Gibbs sampling with an atomic prior

Elana J. Fertig

November 17, 2009

1 Description

We would like to sample from Skilling's atomic domain using Gibbs sampling. We will assume that we are seeking the mass of an atom α at A_{kl} for the $\mathbf{A}$ matrix and P_{lm} for the $\mathbf{P}$ matrix. The initial mass of this atom is α_0, which is 0 if we have decided to birth the atom and > 0 if we have decided to kill it. We retain this term so that we can derive the conditionals for birth and death in a single expression.

Determining the mass of α requires first computing the full conditional distribution $p\left(\alpha|\alpha_0, \mathbf{D}, \mathbf{A}, \mathbf{P}\right)$. To do this, we will first consider $P\left(\mathbf{A}, \mathbf{P}|\mathbf{D}\right)$ and examine the resulting distribution. We will begin by recalling that

$$p\left(\mathbf{A}, \mathbf{P}|\mathbf{D}\right) \propto p\left(\mathbf{D}|\mathbf{A}, \mathbf{P}\right) p\left(\mathbf{A}, \mathbf{P}\right). \tag{1}$$

Putting this in terms of an individual atom, we obtain

$$p\left(\alpha|\alpha_0, \mathbf{D}, \mathbf{A}, \mathbf{P}\right) \propto p\left(\mathbf{D}|\alpha, \alpha_0, \mathbf{A}, \mathbf{P}\right) p\left(\alpha\right). \tag{2}$$

In BD, the atomic prior assumes that

$$p\left(\mathbf{D}|\alpha, \alpha_0, \mathbf{A}, \mathbf{P}\right) \sim N(\mathbf{M}, \mathbf{\Sigma}), \tag{3}$$

where $\mathbf{M}$ is the mock data matrix given by the product of $\mathbf{A}$ and $\mathbf{P}$ that incorporates the change in mass of the atom $\alpha - \alpha_0$ in the updated term. $\mathbf{\Sigma}$ is the covariance matrix for $\mathbf{D}$. The prior for the mass of each atom α is given by an exponential with parameters λ_A and λ_P, respectively.

2 Single matrix element

2.1 Conditional distribution for atoms mapping to elements of A

We will first explore the likelihood in more detail, assuming that the mass of the atom maps to A_{kl}

$$p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) \propto \exp\left\{-\sum_i\sum_j \frac{1}{2\sigma_{ij}^2}\left(D_{ij}-\sum_p A_{ip}P_{pj}-\left(\alpha-\alpha_0\right)P_{lj}\right)^2\right\} \quad (4)$$

Since we are only concerned with computing the conditional for changes to A_{kl} we note that the other terms in $\mathbf{A}$ and $\mathbf{P}$ can be considered as parameters. As a result,

$$\begin{aligned} p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_j \frac{1}{2\sigma_{kj}^2}\left(D_{kj}-\sum_p A_{kp}P_{pj}-\left(\alpha-\alpha_0\right)P_{lj}\right)^2\right\} &(5)\\ &= \exp\left\{-\sum_j \frac{P_{lj}}{2\sigma_{kj}^2}\left(\alpha-\left(\frac{D_{kj}-\sum_p A_{kp}P_{pj}+\alpha_0 P_{lj}}{P_{lj}}\right)\right)^2\right\} &(6)\end{aligned}$$

Let $\mu_{klj}^A = \frac{D_{kj}-\sum_p A_{kp}P_{pj}+\alpha_0 P_{lj}}{P_{lj}}$ and $s_{klj}^A = \frac{P_{lj}^2}{2\sigma_{kj}^2}$. Then, Equation (6) becomes

$$\begin{aligned} p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_j s_{klj}^A\left(\alpha-\mu_{klj}^A\right)^2\right\} &(7)\\ &= \exp\left\{-\sum_j s_{klj}^A\left(\alpha^2-2\mu_{klj}^A\alpha+\mu_{klj}^{A2}\right)\right\} &(8)\\ &= \exp\left\{-\left(\alpha^2\sum_j s_{klj}^A-2\alpha\sum_j s_{klj}^A\mu_{klj}^A+\sum_j s_{klj}^A\mu_{klj}^{A2}\right)\right\} &(9)\\ &\propto \exp\left\{-\sum_j s_{klj}^A\left(\alpha^2-2\alpha\frac{\sum_j s_{klj}^A\mu_{klj}^A}{\sum_j s_{klj}^A}\right)\right\}. &(10)\end{aligned}$$

If we now incorporate the product with the exponential prior distribution for α,

$$\begin{aligned}
p\left(\alpha|\alpha_0,\mathbf{D},\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_j s_{klj}^A\left(\alpha^2-2\alpha\frac{\sum_j s_{klj}^A\mu_{klj}^A}{\sum_j s_{klj}^A}\right)\right\}\exp\left\{-\lambda_A\alpha\right\} && (11)\\
&= \exp\left\{-\sum_j s_{klj}^A\left(\alpha^2-\alpha\left(2\frac{\sum_j s_{klj}^A\mu_{klj}^A}{\sum_j s_{klj}^A}-\frac{\lambda_A}{\sum_j s_{klj}^A}\right)\right)\right\} && (12)\\
&\propto N\left(\frac{2\sum_j s_{klj}^A\mu_{klj}^A-\lambda_A}{2\sum_j s_{klj}^A},\frac{1}{\sqrt{2\sum_j s_{klj}^A}}\right). && (13)
\end{aligned}$$

2.2 Conditional distribution for P

Here, we consider atoms whose mass maps to elements P_{lm}. From the likelihood in Equation (4), we get

$$\begin{aligned}
p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_i\frac{1}{2\sigma_{im}^2}\left(D_{im}-\sum_p A_{ip}P_{pm}-\left(\alpha-\alpha_0\right)A_{il}\right)^2\right\} && (14)\\
&= \exp\left\{-\sum_i\frac{A_{il}}{2\sigma_{im}^2}\left(\alpha-\left(\frac{D_{im}-\sum_p A_{ip}P_{pm}+\alpha_0A_{il}}{A_{il}}\right)\right)^2\right\} && (15)
\end{aligned}$$

If $\mu_{ilm}^P=\frac{D_{im}-\sum_p A_{ip}P_{pm}+\alpha_0A_{il}}{A_{il}}$ and $s_{ilm}^P=\frac{A_{il}^2}{2\sigma_{im}^2}$,

$$\begin{aligned}
p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_i s_{ilm}^P\left(\alpha-\mu_{ilm}^P\right)^2\right\} && (16)\\
&= \exp\left\{-\sum_i s_{ilm}^P\left(\alpha^2-2\mu_{ilm}^P\alpha+\mu_{ilm}^{P2}\right)\right\} && (17)\\
&\propto \exp\left\{-\left(\sum_i s_{ilm}^P\right)\left(\alpha^2-\frac{2\sum_i\mu_{ilm}^P s_{ilm}^P\alpha}{\sum_i s_{ilm}^P}\right)\right\} && (18)
\end{aligned}$$

If we now incorporate the prior distribution for α

$$
\begin{aligned}
p\left(\alpha|\alpha_0,\mathbf{D},\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\left(\sum_i s^P_{ilm}\right)\left(\alpha^2-\left(\frac{2\sum_i \mu^P_{ilm}s^P_{ilm}}{\sum_i s^P_{ilm}}\right)\alpha\right)\right\}\exp\left\{-\lambda^P\alpha\right\} &(19)\\
&= \exp\left\{-\left(\sum_i s^P_{ilm}\right)\left(\alpha^2-\left(\frac{2\sum_i \mu^P_{ilm}s^P_{ilm}-\lambda^P}{\sum_i s^P_{ilm}}\right)\alpha\right)\right\} &(20)\\
&\propto N\left(\frac{2\sum_i \mu^P_{ilm}s^P_{ilm}-\lambda^P}{2\sum_i s^P_{ilm}},\frac{1}{\sqrt{2\sum_i s^P_{ilm}}}\right) &(21)
\end{aligned}
$$

3 Annealing parameter

We in fact wish to sample from the conditional

$$p\left(\alpha|\alpha_0,\mathbf{D},\mathbf{A},\mathbf{P}\right)\propto p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right)^{1/T}p\left(\alpha\right), \tag{22}$$

where T is the annealing temperature. This has the effect of multiplying the term σ in each of the equations by a factor of T. As a result, the standard deviation s of the above terms are the only things to change by as follows.

$$s^A_{klj}=\frac{P_{lj}}{2T\sigma^2_{kj}},\ \text{and} \tag{23}$$

$$s^P_{ilm}=\frac{A_{il}}{2T\sigma^2_{kj}}. \tag{24}$$

4 Mappings to multiple matrix elements

We assume there are K mappings, each of which is represented by M^k.

4.1 A - across multiple rows

If M^k maps across multiple rows, M^k is a row vector with the number of rows equaling the number of genes. That is M^k_i is the coefficient for the mass of the atom that is mapped to gene i in mapping k.

Assume that the new atom of mass α is in the bin that corresponds to map k and pattern l. In this case, Equation (4) becomes

$$p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) \propto \exp\left\{-\sum_i\sum_j\frac{1}{2\sigma_{ij}^2}\left(D_{ij}-\sum_p A_{ip}P_{pj}-(\alpha-\alpha_0)\,M_i^kP_{lj}\right)^2\right\} \quad (25)$$

$$= \sum_i\sum_j\frac{1}{2\sigma_{ij}^2}\left[M_i^kP_{lj}\alpha-\left(D_{ij}-\sum_p A_{ip}P_{pj}-\alpha_0M_i^kP_{lj}\right)\right]^2 \quad (26)$$

$$= \sum_i\sum_j\frac{\left(M_i^kP_{lj}\right)^2}{2\sigma_{ij}^2}\left[\alpha-\frac{D_{ij}-\sum_p A_{ip}P_{pj}+\alpha_0M_i^kP_{lj}}{M_i^kP_{lj}}\right]^2. \quad (27)$$

This is analogous to the previous expression in Section 2.1 where $\mu_{iklj}^A = \frac{D_{ij}-\sum_p A_{ip}P_{pj}+\alpha_0M_i^kP_{lj}}{M_i^kP_{lj}}$ and $s_{iklj}^A = \frac{\left(M_i^kP_{lj}\right)^2}{2\sigma_{ij}^2}$. If simulated annealing is employed, $s_{iklj}^A = \frac{\left(M_i^kP_{lj}\right)^2}{2T\sigma_{ij}^2}$. Following the previous analysis, this will lead to a conditional that is proportional to a normal distribution with mean $\frac{2\sum_i\sum_j s_{iklj}^A\mu_{iklj}^A-\lambda_A}{2\sum_i\sum_j s_{iklj}^A}$ and standard deviation $\frac{1}{\sqrt{2\sum_i\sum_j s_{iklj}^A}}$. Note, the sum over i should only be computed for those genes for which $M_i^k \neq 0$.

4.2 A - across multiple columns

In this case M^k maps across multiple columns. So, M^k is a column vector with the number of columns equaling the number of patterns. That is M_p^k is the coefficient for the mass of the atom that is mapped to pattern p in mapping k.

Assume that the new atom of mass α is in the bin that corresponds to map k and gene l. In this case, Equation (4) becomes

$$
\begin{aligned}
p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_j \frac{1}{2\sigma_{lj}^2}\left(D_{lj}-\sum_p\left\{A_{lp}P_{pj}-\left(\alpha-\alpha_0\right)M_p^kP_{pj}\right\}\right)^2\right\} && (28)\\
&= \sum_j \frac{1}{2\sigma_{lj}^2}\left[\sum_p M_p^kP_{pj}\alpha-\left(D_{lj}-\sum_p\left\{A_{lp}P_{pj}-\alpha_0M_l^kP_{pj}\right\}\right)\right]^2 && (29)\\
&= \sum_j \frac{\left(\sum_p M_p^kP_{pj}\right)^2}{2\sigma_{lj}^2}\left[\alpha-\frac{D_{lj}-\sum_p\left\{A_{lp}P_{pj}+\alpha_0M_p^kP_{pj}\right\}}{\sum_p M_p^kP_{pj}}\right]^2 && (30)
\end{aligned}
$$

This is analogous to the previous expression in Section 2.1 where $\mu_{klj}^A = \frac{D_{lj}-\sum_p\left\{A_{lp}P_{pj}+\alpha_0M_p^kP_{pj}\right\}}{\sum_p M_p^kP_{pj}}$ and $s_{klj}^A = \frac{\left(\sum_p M_p^kP_{pj}\right)^2}{2\sigma_{lj}^2}$. If simulated annealing is employed, $s_{iklj}^A = \frac{\left(\sum_p M_p^kP_{pj}\right)^2}{2T\sigma_{lj}^2}$. The remainder of the analysis follows as before.

4.3 P - across multiple rows

In this case M^k maps across multiple columns. So, M^k is a row vector with the number of rows equaling the number of patterns. That is M_p^k is the coefficient for the mass of the atom that is mapped to pattern p in mapping k.

Assume that the new atom of mass α is in the bin that corresponds to map k and sample m. In this case, Equation (4) becomes

$$
\begin{aligned}
p\left(\mathbf{D}|\alpha,\alpha_0,\mathbf{A},\mathbf{P}\right) &\propto \exp\left\{-\sum_i \frac{1}{2\sigma_{im}^2}\left(D_{im}-\sum_p\left\{A_{ip}P_{pm}-\left(\alpha-\alpha_0\right)M_p^kA_{ip}\right\}\right)^2\right\} && (31)\\
&= \sum_i \frac{1}{2\sigma_{im}^2}\left[\sum_p M_p^kA_{ip}\alpha-\left(D_{im}-\sum_p\left\{A_{ip}P_{pm}-\alpha_0M_l^kA_{ip}\right\}\right)\right]^2 && (32)\\
&= \sum_i \frac{\left(\sum_p M_p^kA_{ip}\right)^2}{2\sigma_{im}^2}\left[\alpha-\frac{D_{im}-\sum_p\left\{A_{ip}P_{pm}+\alpha_0M_p^kA_{ip}\right\}}{\sum_p M_p^kA_{ip}}\right]^2 && (33)
\end{aligned}
$$

This is analogous to the previous expression in Section 2.2 where $\mu^{p}_{ikl} = \frac{D_{im} - \sum_p \left\{ A_{ip} P_{pm} + \alpha_0 M^k_p A_{ip} \right\}}{\sum_p M^k_p A_{ip}}$ and $s^{P}_{klj} = \frac{\left(\sum_p M^k_p A_{ip} \right)^2}{2\sigma^2_{im}}$. If simulated annealing is employed, $s^{P}_{klj} = \frac{\left(\sum_p M^k_p A_{ip} \right)^2}{2T\sigma^2_{im}}$. The remainder of the analysis follows as before.

4.4 P - across multiple columns

In this case M^k maps across multiple columns. So, M^k is a column vector with the number of columns equaling the number of samples. That is M^k_m is the coefficient for the mass of the atom that is mapped to sample m in mapping k.

Assume that the new atom of mass α is in the bin that corresponds to map k and pattern n. In this case, Equation (4) becomes

$$
\begin{aligned}
p\left(\mathbf{D}|\alpha, \alpha_0, \mathbf{A}, \mathbf{P}\right) &\propto \exp\left\{ -\sum_i \sum_j \frac{1}{2\sigma^2_{ij}} \left(D_{ij} - \sum_p A_{ip} P_{pj} - \left(\alpha - \alpha_0\right) M^k_j A_{in} \right)^2 \right\} && (34) \\
&= \sum_i \sum_j \frac{1}{2\sigma^2_{ij}} \left[M^k_j A_{in} \alpha - \left(D_{ij} - \sum_p A_{ip} P_{pj} - \alpha_0 M^k_j A_{in} \right) \right]^2 && (35) \\
&= \sum_i \sum_j \frac{\left(M^k_j A_{in} \right)^2}{2\sigma^2_{ij}} \left[\alpha - \frac{D_{ij} - \sum_p A_{ip} P_{pj} + \alpha_0 M^k_j A_{in}}{M^k_j A_{in}} \right]^2 . && (36)
\end{aligned}
$$

This is analogous to the previous expression in Section 2.2 where $\mu^{p}_{ijkn} = \frac{D_{ij} - \sum_p A_{ip} P_{pj} + \alpha_0 M^k_j A_{in}}{M^k_j A_{in}}$ and $s^{P}_{ijkn} = \frac{\left(M^k_j A_{in} \right)^2}{2\sigma^2_{ij}}$. If simulated annealing is employed, $s^{P}_{ijkn} = \frac{\left(M^k_j A_{in} \right)^2}{2T\sigma^2_{ij}}$. The remainder of the analysis follows as before.