MDSINE 2.0 Supplement

David Kaplan

March 18, 2020

Contents

1 Notation

We parameterize the Scale-Inv-χ^2distribution with degrees of freedom ν and scale τ^2 as

$$\text{Scale-Inv-}\chi^2(x;\nu,\tau^2) = \frac{(\tau^2\nu/2)^{-\nu/2}}{\Gamma(\nu/2)} x^{(-\nu/2)-1} \exp\Big(-\frac{-\nu\tau^2}{2x}\Big) \tag{1}$$

We parameterize the Gamma distribution with shape k and scale θ as

$$\text{Gamma}(x;k,\theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\Big(-\frac{x}{\theta}\Big) \tag{2}$$

We parameterize the Negative Binomial distribution with mean ϕ and dispersion ϵ as

$$\text{NegBin}(y;\phi,\epsilon) = \frac{\Gamma(r+y)}{y!\Gamma(r)}\Big(\frac{\phi}{r+\phi}\Big)^y\Big(\frac{r}{r+\phi}\Big)^r \tag{3}$$

$$r = \frac{1}{\epsilon} \tag{4}$$

We parameterize a truncated normal distribution with mean μ, variance σ^2, lower bound v_1, and upper bound v_2 as

$$\text{TruncNormal}_{(v_1,v_2)}(x;\mu,\sigma^2) = \frac{1}{(2\pi)^{1/2}\sigma}\frac{1}{\Phi\Big(\frac{v_2-\mu}{\sigma}\Big) - \Phi\Big(\frac{v_1-\mu}{\sigma}\Big)} \exp\Big(\frac{-(x-\mu)^2}{2\sigma^2}\Big) \tag{5}$$

where $\Phi(\cdot)$ is the the cumulative distribution function of a standard normal distribution. Normal distributions are written in terms of the variance Normal(μ,σ^2).

2 Model Details

2.1 Dirichlet process for interactions and perturbations

We incorporate a Dirichlet Process (DP)-based clustering technique ([2], [3]) to learn redundant interaction structures (interaction modules) and perturbation effects among OTUs. Let $\mathbf{c}_i \in \mathbb{Z}^+$ be the cluster assignment of OTU i. In the context of our dynamical systems module, only interaction coefficients between modules need to be learned ($\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j}$), rather than interactions between each pair of OTUs ($\mathbf{b}_{i,j}$). If OTU i and OTU j are in different clusters, i.e. $\mathbf{c}_i \neq \mathbf{c}_j$, then $\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j} \in \mathbb{R}$ is the coefficient representing the (interaction) effect from cluster $\mathbf{c}_j$ to cluster $\mathbf{c}_i$. If OTU and OTU j are in the same cluster, i.e. $\mathbf{c}_i = \mathbf{c}_j$, then we assume there is no interaction. If OTU l is in the same cluster as OTU j, different than OTU i, then $\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_l} = \mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j}$ by definition.

Additionally, perturbation effects are modeled on the cluster level: $\gamma^{(j)}$, where $\gamma^{(j)}$ is the j^{th} perturbation for the cluster assigned to OTU i. If OTU i and OTU l are in the same cluster, $\gamma^{(j)}_{\mathbf{c}_i} = \gamma^{(j)}_{\mathbf{c}_l}$. Analogously, if OTU i and OTU l are not in the same cluster, $\gamma^{(j)}_{\mathbf{c}_i} \neq \gamma^{(j)}_{\mathbf{c}_l}$.

We generate the mixture weights $C_1, C_2, ...$ from a "stick-breaking" construction with concentration parameter θ:

$$V_1, V_2, ... \sim \text{Beta}(1,\theta) \tag{6}$$

$$C_k = V_k \prod_{j=1}^{k-1}(1-V_j) \tag{7}$$

Atom locations locations $\eta_1, \eta_2, ...$ are drawn from a base distribution G_0.

$$\eta_1, \eta_2, ... \sim G_0 \tag{8}$$

$$\Theta = \sum_{k=1}^{\infty} C_k \delta_{\eta_k} \tag{9}$$

where Θ is a draw from a Dirchlet process with concentration parameter θ and base distribution G_0, or $\Theta \sim \mathrm{DP}(\theta, G_0)$. δ_{η_k} is an indicator function that is 0 everywhere, except for $\delta_{\eta_k}(C_k) = 1$.

To specify the dirichlet process mixture model, we specify cluster assignment variable $\mathbf{c}_i$ for each OTU i, and a density $\pi_{\mathbf{c}}$ which takes an atom location as a parameter.

$$\mathbf{c}_i \sim \mathrm{Multinomial}(C_k) \tag{10}$$

$$i|\mathbf{c}_i \sim \pi_{\mathbf{c}}(\cdot|\eta_{C_k}) \tag{11}$$

2.2 Baseline logistic growth

We model the dynamics for OTU i in subject s at time t_k with the stochastic version of the generalized Lotka-Volterra (gLV) equations:

$$\begin{aligned}\log(\mu_{si}(k+1)) = \log(\mathbf{x}_{si}(k)) + \Delta_k \Big[\mathbf{a}_{1,i}\Big(1 + \sum_{p=1}^{P} \gamma_{\mathbf{c}_i}^{(p)} \mathbf{z}_{\mathbf{c}_i}^{(\gamma,p)} h_p\Big) \\ + \mathbf{a}_{2,i}\mathbf{x}_{si}(k) + \sum_{\mathbf{c}_i \neq \mathbf{c}_j} \mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j} \mathbf{z}_{\mathbf{c}_i,\mathbf{c}_j}^{(\mathbf{b})} \mathbf{x}_{sj}(k)\Big]\end{aligned} \tag{12}$$

$$\log(\mathbf{x}_{si}(k+1)) \sim \mathrm{Normal}(\log(\mu_{si}(k+1)), \Delta_k \sigma_w^2) \tag{13}$$

where $\mathbf{x}_{si}(k) \in \mathbb{R}^{\geq 0}$ is the latent abundance of OTU i at time t_k for subject s, and $\Delta_k = t_{k+1} - t_k$. The growth $\mathbf{a}_{1,i}$ and self-interaction $\mathbf{a}_{2,i}$ variables are modeled for each OTU and are parameterized with a truncated normal distribution:

$$\mathbf{a}_{1,i} \sim \mathrm{Normal}_{(0,\infty)}(\mu_{\mathbf{a}_1}, \sigma^2_{\mathbf{a}_1}) \tag{14}$$

$$\mathbf{a}_{2,i} \sim \mathrm{Normal}_{(-\infty,0)}(\mu_{\mathbf{a}_2}, \sigma^2_{\mathbf{a}_2}) \tag{15}$$

We place Scale-Inv-χ^2priors on the variances:

$$\sigma^2_{\mathbf{a}_1} \sim \text{Scale-Inv-}\chi^2(\nu_{\mathbf{a}_1}, \tau^2_{\mathbf{a}_1}) \tag{16}$$

$$\sigma^2_{\mathbf{a}_2} \sim \text{Scale-Inv-}\chi^2(\nu_{\mathbf{a}_2}, \tau^2_{\mathbf{a}_2}) \tag{17}$$

The variable $\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j}$ represents the interaction coefficient from cluster $\mathbf{c}_i$ to cluster $\mathbf{c}_j$ and is parameterized with a normal distribution:

$$\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j} \sim \mathrm{Normal}(0, \sigma^2_{\mathbf{b}}) \tag{18}$$

$$\sigma^2_{\mathbf{b}} \sim \text{Scale-Inv-}\chi^2(\nu_{\mathbf{b}}, \tau^2_{\mathbf{b}}) \tag{19}$$

The variable $\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j}$ represents the binary indicator variable that selects the interaction from $\mathbf{c}_i$ to $\mathbf{c}_j$ and is parameterized with a Bernoulli distribution with probability $\pi_{\mathbf{b}}$:

$$\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} \sim \mathrm{Bernoulli}(\pi_{\mathbf{b}}) \tag{20}$$

$$\pi_{\mathbf{b}} \sim \mathrm{Beta}(b_{1,\mathbf{b}}, b_{2,\mathbf{b}}) \tag{21}$$

The process variance σ_w^2 is parameterized with an Scale-Inv-χ^2:

$$\sigma^2_w \sim \text{Scale-Inv-}\chi^2(\nu_w, \tau^2_{\sigma^2_w}) \tag{22}$$

2.3 Perturbations

We assume the baseline growth rate can be modulated by a perturbation. The variable $\gamma_{\mathbf{c}_i}^{(p)}$ represents the magnitude of the p^{th} perturbation on cluster $\mathbf{c}_i$ and is parameterized with a normal distribution:

$$\gamma_{\mathbf{c}_i}^{(p)} \sim \text{Normal}(0, \sigma_{\gamma,p}^2) \tag{23}$$

$$\sigma_{\gamma,p}^2 \sim \text{Scale-Inv-}\chi^2(\nu_\gamma, \tau_\gamma^2) \tag{24}$$

The period that $\gamma^{(p)}$ is active is given by the step function h_p:

$$h_p(k) = \begin{cases} 1 & t_p^{\text{on}} < t_k \leq t_p^{\text{off}} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

Additionally, there is a binary indicator variable $\mathbf{z}_{\mathbf{c}_i}^{(\gamma,p)}$ that selects whether cluster $\mathbf{c}_i$ is affected by perturbation $\gamma^{(p)}$ and is parameterized by a Bernoulli distribution:

$$\mathbf{z}_{\mathbf{c}_i}^{(\gamma,p)} \sim \text{Bernoulli}(\pi_{\gamma,p}) \tag{26}$$

$$\pi_{\gamma,p} \sim \text{Beta}(b_{1,\gamma}, b_{2,\gamma}) \tag{27}$$

2.4 Error model

The observed data are the sequencing counts $\mathbf{y}_{si}(k)$ and qPCR measurements $\mathbf{Q}_s(k)$. We model $\mathbf{y}_{si}(k)$ with a negative binomial distribution [4]:

$$\mathbf{y}_{si}(k) \sim \text{NegBin}\Big(\phi_{si}(\mathbf{x}_s(k), r_s(k)), \epsilon_{si}(\mathbf{x}_s(k), d_0, d_1)\Big) \tag{28}$$

$$\phi_{si}(\mathbf{x}_s(k), r_s(k)) = r_s(k)\frac{\mathbf{x}_{si}(k)}{\sum_i \mathbf{x}_{si}(k)} \tag{29}$$

$$\epsilon_{si}(\mathbf{x}_s(k), d_0, d_1) = \frac{d_0}{\mathbf{x}_{si}(k)/\sum_i \mathbf{x}_{si}(k)} + d_1 \tag{30}$$

where $r_s(k)$ is the total number of reads at time t_k for subject s, and d_0 and d_1 are the negative binomial dispersion scaling parameters which are pre-trained on raw reads. We can model $\mathbf{Q}_s(k)$ with a normal distribution:

$$\mathbf{Q}_s(k) \sim \text{Normal}\Big(\sum_i \mathbf{x}_{si}(k), \sigma^2_{\mathbf{Q}_s(k)}\Big) \tag{31}$$

where $\sigma^2_{\mathbf{Q}_s(k)}$ is the variance over the triplicate measurements of the qPCR measurement at time t_k for subject s.

3 Model inference

Before we perform inference on the model, we learn the negative binomial dispersion parameters (d_0 and d_1) offline and then fix them during inference. To see how inference on these parameters is performed, seed appendix B. Inference on all other parameters is performed using Markov Chain Monte Carlo (MCMC). We use Gibbs steps for variables with conjugate priors and Metropolis-Hastings (MH) steps otherwise. The order of the main sampling loop is as follows:

1. Sample cluster interaction indicators $\mathbf{z}^{(\mathbf{b})}$ and probability $\pi_{\mathbf{b}}$.

2. Sample perturbation indicators $\mathbf{z}^{(\gamma,p)}$ and probabilities $\pi_{\gamma,p}$.

3. Sample interaction magnitudes $\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j}$ and perturbation magnitudes $\gamma_{\mathbf{c}_i}^{(p)}$ jointly.
4. Sample prior variances $\sigma_{\mathbf{b}}^2$ and $\sigma_{\gamma,p}^2$
5. Sample growths $\mathbf{a}_1$ and self-interactions $\mathbf{a}_2$ separately by randomly choosing which one to update first.
6. Sample prior variances $\sigma_{\mathbf{a}_1}^2$ and $\sigma_{\mathbf{a}_2}^2$.
7. Sample process variance σ_w^2.
8. Sample latent trajectory $\mathbf{x}_{si}(k)$.
9. Sample cluster assignments $\mathbf{c}_i$.
10. Sample concentration parameter θ.

3.1 Logistic growth parameters

Since $\mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j}$ and $\gamma_{\mathbf{c}_i}^{(p)}$ have conjugate normal priors, they can be sampled with straight forward Gibbs sampling. The priors of $\mathbf{a}_{1,i}$ and $\mathbf{a}_{2,i}$ are truncated normal but are still conjugate and thus can be sampled with Gibbs sampling. To see a derivation of the posterior see appendix C.

3.2 Prior variances

3.2.1 Perturbation and interaction prior variances

Since $\sigma_{\mathbf{b}}^2$ and $\sigma_{\gamma,p}^2$ have conjugate Scale-Inv-χ^2 priors, they can be sampled with straight forward Gibbs sampling.

3.2.2 Growth and self-interaction prior variances

The prior variances for $\sigma_{\mathbf{a}_1}^2$ and $\sigma_{\mathbf{a}_2}^2$ cannot be sampled directly and thus we do MH step. In the following section, $\mathbf{a}_l, l = 1, 2$ refers to either the growths or the self-interactions (sampling is the same for either). The proposal for the j^{th} MH step is a Scale-Inv-χ^2 distribution if the likelihood was a normal distribution instead of a truncated normal distribution:

$$(\sigma_{\mathbf{a}_l}^2)^{(*)} = \text{Scale-Inv-}\chi^2(\nu, \tau^2) \tag{32}$$

where

$$\nu = \nu_{\mathbf{a}_l} + n_o \tag{33}$$

$$\tau^2 = \frac{\nu_{\mathbf{a}_l}\tau_{\mathbf{a}_l}^2 + \sum_{i=1}^{n_o}(\mathbf{a}_{l,i} - \mu_{\mathbf{a}_l})^2}{\nu_{\mathbf{a}_l} + n_o} \tag{34}$$

The target distribution conditional on all other parameters Ω is:

$$p(\sigma_{\mathbf{a}_l}^2; \Omega) \propto \text{TruncNormal}_{(v_1,v_2)}(\mathbf{a}_l; \mu_{\mathbf{a}l}, \sigma_{\mathbf{a}_l}^2) \tag{35}$$

where v_1 and v_2 are the respective truncation points for the growths or self-interactions. The acceptance probability for the j^{th} MH step is

$$r_{\text{accept}}^{(j)} = \frac{p((\sigma_{\mathbf{a}_l}^2)^{(*)}; \Omega)\text{Scale-Inv-}\chi^2((\sigma_{\mathbf{a}_l}^2)^{(j-1)}; \nu, \tau^2)}{p((\sigma_{\mathbf{a}_l}^2)^{(j-1)}; \Omega)\text{Scale-Inv-}\chi^2((\sigma_{\mathbf{a}_l}^2)^{(*)}; \nu, \tau^2)} \tag{36}$$

The next value for $(\sigma_{\mathbf{a}_l}^2)^{(j)}$ is given by:

$$(\sigma_{\mathbf{a}_l}^2)^{(j)} = \begin{cases} (\sigma_{\mathbf{a}_l}^2)^{(*)} & \text{with probability } \min(1, r_{\text{accept}}^{(j)}) \\ (\sigma_{\mathbf{a}_l}^2)^{(j-1)} & \text{otherwise} \end{cases} \tag{37}$$

3.3 Process variance

Since σ_w^2 has a conjugate prior, it can be updated with straight forward Gibbs sampling.

3.4 Cluster interaction and perturbation indicators

A Gibbs step is used to update the interaction indicator $\mathbf{z}^{(\mathbf{b})}$ and perturbation indicator $\mathbf{z}^{(\gamma)}$ variable to determine if an interaction or perturbation is present or not. Due to conjugacy, $\mathbf{z}^{(\mathbf{b})}$ and $\mathbf{z}^{(\gamma)}$ can be integrated out jointly conditioned on the growth rates and self-interactions ($\mathbf{a}_1$ and $\mathbf{a}_2$). A derivation of this marginalization can be seen in appendix A. In the following equations, let $F_u(\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j}, \mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i}|\mathbf{a}_1, \mathbf{a}_2)$, $u \in \{0, 1\}$ be the marginal likelihood that $\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} = u$ or $\mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} = 0$:

$$F_u(\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j}, \mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i}|\mathbf{a}_1\mathbf{a}_2) = \int \prod_s \prod_k \text{Normal}\left(\log(\mathbf{x}_{si}(k)); \log(\mu_{si}(k)), \Delta_k\sigma_w^2\right) \mathrm{d}(\gamma, \mathbf{b}) \tag{38}$$

where $\log(\mu_{si}(k))$ is the mean (defined in equation 12) calculated for the current assignments of the indicators $\mathbf{z}^{(\gamma)}$ and $\mathbf{z}^{(\mathbf{b})}$.

3.4.1 Sampling indicators of interactions

Let

$$P_0^{(\mathbf{b})}(\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} = 0, \mathbf{z}^{(\gamma)}|\mathbf{a}_1, \mathbf{a}_2) = F_0(\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} = 0)\text{Bernoulli}(0; \pi_{\mathbf{b}}) \tag{39}$$

$$P_1^{(\mathbf{b})}(\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} = 1, \mathbf{z}^{(\gamma)}|\mathbf{a}_1, \mathbf{a}_2) = F_1(\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} = 1)\text{Bernoulli}(1; \pi_{\mathbf{b}}) \tag{40}$$

be the posterior probabilities of assigning $\mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j}$ to 0 and 1, respectively. Given both likelihoods, we sample $u|\Omega$.

3.4.2 Sampling indicators of perturbations

Let

$$P_0^{(\gamma)}(\mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} = 0, \mathbf{z}^{(\mathbf{b})}|\mathbf{a}_1, \mathbf{a}_2) = F_0(\mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} = 0)\text{Bernoulli}(0; \pi_{\gamma,p}) \tag{41}$$

$$P_1^{(\gamma)}(\mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} = 1, \mathbf{z}^{(\mathbf{b})}|\mathbf{a}_1, \mathbf{a}_2) = F_1(\mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} = 1)\text{Bernoulli}(1; \pi_{\gamma,p}) \tag{42}$$

be the posterior probabilities of assigning $\mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i}$ to 0 and 1, respectively. Given both likelihoods, we sample $u|\Omega$.

3.5 Cluster interaction and perturbation probabilities

Since $\pi_{\mathbf{b}}$ and $\pi_{\gamma,p}$ have conjugate beta priors, they can be sampled with straightforward Gibbs sampling.

3.6 Latent trajectory

At each time point t_k, OTU i, subject s, and MH step j, the following jumping proposal is used:

$$\log\left(\mathbf{x}^{(*)}_{si}(k)\right) \sim \text{Normal}\Big(\log\left(\mathbf{x}^{(j-1)}_{si}(k)\right), \sigma^2_{\text{prop}}(s, i, k)\Big) \tag{43}$$

The proposal variance the initial proposal variance is set to $\sigma^2_{\text{prop}}(s, i, k) = \log(\mathbf{x}^2_{si}(k)/100)$ and is tuned during the first half of burn-in to adjust the acceptance rate to 0.44; the optimal rate for a scalar MH step [1]. The unnormalized target distribution conditional on all other parameters Ω is:

$$
\begin{aligned}
p(\mathbf{x}_{si}(k);\Omega) \propto\ & \text{NegBin}\Big(\mathbf{y}_{si}(k);\phi(\mathbf{x}_s(k)),\epsilon(\mathbf{x}_s(k))\Big) && (44)\\
& \times \text{Normal}\Big(\mathbf{Q}_s(k);\sum_i \mathbf{x}_{si}(k),\sigma^2_{\mathbf{Q}_s(k)}\Big) && (45)\\
& \times \text{Normal}\Big(\log\Big(\mathbf{x}_{si}(k)\Big);\log\Big(\mu_{si}(k)\Big),\Delta_k\sigma_w^2\Big) && (46)\\
& \times \text{Normal}\Big(\log\Big(\mathbf{x}_{si}(k+1)\Big);\log\Big(\mu_{si}(k+1)\Big),\Delta_{k+1}\sigma_w^2\Big) && (47)
\end{aligned}
$$

where ϕ and ϵ were defined in section 2.4. Note that when we are sampling the last time point ($k = n_T$) we do not calculate the likelihood of the future time point (set the log-likelihood to 0). when $k = 0$, we cannot calculate the probability from the previous time point so we sample from the prior of $\mathbf{x}$. Additionally, if we are sampling an intermediate time point (There is no qPCR or count measurement for the time and subject we are sampling), we do not calculate the likelihood for the qPCR or counts (set the log-likelihood to 0). The acceptance probability for the j^{th} MH sample is calculated as the ratio:

$$
r^{(j)}_{\text{accept}} = \frac{p\Big(\log\Big(\mathbf{x}^{(*)}_{si}(k)\Big);\Omega\Big)\text{Normal}\Big(\log\Big(\mathbf{x}^{(j-1)}_{si}(k)\Big);\log\Big(\mathbf{x}^{(*)}_{si}(k)\Big),\sigma^2_{\text{prop}}(s,i,k)\Big)}{p\Big(\log\Big(\mathbf{x}^{(j-1)}_{si}(k)\Big);\Omega\Big)\text{Normal}\Big(\log\Big(\mathbf{x}^{(*)}_{si}(k)\Big);\log\Big(\mathbf{x}^{(j-1)}_{si}(k)\Big),\sigma^2_{\text{prop}}(s,i,k)\Big)} \tag{48}
$$

Because our proposal distribution is symmetric, the forward and reverse jumping likelihoods are equal to each other, so our acceptance probability simplifies to:

$$
r^{(j)}_{\text{accept}} = \frac{p\Big(\log\Big(\mathbf{x}^{(*)}_{si}(k)\Big);\Omega\Big)}{p\Big(\log\Big(\mathbf{x}^{(j-1)}_{si}(k)\Big);\Omega\Big)} \tag{49}
$$

where

$$
\begin{aligned}
p\Big(\mathbf{x}_{si}(k);\Omega\Big) \propto\ & \left(\sigma_Q\sigma_w^2\sqrt{\Delta_k\Delta_{k+1}}\right)^{-1} && (50)\\
& \times \exp\left(-\frac{1}{2}\left(\frac{\mathbf{Q}_s(k)-\sum_i \mathbf{x}_{si}(k)}{\sigma_{\mathbf{Q}_s(k)}}\right)^2\right)\\
& \times \exp\left(-\frac{1}{2}\frac{\left(\log\left(\mathbf{x}_{si}(k)\right)-\log\left(\mu_{si}(k)\right)\right)^2}{\Delta_k\sigma_w^2}\right)\\
& \times \exp\left(-\frac{1}{2}\frac{\left(\log\left(\mathbf{x}_{si}(k+1)\right)-\log\left(\mu_{si}(k+1)\right)\right)^2}{\Delta_{k+1}\sigma_w^2}\right)\\
& \times \frac{\Gamma(r+\mathbf{y}_{si}(k))}{\mathbf{y}_{si}(k)!\Gamma(r)}\left(\frac{\phi(\mathbf{x}_{si}(k))}{r+\phi(\mathbf{x}_{si}(k))}\right)^{\mathbf{y}_{si}(k)}\left(\frac{r}{r+\phi(\mathbf{x}_{si}(k))}\right)^{r}
\end{aligned}
$$

where $r = 1/\epsilon(\mathbf{x}_{si}(k))$. If we eliminate variables that are not dependent on $\mathbf{x}_{si}(k)$, we are able to simplify to

$$
\begin{aligned}
p\Big(\mathbf{x}_{si}(k);\Omega\Big) \propto \Big(\sigma_Q\sigma_w^2\Big)^{-1} & \\
\times \exp\left(-\frac{1}{2}\left(\frac{\mathbf{Q}_s(k) - \sum_i \mathbf{x}_{si}(k)}{\sigma_{\mathbf{Q}_s(k)}}\right)^2\right) & \\
\times \exp\left(-\frac{1}{2}\frac{\left(\log\left(\mathbf{x}_{si}(k)\right) - \log\left(\mu_{si}(k)\right)\right)^2}{\Delta_k\sigma_w^2}\right) & \\
\times \exp\left(-\frac{1}{2}\frac{\left(\log\left(\mathbf{x}_{si}(k+1)\right) - \log\left(\mu_{si}(k+1)\right)\right)^2}{\Delta_{k+1}\sigma_w^2}\right) & \\
\times \left(\frac{\phi(\mathbf{x}_{si}(k))}{r + \phi(\mathbf{x}_{si}(k))}\right)^{\mathbf{y}_{si}(k)} \left(\frac{r}{r + \phi(\mathbf{x}_{si}(k))}\right)^{r} &
\end{aligned}
\tag{51}
$$

The next values for $\mathbf{x}_{si}(k)$ are given by:

$$
\mathbf{x}_{si}^{(j)}(k) = \begin{cases} \mathbf{x}_{si}^{(*)}(k) & \text{with probability } \min(1, r_{\text{accept}}^{(j)}) \\ \mathbf{x}_{si}^{(j-1)}(k) & \text{otherwise} \end{cases}
\tag{52}
$$

3.7 Cluster assignments

The cluster assignment parameter $\mathbf{c}$ is sampled suing Gibbs sampling according to algorithm 8 described in [2]. First, we sample the assignments of OTU i by marginalizing out the clusters. In the following equations, $F_m(\gamma_{\mathbf{c}_i}^{(p)}, \mathbf{c}_i|\mathbf{a}_1, \mathbf{a}_2)$ denotes the marginal likelihood for OTU i assigned to cluster m. A derivation for marginalization can be seen in appendix A:

$$
F_m(\mathbf{x}_i|\mathbf{a}_1, \mathbf{a}_2) = \int \prod_{s,k} \text{Normal}\left(\log(\mathbf{x}_{si}(k)); \log(\mu_{si}^{(m)}(k)), \Delta_k\sigma_w^2\right) \mathrm{d}(\gamma, \mathbf{b})
\tag{53}
$$

where $\log(\mu_{si}^{(m)}(k))$ is the log mean of the dynamics when the OTU i is assigned to cluster m

$$
\begin{aligned}
\log(\mu_{si}^{(m)}(k)) = \log(\mathbf{x}_{si}(k-1)) + \Delta_k \Bigg[\mathbf{a}_{1,i}\left(1 + \sum_{p=1}^{P}\gamma_m^{(p)}\mathbf{z}_m^{(\gamma,p)}h_p\right) & \\
+ \mathbf{a}_{2,i}\mathbf{x}_{si}(k-1) + \sum_{m\neq\mathbf{c}_j}\mathbf{b}_{m,\mathbf{c}_j}\mathbf{z}_{m,\mathbf{c}_j}^{(\mathbf{b})}\mathbf{x}_{si}(k-1)\Bigg] &
\end{aligned}
\tag{54}
$$

Conditional on the growth rates and the self-interactions, the posterior distribution for cluster assignment $\mathbf{c}_i$ is

$$
P(m = \mathbf{c}_i|\mathbf{c}_{-i}, \Omega) \propto \begin{cases} n_{-i,m}F_m(\mathbf{x}_{si}) & \text{for } m \in \mathbf{c} \\ \theta F_m(\mathbf{x}_{si}) & \text{for } m \text{ new cluster} \end{cases}
\tag{55}
$$

where $n_{-i,m}$ are the number of OTUs in cluster m excluding OTU i. Given all these likelihoods, we then sample $m|\Omega$. In the case when calculating the likelihood of staying in a singleton cluster ($n_{-i,m} = 0$), then we set $n_{-i,m} = \theta$.

3.8 Cluster concentration

The cluster concentration parameter θ is sampled according to the auxiliary variable method described in [7]. Our posterior is a mixture of Gamma distributions with auxiliary variable η:

$$P(\theta|\eta, n_c) \sim \pi_\eta \text{Gamma}\left(\alpha_\theta + n_c, \hat{\beta}_\theta +\right) \tag{56}$$
$$(1-\pi_\eta)\text{Gamma}\left(\alpha_\theta + n_c - 1, \hat{\beta}_\theta\right)$$

where

$$\hat{\beta}_\theta = \frac{1}{1/\beta_\theta + \log(\eta)} \tag{57}$$

$$\frac{\pi_\eta}{1-\pi_\eta} = \frac{\alpha_\theta + n_c - 1}{n_o(\beta_\theta - \log(\eta))} \tag{58}$$

$$P(\eta|\theta, n_o) \sim \text{Beta}(\theta + 1, n_o) \tag{59}$$

where n_c is the current number of clusters and n_o are the number of OTUs. At each Gibbs step, using the current number of clusters n_c and current value of the cluster concentration θ, we can draw a new value of θ by

1. Sample η using (59).
2. Sample a new θ from the mixture (56) with the current mixture weights.

We repeat the above steps within a single Gibbs step until θ settles on a value. In MDSINE 2.0 we sample 20 times.

4 Initialization

4.1 Growth rates

The prior mean of the growth rates is set to 2:

$$\mu_{\mathbf{a}_1} = 2 \tag{60}$$

The prior variance of the growth rates is set by squaring the mean of the growth rates obtained when doing a simple logistic growth linear regression on the raw data (we exclude time points that are within perturbation periods) and then inflating by 100 to make it more diffuse:

$$Y = \left[\frac{\log(x_{si}(k+1)) - \log(x_{si}(k))}{\Delta_k}\right] \tag{61}$$

$$X = \left[\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_{s1}(k) & & \\ & \ddots & \\ & & x_{sn_o}(k) \end{bmatrix}\right] \tag{62}$$

$$[a_1 \quad a_2] = (X^TX)^{-1}X^TY \tag{63}$$

$$\sigma^2_{\mathbf{a}_1} = 100\left(\frac{\sum_{i=1}^{n_o} a_{1,i}}{n_o}\right)^2 \tag{64}$$

The degrees of freedom is set to be diffuse ($\nu_{\mathbf{a}_1} = 2.5$) and the scale is set so that the mean of the prior is equal to the variance we calculated above:

$$\tau^2_{\mathbf{a}_1} = \frac{\sigma^2_{\mathbf{a}_1}(\nu_{\mathbf{a}_1} - 2)}{\nu_{\mathbf{a}_1}} \tag{65}$$

The initial value of the growth rates are all set to the mean:

$$\mathbf{a}_{1,i} = \mu_{\mathbf{a}_1} \quad i = 1, \ldots, n_o \tag{66}$$

4.2 Self-interactions

The initialization of the self-interactions is similar to that of the growth rates. We perform linear regression on the data conditional on the prior mean of the growth rates and then we set the variance similarly:

$$Y = \left[\frac{\log(x_{si}(k+1)) - \log(x_{si}(k))}{\Delta_k} - \mu_{\mathbf{a}_1} \right] \tag{67}$$

$$X = \begin{bmatrix} x_{s1}(k) & & \\ & \ddots & \\ & & x_{sn_o}(k) \end{bmatrix} \tag{68}$$

$$[a_2] = (X^T X)^{-1} X^T Y \tag{69}$$

$$\sigma^2_{\mathbf{a}_2} = 100 \left(\frac{\sum_{i=1}^{n_o} a_{2,i}}{n_o} \right)^2 \tag{70}$$

The degrees of freedom of the prior variance is set to be diffuse ($\nu_{\mathbf{a}_2} = 2.5$) and the scale is set so that the mean is equal to the variance we calculated above:

$$\tau^2_{\mathbf{a}_2} = \frac{\sigma^2_{\mathbf{a}_2}(\nu_{\mathbf{a}_2} - 2)}{\nu_{\mathbf{a}_2}} \tag{71}$$

The prior mean is set to the mean of the regressed values:

$$\mu_{\mathbf{a}_2} = \frac{\sum_{i=1}^{n_o} a_{2,i}}{n_o} \tag{72}$$

The initial value is set to the mean of the prior:

$$\mathbf{a}_{2,i} = \mu_{\mathbf{a}_2} \quad i = 1, \ldots, n_o \tag{73}$$

4.3 Cluster assignments and concentration

We initialize the clusters using agglomerative clustering with a spearman similarity. The number of clusters is set to the expected number of clusters of a dirichlet process ($n_c = \log(n_o)$). We set the prior of the concentration parameter θ to a Gamma distribution:

$$\theta \sim \text{Gamma}(10^{-5}, 10^5) \tag{74}$$

We set the initial value of θ to the mean of the prior.

4.4 Interactions and perturbations

We initialize the indicators of the interactions and perturbations to be all off. We set the prior of a perturbation and an interaction to be diffuse and a 50% chance of there being there:

$$b_{1,\mathbf{b}} = b_{2,\mathbf{b}} = b_{1,\gamma} = b_{2,\gamma} = 0.5 \tag{75}$$

The initial value of $\pi_{\gamma,p}$ and $\pi_{\mathbf{b}}$ is set to the mean of the prior.

The prior means of the perturbations and the interactions are set to 0:

$$\mu_{\mathbf{b}} = 0 \tag{76}$$
$$\mu_{\gamma} = 0 \tag{77}$$

The prior variance of the perturbations are all set to 10:

$$\sigma^2_{\gamma,p} = 10 \quad p = 1, \ldots, P \tag{78}$$

The prior variance of the interactions is set where it is diffuse and the mean is scaled down from the self-interactions mean:

$$\nu_{\mathbf{b}} = 2.5 \tag{79}$$
$$\tau^2_{\mathbf{b}} = \frac{\tau^2_{\mathbf{a}_2}}{10} \tag{80}$$
$$\sigma^2_{\mathbf{b}} = \mathbb{E}[\text{Scale-Inv-}\chi^2(\nu_{\mathbf{b}}, \tau^2_{\mathbf{b}})] \tag{81}$$

4.5 Latent trajectory

We initialize the latent trajectory with LOESS interpolation from the data and truncate the abundance to be above zero by sampling a tight coupling to the interpolation:

$$\mu_{si}(k) = \text{LOESS}(\mathbf{Q}, \mathbf{y}, s, i, k) \tag{82}$$
$$\mathbf{x}_{si}(k) \sim \text{Normal}\left(\mu_{si}(k), \sigma^2_q\right) \tag{83}$$
$$\sigma^2_q = 10^{-4}\mu^2_{si}(k) + 10^{-4} \tag{84}$$

The prior of the latent trajectory is set to be diffuse:

$$\log(\mathbf{x}_{si}(k)) \sim \text{Normal}(\log(10^7), 10^{10}) \tag{85}$$

A Derivation of marginalization

Let

$$y \sim \text{Normal}(Xw, \Sigma_1) \tag{86}$$

with the prior on w given as

$$w \sim \text{Normal}(\mu_2, \Sigma_2) \tag{87}$$

The design matrix X has dimensions $n \times d$.

$$p_{y|w}p_w = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}} \frac{1}{(2\pi)^{d/2}|\Sigma_2|^{1/2}} \exp\Big(-\frac{1}{2}(y - Xw)^T\Sigma_1^{-1}(y - Xw) - \frac{1}{2}(w - \mu_2)^T\Sigma_1^{-1}(w - \mu_2)\Big) \tag{88}$$

$$= \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}} \frac{1}{(2\pi)^{d/2}|\Sigma_2|^{1/2}} \tag{89}$$
$$\cdot \exp\big(-\frac{1}{2}(w^T(X^T\Sigma_1^{-1}X + \Sigma_2^{-1})w - (y^T\Sigma_1^{-1}X + \mu_2^T\Sigma_2^{-1})w)\big) \tag{90}$$
$$\cdot \exp\big(-\frac{1}{2}(-w^T(X^T\Sigma_1^{-1}y + \Sigma_2^{-1}\mu_2) + \mu_2^T\Sigma_2^{-1}\mu_2 + y^T\Sigma_1^{-1}y)\big) \tag{91}$$

Using the following definitions

$$\Sigma_3^{-1} = X^T\Sigma_1^{-1}X + \Sigma_2^{-1} \tag{92}$$

$$\mu_3 = \Sigma_3(X^T\Sigma_1^{-1}y + \Sigma_2^{-1}\mu_2) \tag{93}$$

we can simplify the above to

$$p_{y|w}p_w = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}}\frac{1}{(2\pi)^{d/2}|\Sigma_2|^{1/2}}\exp\big(-\frac{1}{2}(w^T\Sigma_3^{-1}w - \mu_3^T\Sigma_3^{-1}w - w^T\Sigma_3^{-1}\mu_3)\big) \tag{94}$$

$$\cdot\exp\big(-\frac{1}{2}(\mu_2^T\Sigma_2^{-1}\mu_2 + y^T\Sigma^{-1}y)\big) \tag{95}$$

Completing the square on the first line exponent

$$p_{y|w}p_w = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}}\frac{1}{(2\pi)^{d/2}|\Sigma_2|^{1/2}}\exp\big(-\frac{1}{2}(w-\mu_3)^T\Sigma_3^{-1}(w-\mu_3)\big) \tag{96}$$

$$\cdot\exp\big(-\frac{1}{2}(-\mu_3^T\Sigma_3^{-1}\mu_3 + \mu_2^T\Sigma_2^{-1}\mu_2 + y^T\Sigma_1^{-1}y)\big) \tag{97}$$

Before we integrate out w, we need to multiply and divide by $|\Sigma_3|^{1/2}$ to make this normal with respect to w

$$p_{y|w}p_w = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}}\frac{|\Sigma_3|^{1/2}}{|\Sigma_2|^{1/2}}\frac{1}{(2\pi)^{d/2}|\Sigma_3|^{1/2}}\exp\big(-\frac{1}{2}(w-\mu_3)^T\Sigma_3^{-1}(w-\mu_3)\big) \tag{98}$$

$$\cdot\exp\big(-\frac{1}{2}(-\mu_3^T\Sigma_3^{-1}\mu_3 + \mu_2^T\Sigma_2^{-1}\mu_2 + y^T\Sigma_1^{-1}y)\big) \tag{99}$$

Now we have an exact multivariate Gaussian $\frac{1}{(2\pi)^{d/2}|\Sigma_3|^{1/2}}\exp\big(-\frac{1}{2}(w-\mu_3)^T\Sigma_3^{-1}(w-\mu_3)\big)$. Now if we integrate out w the multivariate normal integral is 1 and

$$\int p_{y|w}p_w \mathrm{d}w = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}}\frac{|\Sigma_3|^{1/2}}{|\Sigma_2|^{1/2}}\exp\big(-\frac{1}{2}(\mu_2^T\Sigma_2^{-1}\mu_2 + y^T\Sigma_1^{-1}y - \mu_3^T\Sigma_3^{-1}\mu_3)\big) \tag{100}$$

B Learning negative binomial dispersion parameters

TODO

C Gibbs sampling of growth and self-interactions

C.1 Sampling growths

The prior of the growths are defined as

$$p(\mathbf{a}_{1,i}) \sim \text{TruncNormal}_{(0,\infty)}(\mu_{\mathbf{a}_1}, \sigma^2_{\mathbf{a}_1}) \tag{101}$$

Equation (101) can be rewritten as:

$$p(\mathbf{a}_{1,i}) = c\text{Normal}(\mu_{\mathbf{a}_1}, \sigma^2_{\mathbf{a}_1})\mathbf{1}\{\mathbf{a}_{1,i} > 0\} \tag{102}$$

where c is the normalizing constant which is independent of $\mathbf{a}_{1,i}$. The likelihood of the data conditional on all other parameters except for growth is as follows:

$$g_{si}(k) = \frac{\frac{\log(\mathbf{x}_{si}(k+1))-\log(\mathbf{x}_{si}(k))}{\Delta_k} - \mathbf{a}_{2,i}\mathbf{x}_{si}(k) - \sum_{\mathbf{c}_i \neq \mathbf{c}_j} \mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j} \mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} \mathbf{x}_{sj}(k)}{1 + \sum_{p=1}^{P} \gamma^{(p)}_{\mathbf{c}_i} \mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} h_p} \tag{103}$$

$$\sigma^2_g(s,i,k) = \frac{\sigma^2_w}{\Delta_k \left(1 + \sum_{p=1}^{P} \gamma^{(p)}_{\mathbf{c}_i} \mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} h_p\right)^2} \tag{104}$$

$$p(g_{si}(k)|\mathbf{a}_{1,i}) \sim \text{Normal}\left(\mathbf{a}_{1,i}, \sigma^2_g(s,i,k)\right) \tag{105}$$

Using Bayes rule, the posterior of $\mathbf{a}_{1,i}$ is as follows:

$$\mathbf{a}_{1,i}|g_i \propto p(\mathbf{a}_{1,i}) \prod_{s,k} p(g_{si}(k)|\mathbf{a}_{1,i}) \tag{106}$$

$$\propto \exp\left(-\frac{(\mathbf{a}_{1,i} - \mu_{\mathbf{a}_1})^2}{2\sigma^2_{\mathbf{a}_1}}\right) \mathbf{1}\{\mathbf{a}_{1,i} > 0\} \prod_{s,k} \exp\left(-\frac{(g_{si}(k) - \mathbf{a}_{1,i})^2}{2\sigma^2_g(s,i,k)}\right) \tag{107}$$

$$\propto \exp\left(-\frac{(\mathbf{a}_{1,i} - \mu_{\mathbf{a}_1})^2}{2\sigma^2_{\mathbf{a}_1}} - \sum_{s,k} \frac{(g_{si}(k) - \mathbf{a}_{1,i})^2}{2\sigma^2_g(s,i,k)}\right) \mathbf{1}\{\mathbf{a}_{1,i} > 0\} \tag{108}$$

$$\propto \exp\Bigg[- \sum_{s,k} \frac{1}{2\sigma^2_{\mathbf{a}_1}\sigma^2_g(s,i,k)/(\sigma^2_{\mathbf{a}_1} + \sigma^2_g(s,i,k))} \tag{109}$$

$$\times \left(\mathbf{a}_{1,i} - \frac{\sigma^2_g(s,i,k)\mu_{\mathbf{a}_1} + \sigma^2_{\mathbf{a}_1} g_{si}(k)}{\sigma^2_{\mathbf{a}_1} + \sigma^2_g(s,i,k)}\right) \Bigg] \mathbf{1}\{\mathbf{a}_{1,i} > 0\} \tag{110}$$

This is the kernel of the normal distribution with the usual mean and variance (as if we had done the derivation for an untruncated prior), but truncated greater than 0:

$$\mu'_{\mathbf{a}_{1,i}} = \frac{1}{\frac{1}{\sigma^2_{\mathbf{a}_1}} + \sum_{s,k} \frac{1}{\sigma^2_g(s,i,k)}} \left(\frac{\mu_{\mathbf{a}_1}}{\sigma^2_{\mathbf{a}_1}} + \sum_{s,k} \frac{g_{si}(k)}{\sigma^2_g(s,i,k)}\right) \tag{111}$$

$$\sigma^{2'}_{\mathbf{a}_{1,i}} = \left(\frac{1}{\sigma^2_{\mathbf{a}_1}} + \sum_{s,k} \frac{1}{\sigma^2_g(s,i,k)}\right)^{-1} \tag{112}$$

All you do is add the indicator function and adjust the normalizing constant.

C.2 Sampling self-interactions

Sampling the values of the self-interactions is very similar to sampling the values of the growths as described in appendix C.1 because the prior is a truncated normal distribution and the likelihood is a normal distribution. The only difference is the truncation ($(-\infty, 0)$ instead of $(0, \infty)$) and how we define $g_{si}(k)$ and $\sigma^2_g(s,i,k)$:

$$g_{si}(k) = \frac{\frac{\log(\mathbf{x}_{si}(k+1))-\log(\mathbf{x}_{si}(k))}{\Delta_k} - \mathbf{a}_{1,i}\left(1 + \sum_{p=1}^{P} \gamma^{(p)}_{\mathbf{c}_i} \mathbf{z}^{(\gamma,p)}_{\mathbf{c}_i} h_p\right) - \sum_{\mathbf{c}_i \neq \mathbf{c}_j} \mathbf{b}_{\mathbf{c}_i,\mathbf{c}_j} \mathbf{z}^{(\mathbf{b})}_{\mathbf{c}_i,\mathbf{c}_j} \mathbf{x}_{sj}(k)}{\mathbf{x}_{si}(k)} \tag{113}$$

$$\sigma^2_g(s,i,k) = \frac{\sigma^2_w}{\Delta_k \mathbf{x}^2_{si}(k)} \tag{114}$$

$$p(g_{si}(k)|\mathbf{a}_{2,i}) \sim \text{Normal}\left(\mathbf{a}_{2,i}, \sigma^2_g(s,i,k)\right) \tag{115}$$

Using the above definitions, we can do an analogous derivation that started at equation (106) to get the following posterior:

$$\mu'_{\mathbf{a}_2,i} = \frac{1}{\frac{1}{\sigma^2_{\mathbf{a}_2}} + \sum_{s,k} \frac{1}{\sigma^2_g(s,i,k)}} \left(\frac{\mu_{\mathbf{a}_2}}{\sigma^2_{\mathbf{a}_2}} + \sum_{s,k} \frac{g_{si}(k)}{\sigma^2_g(s,i,k)} \right) \tag{116}$$

$$\sigma^{2'}_{\mathbf{a}_2,i} = \left(\frac{1}{\sigma^2_{\mathbf{a}_2}} + \sum_{s,k} \frac{1}{\sigma^2_g(s,i,k)} \right)^{-1} \tag{117}$$

References

[1] A. Gelman, H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis Third Edition.* Chapman and Hall/CRC, 2013.

[2] R. M. Neal, "Markov chain sampling methods for dirchlet process mixture models," *Journal of computational and graphical models*, 9(2):249-265, 2000.

[3] C. E. Ramussen, "The infinite gaussian mixture model," *Advances in Information Processing Systems 12*, 2000.

[4] P.J. McMurdie and S. Holmes, "Waste not, want not: why rarefying microbiome data is inadmissible," *PLOS Computational Biology*, vol. 10, no. 4, p. e1003531, 2014.

[5] T. E. Gibson and G. K. Gerber, "Robust and scalable models of microbiome dynamcis", *International Conference on Machine Leanring*, pp. 1758-1767, 2018.

[6] B. Vanni, B. Tzen, N. Ling, M. Simmons, T. Tanoue, E. Bogart, L. Deng, V. Yeliseyev, M. L. Delaney, Q. Liu, B. Olle, R. R. Stein, K. Honda, L. Bry,G. K. Gerber, "Mdsine: Microbial dynamical systems inference engine for microbiome time-series analysis," *Genome Biology*, 17(1):121, 2016.

[7] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, no. 430, pp 577-588, 1995.