Modeling with Mixtures-of-Gammas: Some Practical and Computational Considerations

Tasks for Bill (2017)

- The pdf for a k-component mixture-of-gammas distribution is

$$g(x; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{j=1}^{k} \lambda_j f(x; \alpha_j, \beta_j),$$

  such that f is the pdf for a gamma distribution, $\sum_{j=1}^{k} \lambda_j = 1$, and $\lambda_j > 0$ for all j. In the above notation, we have $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^{\mathrm{T}}$. For our study, we will be looking only at 2-component or 3-component mixture models; i.e., $k \in \{2, 3\}$.

- Load the `mixtools` package and then read in the `gammamixEMnew.R` file that I provided you.

- We will consider these 15 settings:

  (Heavily-Overlapping Components; $k = 2$**)**

  - **Condition 1**: $k = 2$, $\boldsymbol{\alpha} = (2, 5)^{\mathrm{T}}$, $\boldsymbol{\beta} = (3, 4)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.5, 0.5)^{\mathrm{T}}$
  - **Condition 2**: $k = 2$, $\boldsymbol{\alpha} = (2, 5)^{\mathrm{T}}$, $\boldsymbol{\beta} = (3, 4)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.8)^{\mathrm{T}}$

  (Moderately-Overlapping Components; $k = 2$**)**

  - **Condition 3**: $k = 2$, $\boldsymbol{\alpha} = (1, 10)^{\mathrm{T}}$, $\boldsymbol{\beta} = (1, 1)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.5, 0.5)^{\mathrm{T}}$
  - **Condition 4**: $k = 2$, $\boldsymbol{\alpha} = (1, 10)^{\mathrm{T}}$, $\boldsymbol{\beta} = (1, 1)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.8)^{\mathrm{T}}$

  (Well-Separated Components; $k = 2$**)**

  - **Condition 5**: $k = 2$, $\boldsymbol{\alpha} = (2, 30)^{\mathrm{T}}$, $\boldsymbol{\beta} = (3, 2)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.5, 0.5)^{\mathrm{T}}$
  - **Condition 6**: $k = 2$, $\boldsymbol{\alpha} = (2, 30)^{\mathrm{T}}$, $\boldsymbol{\beta} = (3, 2)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.8)^{\mathrm{T}}$

  (Heavily-Overlapping Components; $k = 3$**)**

  - **Condition 7**: $k = 3$, $\boldsymbol{\alpha} = (2, 5, 6)^{\mathrm{T}}$, $\boldsymbol{\beta} = (3, 5, 7)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (1/3, 1/3, 1/3)^{\mathrm{T}}$
  - **Condition 8**: $k = 3$, $\boldsymbol{\alpha} = (2, 5, 6)^{\mathrm{T}}$, $\boldsymbol{\beta} = (3, 5, 7)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.3, 0.5)^{\mathrm{T}}$

  (Moderately-Overlapping Components; $k = 3$**)**

  - **Condition 9**: $k = 3$, $\boldsymbol{\alpha} = (1, 20, 50)^{\mathrm{T}}$, $\boldsymbol{\beta} = (2, 4, 3)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.3, 0.5)^{\mathrm{T}}$
  - **Condition 10**: $k = 3$, $\boldsymbol{\alpha} = (1, 20, 50)^{\mathrm{T}}$, $\boldsymbol{\beta} = (2, 4, 3)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.3, 0.5)^{\mathrm{T}}$

(Well-Separated Components; $k = 3$)

- **Condition 11**: $k = 3$, $\boldsymbol{\alpha} = (2, 50, 180)^{\mathrm{T}}$, $\boldsymbol{\beta} = (1, 2, 3)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.3, 0.5)^{\mathrm{T}}$
- **Condition 12**: $k = 3$, $\boldsymbol{\alpha} = (2, 50, 180)^{\mathrm{T}}$, $\boldsymbol{\beta} = (1, 2, 3)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.2, 0.3, 0.5)^{\mathrm{T}}$

(Common Shape Parameter; Three Different Component Separation)

- **Condition 13**: Heavily-Overlapping, $k = 2$, $\alpha = 5$, $\boldsymbol{\beta} = (5, 10)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.5, 0.5)^{\mathrm{T}}$
- **Condition 14**: Moderately-Overlapping, $k = 2$, $\alpha = 3$, $\boldsymbol{\beta} = (0.50, 3)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.5, 0.5)^{\mathrm{T}}$
- **Condition 15**: Well-Separated, $k = 2$, $\alpha = 1$, $\boldsymbol{\beta} = (0.10, 25)^{\mathrm{T}}$, and $\boldsymbol{\lambda} = (0.5, 0.5)^{\mathrm{T}}$

- For each of the 15 conditions, please do the following:
  - Modify the R script I provided for the settings given in the respective condition. In the script I provided, some lines of code have a # at the end. These are the only lines that need modifying. Create a separate .R file for each simulation and use that file to submit your job to the cluster.
  - Each script needs to be ran for 3 sample sizes: $n \in \{100, 250, 500\}$.
  - For Conditions 1–12, use and modify the file `Condition_1_100.R`. For Conditions 13–15, use and modify the file `Condition_13_100.R`.
  - Save the R workspace (i.e., .RData file) for each of the simulations above.