1 Model in Ash

The model considered in Ash is for each i,

$$x_i \sim \mathcal{N}(\alpha_i, y_i^2). \tag{1}$$

Note that in a PoissonBinomial model, $x_i = \hat{\alpha}_i$ and $y_i = se(\hat{\alpha}_i)$ are estimate (MLE) for $\alpha_i = logit(p_i)$ and its standard error which are estimated by using a glm function in R. Ash considers a mixture of normal distributions as a prior on α_i. Specifically, for each i,

$$\alpha_i \mid \pi, \sigma^2 \sim \sum_{m=1}^{M} \pi_m \mathcal{N}(0, \sigma_m^2), \tag{2}$$

where $\pi = (\pi_1, \ldots, \pi_M)$ are the mixture proportions which are constrained to be non-negative and sum to one, and $\sigma^2 = (\sigma_1^2, \ldots, \sigma_M^2)$ are the variances for each normal distribution. For now, Ash assumes that the Gaussian means are zero, that the variances vector σ^2 is known, and estimates π by using an empirical Bayes procedure - that is we find the maximum-likelihood estimator $\hat{\pi}_{MLE}$.

2 EM algorithm in Ash

The MLE for π can be obtained by using the following EM algorithm. Let $D_i = (x_i, y_i)$ and $D = (D_1, \ldots, D_n)$. Consider unobserved latent variables $Z = (Z_1, \ldots, Z_n)$, where $Z_i \in \{1, \ldots, M\}$ and $\mathsf{P}(Z_i = m) = \pi_m$. Then, a complete data likelihood can be written

$$\begin{aligned}
\mathsf{P}(D, Z \mid \pi) &= \prod_{i=1}^{n} \mathsf{P}(D_i, Z_i \mid \pi) && (3)\\
&= \prod_{i=1}^{n} \prod_{m=1}^{M} \mathsf{P}(D_i, Z_i = m \mid \pi)^{I(Z_i = m)} && (4)\\
&= \prod_{i=1}^{n} \prod_{m=1}^{M} [\mathsf{P}(D_i \mid Z_i = m, \pi)\pi_m]^{I(Z_i = m)}, && (5)
\end{aligned}$$

yielding a log likelihood

$$\log \mathsf{P}(D, Z \mid \pi) = \sum_{i=1}^{n} \sum_{m=1}^{M} I(Z_i = m)\Big[\log \mathsf{P}(D_i \mid Z_i = m) + \log \pi_m\Big]. \tag{6}$$

We denote by π^l the vector of probabilities at step l of the EM algorithm. In each step, we update the vector, i.e. compute in the l-th step π^{l+1} from π^l and the data.

E-step: For each i and m,

$$\begin{aligned}
A^l_{im} \equiv \mathsf{P}(Z_i = m \mid D_i, \pi^l) &= \frac{\mathsf{P}(Z_i = m, D_i \mid \pi^l)}{\sum_{n=1}^M \mathsf{P}(Z_i = n, D_i \mid \pi^l)} && (7)\\
&= \frac{\pi^l_m \mathsf{P}(D_i \mid Z_i = m)}{\sum_{n=1}^M \pi^l_n \mathsf{P}(D_i \mid Z_i = n)} && (8)\\
&= \frac{\pi^l_m \mathrm{BF_i}(\sigma^2_\mathrm{m})}{\sum_{n=1}^M \pi^l_n \mathrm{BF_i}(\sigma^2_\mathrm{n})}, && (9)
\end{aligned}$$

where

$$\mathrm{BF_i}(\sigma^2_\mathrm{m}) = \frac{\mathsf{P}(\mathrm{D_i} \mid \mathrm{Z_i} = \mathrm{m})}{\mathsf{P}(\mathrm{D_i} \mid \alpha_\mathrm{i} = 0)}. \tag{10}$$

M-step: Find the parameters π which maximizes $\mathrm{E}_{Z|D,\pi^l}[\log \mathsf{P}(D, Z \mid \pi)]$.

$$\begin{aligned}
\pi^{l+1} &= \operatorname*{argmax}_\pi \mathrm{E}_{Z|D,\pi^l}[\log \mathsf{P}(D, Z \mid \pi)] && (11)\\
&= \operatorname*{argmax}_\pi \sum_{i=1}^n \sum_{m=1}^M A^l_{im}[\log \mathrm{BF_{im}} + \log \pi_m] && (12)\\
&\equiv \operatorname*{argmax}_\pi Q(\pi \mid \pi^l), && (13)
\end{aligned}$$

where we used the fact that $\log \mathsf{P}(D_i \mid \alpha_i = 0)$ is constant. For each $m = 1, \ldots, M-1$,

$$\begin{aligned}
\frac{\partial Q(\pi \mid \pi^l)}{\partial \pi_m} &= \sum_{i=1}^n [\frac{A^l_{im}}{\pi_m} + \frac{-A^l_{iM}}{\pi_M}] && (14)\\
&= \sum_{i=1}^n [\frac{A^l_{im}\pi_M - A^l_{iM}\pi_m}{\pi_m \pi_M}] && (15)\\
&= \frac{\pi_M \sum_{i=1}^n A^l_{im} - \pi_m \sum_{i=1}^n A^l_{iM}}{\pi_m \pi_M}, && (16)
\end{aligned}$$

where

$$A^l_{iM} = 1 - (A^l_{i1} + \ldots, +A^l_{i(M-1)}), \tag{17}$$
$$\pi_M = 1 - (\pi_1 + \ldots, +\pi_{M-1}). \tag{18}$$

Then,

$$\pi_M \sum_{i=1}^{n} A_{im}^l = \pi_m \sum_{i=1}^{n} A_{iM}^l \quad \text{for} \quad m = 1, \ldots, M-1. \tag{19}$$

Summing $M-1$ equations in (19) leads to

$$\pi_M [\sum_{i=1}^{n} \sum_{m=1}^{M-1} A_{im}^l] = (1 - \pi_M) \sum_{i=1}^{n} A_{iM}^l. \tag{20}$$

Then,

$$\pi_M = \frac{\sum_{i=1}^n A_{iM}^l}{\sum_{i=1}^n \sum_{m=1}^M A_{im}^l} \tag{21}$$

$$= \frac{\sum_{i=1}^n A_{iM}^l}{n}, \tag{22}$$

and for each $m = 1, \ldots, M-1$,

$$\pi_m = \frac{\sum_{i=1}^n A_{im}^l}{n}. \tag{23}$$

To implement the EM algorithm we need an explicit expression for the Bayes Factor in eq. (10). We use an approximate likelihood approach to approximate this factor, shown in the next two paragraphs.

2.1 Likelihood approximation in a PoissonBinomial Model

Under a PoissonBinomial model, a log likelihood function for $\alpha_i = logit(p_i)$ can be written as

$$\mathcal{LL}(\alpha_i) = \log P(D_i \mid \alpha_i) \tag{24}$$

$$= \log \prod_j \binom{n_{ij}}{x_{ij}} \left(\frac{1}{1+e^{-\alpha_i}}\right)^{x_{ij}} \left(\frac{e^{-\alpha_i}}{1+e^{-\alpha_i}}\right)^{n_{ij}-x_{ij}}. \tag{25}$$

Taking three elements in Taylor series of $\mathcal{LL}(\alpha_i)$ about a MLE $\hat{\alpha}_i$, $\mathcal{LL}(\alpha)$ can be approximated by

$$\mathcal{LL}(\alpha_i) \approx \mathcal{LL}(\hat{\alpha}_i) + \frac{\mathcal{LL}''(\hat{\alpha}_i)(\alpha_i - \hat{\alpha}_i)^2}{2} \tag{26}$$

$$\approx \mathcal{LL}(\hat{\alpha}_i) - \frac{(\alpha_i - \hat{\alpha}_i)^2}{2se(\hat{\alpha}_i)^2}. \tag{27}$$

Then, a likelihood function for α_i, $\mathcal{L}(\alpha_i)$ can be approximated by

$$\mathcal{L}(\alpha_i) = exp[\mathcal{LL}(\alpha_i)] \propto \phi(\alpha_i; \hat{\alpha}_i, se(\hat{\alpha}_i)^2) \tag{28}$$

where ϕ is the Gaussian density function, $\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp[\frac{-(x-\mu)^2}{2\sigma^2}]$.

2.2 BF and posterior prob approximation in Ash

This section describes derivations for the approximate Bayes Factor (ABF) and posterior on α_i when $Z_i = m$ in the prior from equation (2).

$$
\begin{aligned}
\mathsf{P}(D_i \mid Z_i = m) &= \int \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)}{\sqrt{2\pi se(\hat{\alpha}_i)^2}} \exp[-\frac{(\alpha_i - \hat{\alpha}_i)^2}{2se(\hat{\alpha}_i)^2}] \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp[-\frac{\alpha_i^2}{2\sigma_m^2}] \mathrm{d}\alpha_i && (29)\\
&= \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)}{2\pi\sqrt{se(\hat{\alpha}_i)^2\sigma_m^2}} \int \exp[-\frac{\sigma_m^2 (\alpha_i - \hat{\alpha}_i)^2 + se(\hat{\alpha}_i)^2\alpha_i^2}{2se(\hat{\alpha}_i)^2\sigma_m^2}] \mathrm{d}\alpha_i && (30)\\
&= \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)}{2\pi\sqrt{se(\hat{\alpha}_i)^2\sigma_m^2}} \int \exp[-\frac{(\sigma_m^2 + se(\hat{\alpha}_i)^2)\,\alpha_i^2 - 2\sigma_m^2\hat{\alpha}_i\alpha_i + \sigma_m^2\hat{\alpha}_i^2}{2se(\hat{\alpha}_i)^2\sigma_m^2}] \mathrm{d}\alpha_i && (31)\\
&= \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)}{2\pi\sqrt{se(\hat{\alpha}_i)^2\sigma_m^2}} \int \exp[-\frac{(\sigma_m^2 + se(\hat{\alpha}_i)^2)\,[\alpha_i - \frac{\sigma_m^2\hat{\alpha}_i}{\sigma_m^2 + se(\hat{\alpha}_i)^2}]^2 + \frac{\sigma_m^2\hat{\alpha}_i^2 se(\hat{\alpha}_i)^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}}{2se(\hat{\alpha}_i)^2\sigma_m^2}] \mathrm{d}\alpha_i && (32)\\
&= \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)\sqrt{2\pi\frac{\sigma_m^2 se(\hat{\alpha}_i)^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}}}{2\pi\sqrt{se(\hat{\alpha}_i)^2\sigma_m^2}} \exp[-\frac{\hat{\alpha}_i^2}{2(se(\hat{\alpha}_i)^2 + \sigma_m^2)}] && (33)\\
&= \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)}{\sqrt{2\pi(\sigma_m^2 + se(\hat{\alpha}_i)^2)}} \exp[-\frac{\hat{\alpha}_i^2}{2(se(\hat{\alpha}_i)^2 + \sigma_m^2)}], && (34)
\end{aligned}
$$

where $C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)$ is a constant depending on $\hat{\alpha}_i, se(\hat{\alpha}_i)^2$ but does not depend on α_i. In fact $C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2) = e^{\mathcal{LL}(\hat{\alpha}_i)}\sqrt{2\pi se(\hat{\alpha}_i)^2}$.

and

$$
\mathsf{P}(D_i \mid \alpha_i = 0) = \frac{C(\hat{\alpha}_i, se(\hat{\alpha}_i)^2)}{\sqrt{2\pi se(\hat{\alpha}_i)^2}} \exp[-\frac{\hat{\alpha}_i^2}{2se(\hat{\alpha}_i)^2}]. \qquad (35)
$$

Then, ABF can be written as

$$
\begin{aligned}
\mathrm{BF_i}(\sigma_\mathrm{m}^2) &= \frac{\mathsf{P}(D_i \mid Z_i = m)}{\mathsf{P}(D_i \mid \alpha_i = 0)} && (36)\\
&= \sqrt{\frac{se(\hat{\alpha}_i)^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}} \exp[\frac{\hat{\alpha}_i^2}{2se(\hat{\alpha}_i)^2} \frac{\sigma_m^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}] && (37)\\
&= \sqrt{\lambda} \exp[T^2(1-\lambda)/2], && (38)
\end{aligned}
$$

where

$$
\begin{aligned}
\lambda &= \frac{se(\hat{\alpha}_i)^2}{se(\hat{\alpha}_i)^2 + \sigma_m^2}, && (39)\\
T &= \frac{\hat{\alpha}_i}{se(\hat{\alpha}_i)}. && (40)
\end{aligned}
$$

And a posterior on α_i is

$$\begin{aligned}
\mathsf{P}(\alpha_i \mid D_i, Z_i = m) &\propto \mathcal{L}(\alpha_i)\mathsf{P}(\alpha_i \mid Z_i) && (41)\\
&\propto \exp[-\frac{(\alpha_i - \hat{\alpha}_i)^2}{2se(\hat{\alpha}_i)^2}] \exp[-\frac{\alpha_i^2}{2\sigma_m^2}] && (42)\\
&\propto \exp[-\frac{(\sigma_m^2 + se(\hat{\alpha}_i)^2)\,[\alpha_i - \frac{\sigma_m^2 \hat{\alpha}_i}{\sigma_m^2 + se(\hat{\alpha}_i)^2}]^2}{2se(\hat{\alpha}_i)^2 \sigma_m^2}], && (43)
\end{aligned}$$

leading to

$$\alpha_i \mid D_i, Z_i = m \sim \mathcal{N}(\frac{\sigma_m^2 \hat{\alpha}_i}{\sigma_m^2 + se(\hat{\alpha}_i)^2}, \frac{\sigma_m^2 se(\hat{\alpha}_i)^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}). \quad (44)$$

$$\begin{aligned}
Pr(\alpha_i \mid D_i) &= \sum_{m=1}^{M} Pr(Z_i = m|D_i) Pr(\alpha_i \mid D_i, Z_i = m) && (45)\\
&= \sum_{m=1}^{M} \frac{\pi_m Pr(D_i|Z_i = m)}{\sum_n \pi_n Pr(D_i|Z_i = n)} Pr(\alpha_i \mid D_i, Z_i = m) && (46)\\
&= \sum_{m=1}^{M} A_{im}^l \mathcal{N}(\frac{\sigma_m^2 \hat{\alpha}_i}{\sigma_m^2 + se(\hat{\alpha}_i)^2}, \frac{\sigma_m^2 se(\hat{\alpha}_i)^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}) && (47)
\end{aligned}$$

2.3 EM-Algorithm Description

We can now write explicitly the EM-algorithm:
Algorithm: EM-Ash
Input: Data $D = (\hat{\alpha}_i, se(\hat{\alpha}_i)^2), i = 1, .., n$
Output: $\hat{\pi}_{MLE}$

1. Initialize $\pi_m^1 = \frac{1}{M}$ for $m = 1, .., M$. Set $l = 1$.
2. While $||\pi^{l+1} - \pi^l|| > \epsilon$ (i.e. repeat until π^l value converges)
   (a) E-Step: For each $i = 1, .., n$ set

   $$\mathrm{BF_i}(\sigma_\mathrm{m}^2) = \sqrt{\frac{se(\hat{\alpha}_i)^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}} \exp[\frac{\hat{\alpha}_i^2}{2se(\hat{\alpha}_i)^2} \frac{\sigma_m^2}{\sigma_m^2 + se(\hat{\alpha}_i)^2}] \quad (48)$$

   For each $i = 1, .., n$ and $m = 1, .., M$ set

   $$A_{im}^l = \frac{\pi_m^l \mathrm{BF_i}(\sigma_\mathrm{m}^2)}{\sum_{n=1}^{M} \pi_n^l \mathrm{BF_i}(\sigma_\mathrm{n}^2)} \quad (49)$$

(b) M-Step: For each $m = 1, .., M$, set

$$\pi_m^{l+1} \quad = \quad \frac{\sum_{i=1}^n A_{im}^l}{n}. \tag{50}$$

set $l = l + 1$.

3. Output $\hat{\pi}_{MLE} = \pi^l$.

3 Mean and Variance for a mixture of distributions

We will describe how to get mean and variance for a mixture of distributions in a general context. Consider a mixture of distributions whose mean and variance are given. Specifically,

$$Y \mid \pi, \mu, \sigma^2 \sim \sum_{m=1}^{M} \pi_m D(\mu_m, \sigma_m^2), \tag{51}$$

where $\pi = (\pi_1, \ldots, \pi_M)$ are the mixture proportions which are constrained to be non-negative and sum to one and $\mu = (\mu_1, \ldots, \mu_M)$ and $\sigma^2 = (\sigma_1^2, \ldots, \sigma_M^2)$ are the means and variances for each distribution. This distribution can be written as

$$Y \mid Z = m \quad \sim \quad D(\mu_m, \sigma_m^2) \tag{52}$$

where a latent variable $Z \in \{1, \ldots, M\}$ has a distribution $\mathsf{P}(Z = m) = \pi_m$. Then,

$$\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}(\mathrm{E}(Y \mid Z)) && (53)\\
&= \sum_{m=1}^{M} \pi_m \mu_m, && (54)\\
\mathrm{E}(Y^2) &= \mathrm{E}(\mathrm{E}(Y^2 \mid Z)) && (55)\\
&= \sum_{m=1}^{M} \pi_m (\mu_m^2 + \sigma_m^2), && (56)\\
\mathrm{Var}(Y) &= \mathrm{E}(Y^2) - \mathrm{E}(Y)^2 && (57)\\
&= \sum_{m=1}^{M} \pi_m (\mu_m^2 + \sigma_m^2) - \big(\sum_{m=1}^{M} \pi_m \mu_m\big)^2 && (58)
\end{aligned}$$