Motivation: Our scientist friends have noticed that the velocity for the molecular motor appears to be bimodal.

Goal: Estimate the velocity distribution using a mixture model.

Method: Gibbs Sampling/Metropolis Hasting Hybrid algorithm with to estimate the parameters of the mixture model.

Let $\vec{\nu} = (\nu_1, \dots \nu_N)$ denote the velocity data, where $\nu_i > 0$ for $i = 1, \dots, N$. We restrict ν_j to positive values because we consider the angle in which the motor is processing.

1 Model: Mixture Model

We assume the velocity from the mixture model

$$\nu_j \sim p_0 \text{ Gamma}(\alpha_0, \beta_0) + p_1 \text{Gamma}(\alpha_1, \beta_1). \tag{1}$$

where $p_0 + p_1 = 1$. Let f_k denote the density of the k th state for $i = 0, 1$. Under this model, our likelihood take the form

$$\ell(\vec{\nu}; \vec{\alpha}, \vec{\beta}, \vec{p}) = \prod_{j=1}^{N} \Big(\sum_{k=0}^{1} p_k f(\nu_j; \alpha_k, \beta_k) \Big) = \prod_{j=1}^{N} \Big(p_0 \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \nu_j^{\alpha_0 - 1} e^{\beta_0 \nu_j} + p_1 \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \nu_j^{\alpha_1 - 1} e^{\beta_1 \nu_j} \Big)$$

To make the computation more feasible, we introduce the"missing" state data $Z = (z_1, \dots, z_N)$ where

$$z_j = \begin{cases} 0 & \text{if } v_j \sim \text{G}(\alpha_0, \beta_0) \\ 1 & \text{if } v_j \sim \text{G}(\alpha_1, \beta_1), \end{cases}$$

If we condition on the parameters and the state vector factors the likelihood of observing the data is

$$\ell(\vec{\nu}; Z, \vec{\alpha}, \vec{\beta}, \vec{p}) = \Bigg(\prod_{j: z_j = 0} p_0 \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \nu_j^{\alpha_0 - 1} e^{\beta_0 \nu_j} \Bigg) \Bigg(\prod_{j: z_j = 1} p_1 \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \nu_j^{\alpha_1 - 1} e^{\beta_1 \nu_j} \Bigg).$$

2 MCMC Algorithm

We utilize a conjugacy structure on the priors of the parameter of the mixture model, so that $f_i(\alpha_i, \beta_i)$ for $i = 0, 1$ are hyper distributions where α_i and β_i follow different distributions with specified hyper parameters.

2.0.1 Rate Parameter β_i for $i = 0, 1$

Prior: $\beta_i \sim Gamma(\underline{a}_k, \underline{b}_k)$ for $k = 0, 1$, where the hyper parameters are selected so that $\hat{\beta}_k^{(0)} = \underline{a}_k / \underline{b}_k$ and the prior variance optimizes convergence (which is more of a constraint on $\underline{b}_k$).

ith iteration: $\hat{\beta}_k \sim Gamma(\underline{a}_k + n_k \hat{\alpha}_k,\ \underline{b}_k + \sum_{j: s_j = k} v_j)$ for k = 0,1

Initializing: Let $n_k = \sum_{j=1}^{N} 1_{z_j = k}$ and $\bar{\nu}_k = \frac{1}{n_k} \sum_{j=1}^{N} \nu_j 1_{\{z_j = k\}}$ denote the mean of the kth state. Using method of moments we initially estimate

$$\hat{\beta}_k^{(0)} = \frac{\bar{\nu}_k}{\frac{1}{n_k - 1} \sum_{j: z_j = k}^{N} (\nu_j - \bar{\nu}_k)^2}.$$

2.0.2 Shape Parameter α_i for $i = 0, 1$

Prior: There exists a conjugate prior for α_k with hyper parameters, $\underline{a}_k, \underline{b}_k, \underline{c}_k$

$$\alpha_k \sim \frac{\underline{a}_k^{\alpha-1}\beta^{\alpha\underline{c}_k}}{\Gamma(\alpha)^{\underline{b}_k}}.$$

ith iteration:

$$\hat{\alpha}_k \sim \frac{(\underline{a}_k \prod_{j:z_j=k} v_j)^{\hat{\alpha}_k-1}\hat{\beta}_k^{\hat{\alpha}_k\underline{c}_k+n_k}}{\Gamma(\hat{\alpha}_k)^{\underline{b}_k+n_k}}.$$

That is the posterior parameters are updated via

$$\underline{a}_k \to \underline{a}_k \prod_{j:z_j=k} v_j, \quad \underline{b}_k \to \underline{b}_k + n_k, \quad \underline{c}_k \to \underline{c}_k + n_k.$$

Although this is not a well known density, we sample from the posterior distribution using a Metropolis Hasting algorithm with a random walk transition density. Note that for x large, the gamma function in R does not work.

Initializing: Let $n_k = \sum_{j=1}^N 1_{z_j=k}$ and $\bar{\nu}_k = \frac{1}{n_k}\sum_{j=1}^N \nu_j 1_{\{z_j=k\}}$ denote the mean of the kth state. Using method of moments we initially estimate

$$\hat{\alpha}_k^{(0)} = \frac{\bar{\nu}_k^2}{\frac{1}{n_k-1}\sum_{j:z_j=k}^N(\nu_j - \bar{\nu}_k)^2}.$$

Understanding the conjugate prior: Since this is not a well known distribution, how do we pick the hyper parameters? One method is to make the method of moments estimate for α_i be close to the maximum of the the log likelihood, which has form:

$$\begin{aligned}\log(f(\alpha_k)) &\overset{c}{=} \alpha(\log(a) + c\log(\beta)) - b\log(\Gamma(\alpha) - \log(a) \\ &\approx \alpha(\log(a) + c\log(\beta)) - b(\alpha\log(\alpha) - \alpha) - \log(a)\end{aligned}$$

We see for small α values the log likelihood linearly depends on α with slope depending on a and c and for larger values of α the hyper parameter b influences the value of the log likelihood.

2.1 Other parameters in the model

2.1.1 The State Vector, Z

Initializing: We pick an arbitrary threshold velocity, ν^*, based on the spread/histogram of the velocity. If

$$z_j = \begin{cases} 0 & \text{if } \nu_j \le \nu^* \\ 1 & \text{if } \nu_j > \nu^* \end{cases}$$

ith iteration: For $jin1 : N$, generate $u \sim Unif(0, 1)$ and set $z_j^{(i)} = 0$ if $u \le p_{0j}$ where

$$p_{0j} = \frac{p_0 f(\nu_j : \alpha_0^{(i)}, \beta_0^{(i)})}{p_0 f(\nu_j : \alpha_0^{(i)}, \beta_0^{(i)}) + p_1 f(\nu_j : \alpha_1^{(i)}, \beta_1^{(i)})}$$

2.1.2 Probability Parameter, p

Initializing: Let $n_k = \sum_{j=1}^N 1_{z_j=k}$ We initially estimate

$$\hat{p}_k^{(0)} = \frac{\sum_{j=1}^N 1_{z_j=k}}{N}.$$

Prior: $p_0 \sim Beta(\underline{a}, \underline{b})$. In the case that $k > 2$ the conjugate prior would be Dirichlet.

ith iteration: $\hat{p}_0 \sim beta(\underline{a}_0 + n_0\hat{\alpha}_0,\ \underline{b}_0 + \sum_{j:s_j=0} v_j)$ and $\hat{p}_1 = 1 - \hat{p}_0$.

3 Inference on Model

Assuming all Markov Chains have converged to their stationary distribution, we simulate a sample of 10,000 velocities from the posterior predictive distribution. The ith velocity is obtained by

1. Setting the ith parameter vector $\hat{\theta} = (\hat{p}, \hat{\alpha}_0, \hat{\beta}_0, \hat{\alpha}_1, \hat{\beta}_1)$ from the Gibbs Sampling.
2. Draw a uniform random variable on (0,1), u, and draw velocity

$$\hat{v}_i \sim \begin{cases} Gamma(\hat{\alpha}_0, \hat{\beta}_0,) & \text{if } u \leq \hat{p}_0 \\ Gamma(\hat{\alpha}_1, \hat{\beta}_1,) & \text{if } u > \hat{p}_0. \end{cases}$$

We test whether the simulate sample, $\hat{v}$ comes from the underlying distribution using the Kolmogorov Smirnov test statistic:

$$D = \sup\left|F_N^{\nu}(u) - F_N^{\hat{\nu}}(u)\right|$$

where

$$F_n^{\nu}(u) = \frac{1}{n}\sum_{j=1}^{N} 1\{\nu_j \leq u\} \quad \text{and} \quad F_{\hat{N}}^{\hat{\nu}}(u) = \frac{1}{n}\sum_{j=1}^{\hat{N}} 1\{\hat{\nu}_j \leq u\},$$

the empirical distribution function of the velocity data and the simulated velocities, respectively. We found the posterior predictive velocities and the sample velocity data has the same distribution, while we are unable to conclude the posterior predictive velocities and the underlying distribution (the one used to generate the fake velocity data) were from the same distribution.