Model-based Rank Aggregation (Normal)

Based on discussions with Tom Dietterich and Andrew Emmott

Generative Process with Separate λ for each detector

1. Draw $\pi \sim Beta(shape = 0.05, rate = 100)$. This is the proportion of anomalous points
2. Draw $\lambda \sim Gamma(shape = 1.1, rate = 1.0)$. This is the score "bonus" for being anomalous
3. For each point x,
   a. Draw $\theta_x \sim Bern(\pi)$
   b. If $\theta_x = 0$ then x is "normal" and $score(x, i) \sim Norm(\lambda_r, \sigma^2)$
   c. Else x is "anomalous", and $score(x, i) \sim Norm(\lambda_a, \sigma^2)$
4. Sort the {x} into descending order and assign ranks such that rank(x, i) is the position of x in the sorted list.

Assume there are *D* detectors. $\boldsymbol{x}_i = \{x_{i1}, \dots, x_{iD}\}$ are the scores reported by detectors $d_1, \dots d_D$. Let $\boldsymbol{\lambda}_r = \{\lambda_{r1}, \dots, \lambda_{rD}\}$ be the mean scores for regular instances and let $\boldsymbol{\lambda}_a = \lambda_{a1}, \dots, \lambda_{aD}$ be the means for anomaly scores. The likelihood of a score under the distribution for 'regular' scores is:

$$f_r(\boldsymbol{x}_i|\boldsymbol{\lambda}_r, \sigma^2) \sim Norm(\boldsymbol{\lambda}_r, \sigma^2) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{id}-\lambda_{rd})^2}{2\sigma^2}\right) \tag{1}$$

And the likelihood of a score under the distribution for 'anomalous' scores is:

$$f_a(\boldsymbol{x}_i|\boldsymbol{\lambda}_a, \sigma^2) \sim Norm(\boldsymbol{\lambda}_a, \sigma^2) = \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{id}-\lambda_{ad})^2}{2\sigma^2}\right) \tag{2}$$

The likelihood of each score assuming we know whether it is normal or anomalous is:

$$f(\boldsymbol{x}_i|\pi, \boldsymbol{\lambda}_a, \boldsymbol{\lambda}_r) = \big((1-\pi) f_r(\boldsymbol{x}_i|\boldsymbol{\lambda}_r)\big)^{1-\theta_i} \big(\pi f_a(\boldsymbol{x}_i|\boldsymbol{\lambda}_a)\big)^{\theta_i} \tag{3}$$

The priors on $\pi, \boldsymbol{\lambda}_a, \boldsymbol{\lambda}_r$ are:

$$f_\pi(\pi) = \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha, \beta)}, \qquad f_{\lambda_a}(\lambda_{ad}|\alpha_a, \beta_a) = \frac{{\beta_a}^{\alpha_a}}{\Gamma(\alpha_a)} {\lambda_{ad}}^{\alpha_a-1} e^{-\beta_a\lambda_{ad}},$$
$$f_{\lambda_r}(\lambda_{rd}|\alpha_r, \beta_r) = \frac{{\beta_r}^{\alpha_r}}{\Gamma(\alpha_r)} {\lambda_{rd}}^{\alpha_r-1} e^{-\beta_r\lambda_{rd}}$$

Where we assume that α, β, α_a, β_a, α_r and β_r are known constants; $B(\alpha, \beta)$ is the Beta function and $\Gamma(\alpha_a)$, $\Gamma(\alpha_r)$ are Gamma functions.

Therefore,

$$f(\boldsymbol{x}_i, \pi, \boldsymbol{\lambda}_a, \boldsymbol{\lambda}_r; \sigma^2) = f(\boldsymbol{x}_i|\pi, \boldsymbol{\lambda}_a, \boldsymbol{\lambda}_r; \sigma^2) f_{\lambda_a}(\boldsymbol{\lambda}_a) f_{\lambda_r}(\boldsymbol{\lambda}_r) f_\pi(\pi)$$
$$= \left(\pi \prod_{d=1}^{D} f_a(x_{id}|\lambda_{ad})\right)^{\theta_i} \left((1 - \pi)\prod_{d=1}^{D} f_r(x_{id}|\lambda_{rd})\right)^{1-\theta_i} \prod_{d=1}^{D} \left\{\frac{{\beta_a}^{\alpha_a}}{\Gamma(\alpha_a)} {\lambda_{ad}}^{\alpha_a-1} e^{-\beta_a\lambda_{ad}} \frac{{\beta_r}^{\alpha_r}}{\Gamma(\alpha_r)} {\lambda_{rd}}^{\alpha_r-1} e^{-\beta_r\lambda_{rd}}\right\} \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha, \beta)} \tag{4}$$

The complete-data likelihood of all scores across all detectors is (where *n* is the number of instances):

$$L(\boldsymbol{x}; \pi, \lambda_{da}, \lambda_{dr}, \sigma^2) = \left[\left\{\prod_{i=1}^{n} \prod_{d=1}^{D} f(x_{id}|\pi, \lambda_{ad})\right\}\left\{\prod_{d=1}^{D} f_{\lambda_a}(\lambda_{ad}) f_{\lambda_r}(\lambda_{rd})\right\}\right] f_\pi(\pi) \tag{5}$$

$$= \left[\prod_{i=1}^{n}\left\{\left(\pi\prod_{d=1}^{D}f_a(x_{id}|\lambda_{ad})\right)^{\theta_i}\left((1 - \pi)\prod_{d=1}^{D}f_r(x_{id}|\lambda_{rd})\right)^{1-\theta_i}\right\}\right]\prod_{d=1}^{D}\left\{\frac{\beta_a{}^{\alpha_a}}{\Gamma(\alpha_a)}\lambda_{ad}{}^{\alpha_a-1}e^{-\beta_a\lambda_{ad}}\frac{\beta_r{}^{\alpha_r}}{\Gamma(\alpha_r)}\lambda_{rd}{}^{\alpha_r-1}e^{-\beta_r\lambda_{rd}}\right\}\frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} \tag{6}$$

Instead of assigning the hard class labels $((1-\theta_i), \theta_i)$ which will be hard to infer, we will use soft-assignments (i.e., responsibilities) denoted by z_{ai} and z_{ri} which refer to the probability of assigning *i*-th score of *d*-th detector to 'anomaly' and 'normal' classes respectively; $z_{ai} + z_{ri} = 1$. We rewrite $L(\boldsymbol{x}_i; \pi, \boldsymbol{\lambda})$:

$$= \left[\prod_{i=1}^{n}\left\{\left(\pi\prod_{d=1}^{D}f_a(x_{id}|\lambda_{ad})\right)^{z_{ai}}\left((1 - \pi)\prod_{d=1}^{D}f_r(x_{id}|\lambda_{rd})\right)^{z_{ri}}\right\}\right]\left[\prod_{d=1}^{D}\left\{\prod_{k\in\{a,r\}}\frac{\beta_k{}^{\alpha_k}}{\Gamma(\alpha_k)}\lambda_{kd}{}^{\alpha_k-1}e^{-\beta_k\lambda_{kd}}\right\}\right]\frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} \tag{7}$$

The log-likelihood is:

$$l(\boldsymbol{x};\pi,\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_r,\sigma^2) = \log\big(L(\boldsymbol{x};\pi,\boldsymbol{\lambda}_a,\boldsymbol{\lambda}_r,\sigma^2)\big)$$

$$= \sum_{i=1}^{n} z_{ai}\left\{\log(\pi) + \sum_{d=1}^{D}\log\big(f_a(x_{id}|\lambda_{ad})\big)\right\} + \sum_{i=1}^{n} z_{ri}\left\{\log(1-\pi) + \sum_{d=1}^{D}\log\big(f_r(x_{id}|\lambda_{rd})\big)\right\} + \sum_{d=1}^{D}\sum_{k\in\{a,r\}}\left\{(\alpha_\lambda - 1)\log(\lambda_{kd}) - \beta_\lambda\lambda_{kd} + \log\left(\frac{\beta_\lambda{}^{\alpha_\lambda}}{\Gamma(\alpha_\lambda)}\right)\right\} + (\alpha-1)\log(\pi) + (\beta-1)\log(1-\pi) - \log\big(B(\alpha,\beta)\big) \tag{8}$$

After substituting $f_a(x_{id}|\lambda_{ad})$ and $f_r(x_{id}|\lambda_{rd})$ in the above equation:

$$l(\boldsymbol{x};\pi,\lambda_{ad},\lambda_{rd},\sigma^2) = \sum_{i=1}^{n} z_{ai}\left\{\log(\pi) + \sum_{d=1}^{D}\left\{-\frac{(x_{id}-\lambda_{ad})^2}{2\sigma^2} - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(2\pi)\right\}\right\} + \sum_{i=1}^{n} z_{ri}\left\{\log(1-\pi) + \sum_{d=1}^{D}\left\{-\frac{(x_{id}-\lambda_{rd})^2}{2\sigma^2} - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\log(2\pi)\right\}\right\} + \sum_{d=1}^{D}\sum_{k\in\{a,r\}}\left\{(\alpha_k - 1)\log(\lambda_{kd}) - \beta_k\lambda_{kd} + \log\left(\frac{\beta_k{}^{\alpha_k}}{\Gamma(\alpha_k)}\right)\right\} + (\alpha-1)\log(\pi) + (\beta-1)\log(1-\pi) - \log\big(B(\alpha,\beta)\big) \tag{9}$$

M-Step: Derive the MLE of parameters by differentiation.

MLE for π:

$$\frac{\partial l(\boldsymbol{x};\pi,\lambda_{ad},\lambda_{rd},\sigma^2)}{\partial\pi} = \frac{1}{\pi}\sum_{i=1}^{n} z_{ai} - \frac{1}{1-\pi}\sum_{i=1}^{n} z_{ri} + \frac{(\alpha-1)}{\pi} - \frac{(\beta-1)}{1-\pi} = 0$$

$$\Rightarrow (1-\pi)\sum_{i=1}^{n} z_{ai} - \pi\sum_{i=1}^{n} z_{ri} + (1-\pi)(\alpha-1) - \pi(\beta-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_{ai} + (\alpha-1) = \pi\left(\sum_{i=1}^{n} z_{ai} + \sum_{i=1}^{n} z_{ri} + (\alpha-1) + (\beta-1)\right)$$

$$\Rightarrow \hat{\pi} = \frac{\sum_{i=1}^{n} z_{ai} + (\alpha-1)}{\left(\sum_{i=1}^{n} z_{ai} + \sum_{i=1}^{n} z_{ri} + (\alpha-1) + (\beta-1)\right)}$$

$$\Rightarrow \hat{\pi} = \frac{\sum_{i=1}^{n} z_{ai} + (\alpha-1)}{\left(n + (\alpha-1) + (\beta-1)\right)} \tag{10}$$

The above follows from the observation that: $z_{ai} + z_{ri} = 1$.

MLE for λ_{kd} for $k \in \{a, r\}$:

$$\frac{\partial l(\boldsymbol{x}; \pi, \lambda_{ad}, \lambda_{rd}, \sigma^2)}{\partial \lambda_{kd}} = \sum_{i=1}^{n} z_{ki}\left\{\frac{(x_{id}-\lambda_{kd})}{\sigma^2}\right\} + \frac{(\alpha_k - 1)}{\lambda_{kd}} - \beta_k = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_{ki}\lambda_{kd}(x_{id}-\lambda_{kd}) + (\alpha_k-1)\sigma^2 - \beta_k\sigma^2\lambda_{kd} = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_{ki}\lambda_{kd}^2 - \left(\sum_{i=1}^{n} z_{ki}x_{id} - \beta_k\sigma^2\right)\lambda_{kd} - (\alpha_k-1)\sigma^2 = 0$$

$$\Rightarrow \hat{\lambda}_{kd} = \frac{\sum_{i=1}^{n} z_{ki}x_{id} - \beta_k\sigma^2 \pm \sqrt{(\sum_{i=1}^{n} z_{ki}x_{id} - \beta_k\sigma^2)^2 + 4(\sum_{i=1}^{n} z_{ki})(\alpha_k-1)\sigma^2}}{2\sum_{i=1}^{n} z_{ki}} \tag{11}$$

MLE for σ^2:

$$\frac{\partial l(\boldsymbol{x}; \pi, \lambda_{ad}, \lambda_{rd}, \sigma^2)}{\partial \sigma^2} = \sum_{i=1}^{n}\sum_{k\in\{a,r\}} z_{ki}\sum_{d=1}^{D}\left\{\frac{(x_{id}-\lambda_{kd})^2}{2(\sigma^2)^2} - \frac{1}{2\sigma^2}\right\} = 0$$

$$\Rightarrow \sum_{i=1}^{n}\sum_{k\in\{a,r\}} z_{ki}\sum_{d=1}^{D}((x_{id}-\lambda_{kd})^2 - \sigma^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\sum_{k\in\{a,r\}} z_{ki}\sum_{d=1}^{D}(x_{id}-\lambda_{kd})^2 = \sigma^2\left(\sum_{i=1}^{n}\sum_{k\in\{a,r\}} D z_{ki}\right)$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\sum_{k\in\{a,r\}} z_{ki}\sum_{d=1}^{D}(x_{id}-\lambda_{kd})^2}{\sum_{i=1}^{n} D\sum_{k\in\{a,r\}} z_{ki}} = \left(\frac{\sum_{i=1}^{n}\sum_{k\in\{a,r\}} z_{ki}\sum_{d=1}^{D}(x_{id}-\lambda_{kd})^2}{nD}\right) \tag{12}$$

The above follows from the observation that: $z_{ai} + z_{ri} = 1$.

E-Step: Compute:

$$E[z_{ai} = 1|\boldsymbol{x}] = P(z_{ai} = 1) = \frac{\pi\prod_{d=1}^{D} f_a(x_{id}|\lambda_{ad},\sigma^2)}{\pi\prod_{d=1}^{D} f_a(x_{id}|\lambda_{ad},\sigma^2) + (1-\pi)\prod_{d=1}^{D} f_r(x_{id}|\lambda_{rd},\sigma^2)} \tag{13}$$

and,

$$E[z_{ri} = 1|\boldsymbol{x}] = P(z_{ri} = 1) = \frac{(1-\pi)\prod_{d=1}^{D} f_r\left(x_{id}|\lambda_{rd},\sigma^2\right)}{\pi\prod_{d=1}^{D} f_a(x_{id}|\lambda_{ad},\sigma^2)+(1-\pi)\prod_{d=1}^{D} f_r(x_{id}|\lambda_{rd},\sigma^2)} \tag{14}$$