *if:*

$$\hat{\sigma}^2 < 2\sigma^2 \tag{7}$$

Proof. Since we assumed an Oracle model (constant guesses of σ^2) and showed that $\hat{X}^A$ is following a normal distribution applying corollary 2.1.1 yields the result. $\qquad\square$