RVD2 Project Notes
November 8, 2013

RVD2.1: Independent variational parameters

Consider the following model

$$\mu_j \sim \text{Gaussian}(u_0, \sigma_0^2) \tag{1}$$
$$\theta_{kj} \sim \text{Beta}(\mu_j, M_0) \tag{2}$$
$$r_{kj} \sim \text{Bino}(\theta_{kj}, n_{kj}) \tag{3}$$

The inferential object of interest is the joint posterior distribution over the latent variables, $p(\mu, \theta | r)$. From this we can compute the marginal posterior distribution over the site-specific error rate - the null hypothesis distribution.

$$p(\mu|r) = \int_\theta p(\mu, \theta|r) d\theta \tag{4}$$

Writing out the joint posterior distribution gives

$$\begin{aligned}
\log p(\mu, \theta|r) &= \log \frac{p(\mu, \theta, r; u_o, \sigma_0^2, M_0)}{p(r|u_o, \sigma_0^2, M_0)} \\
&= \log p(\mu, \theta, r; u_o, \sigma_0^2, M_0) - \log p(r; u_o, \sigma_0^2, M_0) \\
&= \log \prod_{j=1}^{J} \prod_{k=1}^{K} p(\mu_j, \theta_{kj}, r_{kj}) - \log \prod_{j=1}^{J} \prod_{k=1}^{K} \int_{\theta_{kj}} \int_{\mu_j} p(\mu_j, \theta_{kj}, r_{kj}) d\mu_j d\theta_{kj}
\end{aligned} \tag{5}$$

$$\begin{aligned}
\log p(\mu, \theta|r) = K \sum_{j=1}^{J} \log p(\mu_j; \mu_0, \sigma_0^2) + \sum_{j=1}^{J} \sum_{k=1}^{K} \log p(\theta_{kj}|\mu_j; M_0) \\
+ \sum_{j=1}^{J} \sum_{k=1}^{K} \log p(r_{kj}|\theta_{kj}, n_{kj}) - \sum_{j=1}^{J} \sum_{k=1}^{K} \log \int_{\theta_{kj}} \int_{\mu_j} p(r_{kj}; \mu_0, \sigma_0^2, M_0)
\end{aligned} \tag{6}$$

C The computation of $p(r; u_0, \sigma_0^2, M_0)$ is intractable due to the coupling in the latent variables induced by the integral so we must use an approximation or sampling method.

Variational approximation

$$\log p(r_{kj}; u_0, \sigma_0^2, M_0) = \log \int_{\theta_{kj}} \int_{\mu_j} p(\mu_j, \theta_{kj}, r_{kj}) d\mu_j d\theta_{kj} \tag{7}$$
$$= \log \int_{\theta_{kj}} \int_{\mu_j} p(\mu_j, \theta_{kj}, r_{kj}) \frac{q(\mu_j, \theta_{kj})}{q(\mu_j, \theta_{kj})} d\mu_j d\theta_{kj} \tag{8}$$
$$\geq \int_{\theta_{kj}} \int_{\mu_j} q(\mu_j, \theta_{kj}) \log p(\mu_j, \theta_{kj}, r_{kj}) - \int_{\theta_{kj}} \int_{\mu_j} q(\mu_j, \theta_{kj}) \log q(\mu_j, \theta_{kj}) \tag{9}$$
$$= \mathbb{E}_q [\log p(r_{kj}, \mu_j, \theta_{kj})] - \mathbb{E}_q [\log q(\mu_j, \theta_{kj})] = \mathcal{L}_{kj} \tag{10}$$

The bound on the log likelihood of the data, $\mathcal{L}$, is seen to be the sum of the expected complete data log-likelihood and the entropy of the variational approximating distribution.

We now specify an averaging(variational) distribution

$$q(\mu_j|\gamma_{1j}, \gamma_{2j}) \sim \text{Gaussian}(\gamma_{1j}, \gamma_{2j}) \tag{11}$$
$$q(\theta_{kj}|\alpha_{kj}, \beta_{kj}) \sim \text{Beta}(\alpha_{kj}, \beta_{kj}) \tag{12}$$

We are now able to decompose $\mathcal{L}$ into its constituent parts

$$\mathcal{L} = K\sum_{j=1}^{J}\mathbb{E}_q \log p(\mu_j|u_0,\sigma_0^2) + \sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}_q \log p(\theta_{kj}|\mu_j;M_0) + \sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}_q \log p(r_{kj}|\theta_{kj},n_{kj}) - K\sum_{j=1}^{J}\mathbb{E}_q \log q(\mu) - \sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{E}_q \log q(\theta_{kj}) \quad (13)$$

Taking each term in (13) individually we can cast the bound on the log-likelihood in terms of original and variational parameters only.

$$\mathbb{E}_q \log p(\mu_j;u_0,\sigma_0^2) = -\frac{1}{2}\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\left[\gamma_{2j} + (\gamma_{1j}-u_0)^2\right] \quad (14)$$

$$\begin{aligned}\mathbb{E}_q \log p(\theta_{kj}|\mu_j;M_0) &= \int_{\mu_j} q(\mu_j)\int_{\theta_{kj}} q(\theta_{kj})\log p(\theta_{kj}|\mu_j,M_0)\\ &= \log\Gamma(M_0) - \mathbb{E}_q \log\Gamma(\mu_j M_0) - \mathbb{E}_q\log\Gamma((1-\mu_j)M_0)\\ &\quad + (\gamma_{1j}M_0 - 1)\left(\psi(\alpha_{kj}) - \psi(\alpha_{kj}+\beta_{kj})\right)\\ &\quad + ((1-\gamma_{1j})M_0 - 1)\left(\psi(\beta_{kj}) - \psi(\alpha_{kj}+\beta_{kj})\right)\end{aligned} \quad (15)$$

where $\mathbb{E}_q \log\Gamma(\mu_j M_0)$ is easily computed numerically. In general, we will be able to differentiate under the integral by Leibniz integral rule and the fact that the bounds are constants $\mu_j \in [0,1]$.

$$\begin{aligned}\mathbb{E}_q \log p(r_{kj}|\theta_{kj},n_{kj}) &= \int_{\theta_{kj}} q(\theta_{kj})\log p(r_{kj}|\theta_{kj},n_{kj})\\ &= \log\Gamma(n_{kj}+1) - \log\Gamma(r_{kj}+1) - \log\Gamma(n_{kj}-r_{kj}+1)\\ &\quad + r(\psi(\alpha_{kj}) - \psi(\alpha_{kj}+\beta_{kj})) + (n_{kj}-r_{kj})(\psi(\beta_{kj}) - \psi(\alpha_{kj}+\beta_{kj}))\end{aligned} \quad (16)$$

$$\mathbb{E}_q \log q(\mu_j) = -\frac{1}{2}\log(2\pi e\gamma_{2j}) \quad (17)$$

$$\begin{aligned}\mathbb{E}_q \log q(\theta_{kj}) &= \log\Gamma(\alpha_{kj}+\beta_{kj}) - \log\Gamma(\alpha_{kj}) - \log\Gamma(\beta_{kj})\\ &\quad + (\alpha_{kj}-1)\psi(\alpha_{kj}) + (\beta_{kj}-1)\psi(\beta_{kj}) - (\alpha_{kj}+\beta_{kj}-2)\psi(\alpha_{kj}+\beta_{kj})\end{aligned} \quad (18)$$

Now that (13) has been written in terms of only the variational and model parameters, we can numerically optimize the system with respect to variational and the model parameters and do coordinate ascent. Maximizing with respect to the variational parameters tightens the bound on the data log-likelihood and maximizing with respect to the model parameters maximizes the likelihood.

Update Equation for u_0

Isolating the terms of (13) that involve u_0 gives

$$\mathcal{L}_{[u_0]} = K\sum_{j=1}^{J}\left(-\frac{1}{2\sigma_0^2}(\gamma_{1j}-u_0)^2\right)$$

Taking the derivative and solving for u_0 gives

$$u_0 \leftarrow \frac{1}{J}\sum_{j=1}^{J}\gamma_{1j} \quad (19)$$

Update Equation for σ_0^2

Isolating the terms of (13) that involve σ_0^2 gives

$$\mathcal{L}_{[\sigma_0^2]} = -\frac{J}{2}\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\sum_{j=1}^{J}\left(\gamma_{2j} + (\gamma_{1j} - u_0)^2\right)$$

Taking the derivative with respect to σ_0^2 gives

$$\frac{\partial \mathcal{L}}{\partial \sigma_0^2} = -\frac{J}{2\sigma_0^2} + \frac{1}{\sigma_0^4}\sum_{j=1}^{J}\left(\gamma_{2j} + (\gamma_{1j} - u_0)^2\right)$$

Setting the derivative equal to zero and solving for σ_0^2 gives

$$\sigma_0^2 \leftarrow \frac{1}{J}\sum_{j=1}^{J}\left(\gamma_{2j} + (\gamma_{1j} - u_0)^2\right) \tag{20}$$

Update Equation for M_0
Isolating terms that involve M_0 gives

$$\begin{aligned}\mathcal{L}_{[M_0]} = \log\Gamma(M_0) - \mathbb{E}_q \log\Gamma(\mu_j M_0) - \mathbb{E}_q \log\Gamma((1-\mu_j)M_0) + \\ + \gamma_{1j}M_0(\psi(\alpha_{kj}) - \psi(\alpha_{kj} + \beta_{kj})) + (1-\gamma_{1j})M_0(\psi(\beta_{kj}) - \psi(\alpha_{kj} + \beta_{kj}))\end{aligned}$$

Taking the derivative with respect to M_0 gives

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial M_0} = \psi(M_0) - \frac{\partial}{\partial M_0}\mathbb{E}_q \log\Gamma(\mu_j M_0) - \frac{\partial}{\partial M_0}\mathbb{E}_q \log\Gamma((1-\mu_j)M_0) \\ + \gamma_{1j}(\psi(\alpha_{kj}) - \psi(\alpha_{kj} + \beta_{kj})) + (1-\gamma_{1j})(\psi(\beta_{kj}) - \psi(\alpha_{kj} + \beta_{kj}))\end{aligned}$$

The partial derivative can be taken inside of the integral which gives $\frac{\partial}{\partial M_0}\mathbb{E}_q \log\Gamma(\mu_j M_0) = \mathbb{E}_q\left[\mu_j \psi(\mu_j M_0)\right]$ and a similar expression for the second term $\frac{\partial}{\partial M_0}\mathbb{E}_q \log\Gamma((1-\mu_j)M_0) = \mathbb{E}_q\left[(1-\mu_j)\psi((1-\mu_j)M_0)\right]$.
Plugging these simplifications into the partial with respect to M_0 gives

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial M_0} = \psi(M_0) - \mathbb{E}_q\left[\mu_j \psi(\mu_j M_0)\right] - \mathbb{E}_q\left[(1-\mu_j)\psi((1-\mu_j)M_0)\right] \\ + \gamma_{1j}(\psi(\alpha_{kj}) - \psi(\alpha_{kj} + \beta_{kj})) + (1-\gamma_{1j})(\psi(\beta_{kj}) - \psi(\alpha_{kj} + \beta_{kj}))\end{aligned} \tag{21}$$

Update Equation for α_{kj} and β_{kj}
Isolating terms that involve α_{kj} gives

$$\begin{aligned}\mathcal{L}_{[\alpha_{kj}]} = (u_0 M_0 - 1)(\psi(\alpha_{kj}) - \psi(\alpha_{kj} + \beta_{kj})) - ((1-u_0)M_0 - 1)\psi(\alpha_{kj} + \beta_{kj}) \\ + r_{kj}\psi(\alpha_{kj}) - n_{kj}\psi(\alpha_{kj} + \beta_{kj}) \\ - \log\Gamma(\alpha_{kj} + \beta_{kj}) + \log\Gamma(\alpha_{kj}) - (\alpha_{kj} - 1)\psi(\alpha_{kj}) + (\alpha_{kj} + \beta_{kj} - 2)\psi(\alpha_{kj} + \beta_{kj})\end{aligned}$$

Setting the derivative equal to zero and solving for α_{kj} gives

$$\frac{\partial \mathcal{L}}{\partial \alpha_{kj}} = \psi_1(\alpha_{kj})(u_0 M_0 + r - \alpha_{kj}) - \psi_1(\alpha_{kj} + \beta_{kj})(M_0 + n - (\alpha_{kj} + \beta_{kj})) \tag{22}$$

A similar procedure holds for β_{kj} and the resulting update equation bears symmetry to that for α_{kj},

$$\frac{\partial \mathcal{L}}{\partial \beta_{kj}} = \psi_1(\beta_{kj})\left[(1-u_0)M_0 - 1 + (n_{kj} - r_{kj}) - \beta_{kj}\right] - \psi_1(\alpha_{kj} + \beta_{kj})(M_0 + n - (\alpha_{kj} + \beta_{kj})) \tag{23}$$

Update equation for γ_{1j}

Isolating terms that involve γ_{1j} gives

$$\mathcal{L}_{[\gamma_{1j}]} = -\frac{K}{2\sigma_0^2}(\gamma_{1j} - u_0)^2 - \mathbb{E}_q \log\Gamma(\mu_j M_0) - \mathbb{E}_q \log\Gamma((1-\mu_j)M_0) + \sum_{k=1}^{K} \gamma_{1j} M_0(\psi(\alpha_k j) - \psi(\alpha_{kj} + \beta_{kj})) - \sum_{k=1}^{K} \gamma_{1j} M_0(\psi(\beta_k j) - \psi(\alpha_{kj} + \beta_{kj}))$$

Taking the derivative with respect to γ_{1j} gives

$$\frac{\partial \mathcal{L}}{\partial \gamma_{1j}} = -\frac{K}{\sigma_0^2}(\gamma_{1j} - u_0) + M_0 \sum_{k=1}^{K} (\psi(\alpha_{kj}) - \psi(\beta_{kj})) - K\mathbb{E}_q\left[\frac{1}{\gamma_{2j}} \log\Gamma(\mu_j M_0)\right] - K\mathbb{E}_q\left[\frac{1}{\gamma_{2j}} \log\Gamma((1-\mu_j) M_0)\right] \tag{24}$$

Update Equation for γ_{2j}
Isolating terms that involve γ_{2j} gives

$$\mathcal{L}_{[\gamma_{2j}]} = -\frac{K}{2\sigma_0^2}\gamma_{2j} - \mathbb{E}_q \log\Gamma(\mu_j M_0) - \mathbb{E}_q \log\Gamma((1-\mu_j)M_0) + \frac{1}{2}\log(2\pi e \gamma_{2j})$$

Taking the derivative with respect to γ_{2j} gives

$$\frac{\partial \mathcal{L}}{\partial \gamma_{2j}} = -\frac{K}{2\sigma_0^2} - \frac{\partial}{\partial \gamma_{2j}}\mathbb{E}_q \log\Gamma(\mu_j M_0) - \frac{\partial}{\partial \gamma_{2j}}\mathbb{E}_q \log\Gamma((1-\mu_j)M_0) + \frac{1}{2\gamma_{2j}} \tag{25}$$

The two partials are very similar, so we show only the derivation of the first and apply the solution to the second replacing μ_j with $(1-\mu_j)$ in the gamma function.

$$\frac{\partial}{\partial \gamma_{2j}}\mathbb{E}_q \log\Gamma(\mu_j M_0) = \frac{\partial}{\partial \gamma_{2j}} \int_0^1 (2\pi\gamma_{2j})^{-1/2} \exp\left(-\frac{1}{2\gamma_{j2}}(\mu_j - \gamma_{1j})^2\right) \log\Gamma(\mu_j M_0) d\mu_j \tag{26}$$

$$= \frac{1}{2\gamma_{2j}}\left(\frac{1}{\gamma_{2j}}\mathbb{E}_q\left[(\mu_j - \gamma_{1j})^2 \log\Gamma(\mu_j M_0)\right] - \mathbb{E}_q\left[\log\Gamma(\mu_j M_0)\right]\right) \tag{27}$$

Replacing (26) in (25) gives

$$\frac{\partial \mathcal{L}}{\partial \gamma_{2j}} = -\frac{K}{2\sigma_0^2} - \frac{1}{2\gamma_{2j}}\left(\frac{1}{\gamma_{2j}}\mathbb{E}_q\left[(\mu_j - \gamma_{1j})^2 \log\Gamma(\mu_j M_0)\right] - \mathbb{E}_q\left[\log\Gamma(\mu_j M_0)\right]\right) - \frac{1}{2\gamma_{2j}}\left(\frac{1}{\gamma_{2j}}\mathbb{E}_q\left[(\mu_j - \gamma_{1j})^2 \log\Gamma((1-\mu_j) M_0)\right] - \mathbb{E}_q\left[\log\Gamma((1-\mu_j) M_0)\right]\right) + \frac{1}{2\gamma_{2j}} \tag{28}$$

Solving for the root gives the maximizing value of γ_{2j} and the update.

1 RVD 2.3: Gamma Prior

Consider the following model

$$\alpha_j \sim \text{Gamma}(a, b) \tag{29}$$
$$\theta_{ij} \sim \text{Dirichlet}(\alpha_j) \tag{30}$$
$$r_{ij} \sim \text{Multinomial}(\theta_{ij}, n_{ij}) \tag{31}$$

For each genomic location, $j = 1, \ldots, M$, K random variables are independently chosen from a Gamma distribution with shape parameter a and scale parameter b. Within each location, N Dirichlet random variables, θ_{ij} are chosen with parameter α_j for each replicate. Finally, a read count variable r_{ij} is drawn from a multinomial distribution with probability parameter θ_{ij} and total count n_{ij}. The count of nucleotide $k \in \{A, C, T, G\}$ at position j in replicate i is then r_{ij}^k.

The inferential object of interest here is the joint posterior distribution over the latent variables $p(\theta, \alpha | r)$. From the posterior distribution we can estimate the location-specific Dirichlet parameters $\hat{\alpha}_j$ and thus the location-specific nucleotide probabilities $\frac{\hat{\alpha}_j^k}{\hat{\alpha}_j^0}$, where $\hat{\alpha}_j^0 = \sum_k \hat{\alpha}_j^k$.

The joint posterior distribution is

$$p(\alpha, \theta | r; a, b) = \prod_{i=1}^{N} \prod_{j=1}^{M} \frac{p(r_{ij}|\theta_{ij}) p(\theta_{ij}|\alpha_j) p(\alpha_j; a, b)}{p(r_{ij}; a, b)}. \tag{32}$$

But, normalization factor in the posterior distribution,

$$p(r_{ij}; a, b) = \int_{\theta_{ij}} \int_{\alpha_j} p(r_{ij}|\theta_{ij}) p(\theta_{ij}|\alpha_j) p(\alpha_j; a, b), \tag{33}$$

is computationally intractable due to the coupling between θ_{ij} and α_j.

Instead of attempting to directly compute the posterior distribution, we instead sample it using a Metropolis-Hastings within Gibbs scheme.

Data: Counts r_{ij} for experimental replicate i at genomic location j
Result: MLE estimates for hyper-parameters $\hat{a}$ and $\hat{b}$. Samples from joint posterior distribution $p(\theta_{ij}, \alpha_j | r_{ij}; \hat{a}, \hat{b})$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$
Initialize a and b;
repeat
  Sample α_j from $p(\alpha_j | a, b, \theta_{1j}, \ldots, \theta_{Nj})$ using Metropolis-Hastings;
  Sample θ_{ij} from $p(\theta_{ij}|\alpha_j, r_{ij})$;
  Compute maximum-likelihood estimates $\hat{a}$ and $\hat{b}$
until *convergence*;

Algorithm 1: Metropolis-within-Gibbs Sampling for RVD2.3 Model

The maximum-likelihood estimates $\hat{a}$ and $\hat{b}$ are computed using generalized Newton-Raphson as outlined in http://research.microsoft.com/en-us/um/people/minka/papers/minka-gamma.pdf

Samples from $p(\theta_{ij}|\alpha_j, r_{ij})$ can be obtained directly because the Dirichlet is conjugate to the Multinomial. The posterior sampling distribution is

$$\theta_{ij} | \alpha_j, r_{ij} \sim \text{Dirichlet}(\alpha_j + r_{ij}). \tag{34}$$

The Gamma distribution is not conjugate to the Dirichlet, so we draw samples from $p(\alpha_j | a, b, \theta_{1j}, \ldots, \theta_{Nj})$ by Metropolis-Hastings. We use a proposal distribution,

$$\alpha_j^{k(\text{new})} | \alpha_j^k \sim \text{Gaussian}(\alpha_j^k, 0.05), \tag{35}$$

which we find mixes sufficiently fast to sample from the posterior after a burn-in of length X.