

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

Collaborative Place Models

Supplement 2

Anonymous Author(s)
Affiliation
Address
email

1 Bayesian Gaussian Mixture Model (BGMM)

In this section, we provide a formal description of BGMM and describe its generative process. The model is depicted in Figure 1.

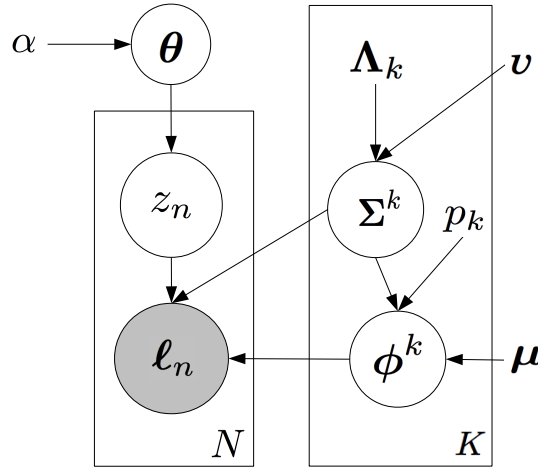


Figure 1: Graphical model representation of BGMM. The geographic coordinates, denoted by ℓ , are the only observed variables.

Let \mathcal{N} denote the normal distribution and IW denote the inverse-Wishart distribution. Let $\text{Dirichlet}_K(\cdot)$ denote a symmetric Dirichlet if its parameter is a scalar and a general Dirichlet if its parameter is a vector. The generative process of BGMM is described in more formal terms below. Further technical details about the distributions we use in our model can be found in the appendix.

1. Draw a place distribution $\theta \sim \text{Dirichlet}_K(\alpha)$.
2. For each place k ,
 - (a) Draw a place covariance $\Sigma^k \sim IW(\Lambda_k, v)$.
 - (b) Draw a place mean $\phi^k \sim \mathcal{N}(\mu, \frac{\Sigma^k}{p_k})$.
3. For each observation index n ,
 - (a) Draw a place assignment $z_n \sim \text{Categorical}(\theta)$.
 - (b) Draw a location $\ell_n \sim \mathcal{N}(\phi^{z_n}, \Sigma^{z_n})$.

2 Inference

In this section, we derive our inference algorithm, and we present our derivation in multiple steps. First, we use a strategy popularized by Griffiths and Steyvers [1], and derive a collapsed Gibbs sampler to sample from the posterior distribution of the categorical random variables conditioned on the observed geographic coordinates. Second, we derive the conditional likelihood of the posterior samples, which we use to determine the sampler’s convergence. Finally, we derive formulas for approximating the posterior expectations of the non-categorical random variables conditioned on the posterior samples.

2.1 Collapsed Gibbs Sampler

In Lemma 1, we derive the collapsed Gibbs sampler for variable z .

Given a vector \mathbf{x} and an index k , let \mathbf{x}_{-k} indicate all the entries of the vector excluding the one at index k . For Lemma 1, assume i denotes the index of the variable that will be sampled.

Lemma 1. *The unnormalized probability of z_i conditioned on the observed location data and remaining categorical variables is*

$$p(z_i = k \mid \mathbf{z}_{-i}, \boldsymbol{\ell}) \propto t_{\tilde{v}_k - 1} \left(\boldsymbol{\ell}_i \mid \tilde{\boldsymbol{\mu}}_k, \frac{\tilde{\boldsymbol{\Lambda}}^k (\tilde{p}_k + 1)}{\tilde{p}_k (\tilde{v}_k - 1)} \right) (\alpha + \tilde{m}_{:,i}^{k,\cdot}).$$

The parameters \tilde{v}_k , $\tilde{\boldsymbol{\mu}}_k$, $\tilde{\boldsymbol{\Lambda}}^k$, and \tilde{p}_k are defined in the proof. t denotes the bivariate t -distribution and $\tilde{m}_{:,i}^{k,\cdot}$ denotes counts, both of which are defined in the appendix.

Proof. We decompose the probability into two components using Bayes’ theorem:

$$\begin{aligned} p(z_i = k \mid \mathbf{z}_{-i}, \boldsymbol{\ell}) &= p(\boldsymbol{\ell}_i \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) \\ &\quad \times \frac{p(z_i = k \mid \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i})}{p(\boldsymbol{\ell}_i \mid \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i})} \\ &= p(\boldsymbol{\ell}_i \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) \\ &\quad \times \frac{p(z_i = k \mid \mathbf{z}_{-i})}{p(\boldsymbol{\ell}_i \mid \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i})} \\ &\propto p(\boldsymbol{\ell}_i \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) \\ &\quad \times p(z_i = k \mid \mathbf{z}_{-i}). \end{aligned} \tag{1}$$

In the first part of the derivation, we operate on (1). We augment it with $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$:

$$\begin{aligned} p(\boldsymbol{\ell}_i \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) &= \int \int p(\boldsymbol{\ell}_i \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}, \boldsymbol{\phi}^k, \boldsymbol{\Sigma}^k) \\ &\quad \times p(\boldsymbol{\phi}^k, \boldsymbol{\Sigma}^k \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) \, d\boldsymbol{\phi}^k \, d\boldsymbol{\Sigma}^k \\ &= \int \int p(\boldsymbol{\ell}_i \mid z_i = k, \boldsymbol{\phi}^k, \boldsymbol{\Sigma}^k) \\ &\quad \times p(\boldsymbol{\phi}^k, \boldsymbol{\Sigma}^k \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) \, d\boldsymbol{\phi}^k \, d\boldsymbol{\Sigma}^k \\ &= \int \int \mathcal{N}(\boldsymbol{\ell}_i \mid \boldsymbol{\phi}^k, \boldsymbol{\Sigma}^k) \\ &\quad \times p(\boldsymbol{\phi}^k, \boldsymbol{\Sigma}^k \mid z_i = k, \mathbf{z}_{-i}, \boldsymbol{\ell}_{-i}) \, d\boldsymbol{\phi}^k \, d\boldsymbol{\Sigma}^k. \end{aligned} \tag{3}$$

We convert (4) into a more tractable form. Let $\tilde{M}_{:,i}^{k,\cdot}$ be a set of indices, which we define in the appendix, and let $\boldsymbol{\ell}_{\tilde{M}_{:,i}^{k,\cdot}}$ denote the subset of observations whose indices are in $\tilde{M}_{:,i}^{k,\cdot}$. In the derivation

below, we treat all variables other than ϕ^k and Σ^k as a constant:

$$\begin{aligned}
p\left(\phi^k, \Sigma^k \mid z_i = k, \mathbf{z}_{-i}, \ell_{-i}\right) &= p\left(\phi^k, \Sigma^k \mid z_i = k, \mathbf{z}_{-i}, \ell_{\tilde{M}^k, \cdot}\right) \\
&= p\left(\phi^k, \Sigma^k \mid \mathbf{z}_{-i}, \ell_{\tilde{M}^k, \cdot}\right) \\
&= p\left(\ell_{\tilde{M}^k, \cdot} \mid \phi^k, \Sigma^k, \mathbf{z}_{-i}\right) \frac{p\left(\phi^k, \Sigma^k \mid \mathbf{z}_{-i}\right)}{p\left(\ell_{\tilde{M}^k, \cdot} \mid \mathbf{z}_{-i}\right)} \\
&\propto p\left(\ell_{\tilde{M}^k, \cdot} \mid \phi^k, \Sigma^k, \mathbf{z}_{-i}\right) p\left(\phi^k, \Sigma^k\right) \\
&= \left(\prod_{j \in \tilde{M}^k, \cdot} p\left(\ell_j \mid \phi^k, \Sigma^k, \mathbf{z}_{-i}\right)\right) p\left(\phi^k, \Sigma^k\right) \\
&= \left(\prod_{j \in \tilde{M}^k, \cdot} p\left(\ell_j \mid \phi^k, \Sigma^k, z_j\right)\right) p\left(\phi^k, \Sigma^k\right) \\
&= \left(\prod_{j \in \tilde{M}^k, \cdot} \mathcal{N}\left(\ell_j \mid \phi^k, \Sigma^k\right)\right) \mathcal{N}\left(\phi^k \mid \boldsymbol{\mu}, \frac{\Sigma^k}{p_k}\right) \\
&\quad \times IW\left(\Sigma^k \mid \boldsymbol{\Lambda}_k, \nu\right).
\end{aligned}$$

Since the normal-inverse-Wishart distribution is the conjugate prior of the multivariate normal distribution, the posterior is also a normal-inverse-Wishart distribution,

$$p\left(\phi^k, \Sigma^k \mid z_i = k, \mathbf{z}_{-i}, \ell_{-i}\right) = \mathcal{N}\left(\phi^k \mid \tilde{\boldsymbol{\mu}}_k, \frac{\Sigma^k}{\tilde{p}_k}\right) IW\left(\Sigma^k \mid \tilde{\boldsymbol{\Lambda}}^k, \tilde{\nu}_k\right), \quad (5)$$

whose parameters are defined as

$$\begin{aligned}
\tilde{p}_k &= p_k + \tilde{m}_{\cdot, \cdot}^{k, \cdot}, \\
\tilde{\nu}_k &= \nu + \tilde{m}_{\cdot, \cdot}^{k, \cdot}, \\
\tilde{\boldsymbol{\ell}}_k &= \frac{1}{\tilde{m}_{\cdot, \cdot}^{k, \cdot}} \sum_{j \in \tilde{M}^k, \cdot} \ell_j, \\
\tilde{\boldsymbol{\mu}}_k &= \frac{p_k \boldsymbol{\mu} + \tilde{m}_{\cdot, \cdot}^{k, \cdot} \tilde{\boldsymbol{\ell}}_k}{\tilde{p}_k}, \\
\tilde{S}^k &= \sum_{j \in \tilde{M}^k, \cdot} \left(\ell_j - \tilde{\boldsymbol{\ell}}_k\right) \left(\ell_j - \tilde{\boldsymbol{\ell}}_k\right)^T, \\
\tilde{\boldsymbol{\Lambda}}^k &= \boldsymbol{\Lambda}_k + \tilde{S}^k + \frac{p_k \tilde{m}_{\cdot, \cdot}^{k, \cdot}}{p_k + \tilde{m}_{\cdot, \cdot}^{k, \cdot}} \left(\tilde{\boldsymbol{\ell}}_k - \boldsymbol{\mu}\right) \left(\tilde{\boldsymbol{\ell}}_k - \boldsymbol{\mu}\right)^T.
\end{aligned}$$

The posterior parameters depicted above are derived based on the conjugacy properties of Gaussian distributions, as described in [2]. We rewrite (1) by combining (3), (4), and (5) to obtain

$$\begin{aligned}
p(\ell_i | z_i = k, \mathbf{z}_{-i}, \ell_{-i}) &= \int \int \mathcal{N}(\ell_i | \phi^k, \Sigma^k) \\
&\quad \times p(\phi^k, \Sigma^k | z_i = k, \mathbf{z}_{-i}, \ell_{-i}) d\phi^k d\Sigma^k \\
&= \int \int \mathcal{N}(\ell_i | \phi^k, \Sigma^k) \mathcal{N}\left(\phi^k | \tilde{\boldsymbol{\mu}}_k, \frac{\tilde{\boldsymbol{\Sigma}}^k}{\tilde{p}_k}\right) \\
&\quad \times IW(\Sigma^k | \tilde{\boldsymbol{\Lambda}}^k, \tilde{v}_k) d\phi^k d\Sigma^k \\
&= t_{\tilde{v}_k-1}\left(\ell_i | \tilde{\boldsymbol{\mu}}_k, \frac{\tilde{\boldsymbol{\Lambda}}^k (\tilde{p}_k + 1)}{\tilde{p}_k (\tilde{v}_k - 1)}\right), \tag{6}
\end{aligned}$$

where t is the bivariate t -distribution. (6) is derived by applying Equation 258 from [2].

Now, we move onto the second part of the derivation. We operate on (2) and augment it with $\boldsymbol{\theta}$:

$$\begin{aligned}
p(z_i = k | \mathbf{z}_{-i}) &= \int p(z_i = k | \mathbf{z}_{-i}, \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{z}_{-i}) d\boldsymbol{\theta} \\
&= \int p(z_i = k | \boldsymbol{\theta}) \\
&\quad \times p(\boldsymbol{\theta} | \mathbf{z}_{-i}) d\boldsymbol{\theta}. \tag{7}
\end{aligned}$$

We convert (8) into a more tractable form. As before, let $\tilde{M}_{:,i}$ be a set of indices, which we define in the appendix, and let $\mathbf{z}_{\tilde{M}_{:,i}}$ denote the subset of place assignments whose indices are in $\tilde{M}_{:,i}$. In the derivation below, we treat all variables other than $\boldsymbol{\theta}$ as a constant:

$$\begin{aligned}
p(\boldsymbol{\theta} | \mathbf{z}_{-i}) &= p(\boldsymbol{\theta} | \mathbf{z}_{\tilde{M}_{:,i}}) \\
&= \frac{p(\mathbf{z}_{\tilde{M}_{:,i}} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{z}_{\tilde{M}_{:,i}})} \\
&\propto \prod_{j \in \tilde{M}_{:,i}} p(z_j | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\
&= \prod_{j \in \tilde{M}_{:,i}} \text{Categorical}(z_j | \boldsymbol{\theta}) \text{Dirichlet}_K(\boldsymbol{\theta} | \boldsymbol{\alpha}) \\
\implies p(\boldsymbol{\theta} | \mathbf{z}_{-i}) &= \text{Dirichlet}_K(\boldsymbol{\theta} | \boldsymbol{\alpha} + \tilde{m}_{:,i}^1, \dots, \boldsymbol{\alpha} + \tilde{m}_{:,i}^K), \tag{9}
\end{aligned}$$

where the last step follows because Dirichlet distribution is the conjugate prior of the categorical distribution. We rewrite (2) by combining (7), (8), and (9):

$$\begin{aligned}
p(z_i = k | \mathbf{z}_{-i}) &= \int p(z_i = k | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{z}_{-i}) d\boldsymbol{\theta} \\
&= \int \theta_k \text{Dirichlet}_K(\boldsymbol{\theta} | \boldsymbol{\alpha} + \tilde{m}_{:,i}^1, \dots, \boldsymbol{\alpha} + \tilde{m}_{:,i}^K) d\boldsymbol{\theta} \\
&= \frac{\alpha + \tilde{m}_{:,i}^k}{K\alpha + \tilde{m}_{:,i}^{\cdot}}. \tag{10}
\end{aligned}$$

The last step follows because it is the expected value of the Dirichlet distribution.

Finally, we combine (1), (2), (6), and (10) to obtain the unnormalized probability distribution:

$$\begin{aligned}
p(z_i = k | \mathbf{z}_{-i}, \boldsymbol{\ell}) &\propto p(\ell_i | z_i = k, \mathbf{z}_{-i}, \ell_{-i}) \\
&\quad \times p(z_i = k | \mathbf{z}_{-i}) \\
&= t_{\tilde{v}_k-1}\left(\ell_i | \tilde{\boldsymbol{\mu}}_k, \frac{\tilde{\boldsymbol{\Lambda}}^k (\tilde{p}_k + 1)}{\tilde{p}_k (\tilde{v}_k - 1)}\right) \frac{\alpha + \tilde{m}_{:,i}^k}{K\alpha + \tilde{m}_{:,i}^{\cdot}}.
\end{aligned}$$

□

216
217

2.2 Likelihoods

218
219
220

In this subsection, we derive the conditional likelihoods of the posterior samples conditioned on the observed geographical coordinates. We use these conditional likelihoods to determine the sampler's convergence.

221
222

We present the derivations in multiple lemmas and combine them in a theorem at the end of the subsection. Let Γ denote the gamma function.

223
224

Lemma 2. *The marginal probability of the categorical random variable \mathbf{z} is*

225
226
227

$$p(\mathbf{z}) = \frac{\Gamma(K\alpha) \prod_{k=1}^K \Gamma(\alpha + m_{:,k}^{\cdot})}{\Gamma(\alpha)^K \Gamma(K\alpha + m_{:, \cdot}^{\cdot})},$$

228
229

where the counts $m_{:,k}^{\cdot}$ and $m_{:, \cdot}^{\cdot}$ are defined in the appendix.

230
231
232

Proof. Below, we will augment the marginal probability with $\boldsymbol{\theta}$, and then factorize it based on the conditional independence assumptions made by our model:

233
234
235
236
237
238
239
240
241

$$\begin{aligned} p(\mathbf{z}) &= \int p(\mathbf{z} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int p(\boldsymbol{\theta}) \left(\prod_j p(z_j | \boldsymbol{\theta}) \right) d\boldsymbol{\theta} \\ &= \int \left(\text{Dirichlet}_K(\boldsymbol{\theta} | \alpha) \prod_j \text{Categorical}(z_j | \boldsymbol{\theta}) \right) d\boldsymbol{\theta}. \end{aligned} \quad (11)$$

242
243

Now, we substitute the probabilities in (11) with Dirichlet and categorical distributions, which are defined in more detail in the appendix:

244
245
246
247
248
249
250
251
252
253
254
255
256
257

$$\begin{aligned} p(\mathbf{z}) &= \int \left(\text{Dirichlet}_K(\boldsymbol{\theta} | \alpha) \prod_j \text{Categorical}(z_j | \boldsymbol{\theta}) \right) d\boldsymbol{\theta} \\ &= \int \left(\frac{1}{B(\alpha)} \prod_{k=1}^K \theta_k^{\alpha-1} \right) \left(\prod_{k=1}^K \theta_k^{m_{:,k}^{\cdot}} \right) d\boldsymbol{\theta} \\ &= \int \left(\frac{1}{B(\alpha)} \prod_{k=1}^K \theta_k^{\alpha-1+m_{:,k}^{\cdot}} \right) d\boldsymbol{\theta} \\ &= \frac{1}{B(\alpha)} B(\alpha + m_{:,1}^{\cdot}, \dots, \alpha + m_{:,K}^{\cdot}) \\ &= \frac{\Gamma(K\alpha) \prod_{k=1}^K \Gamma(\alpha + m_{:,k}^{\cdot})}{\Gamma(\alpha)^K \Gamma(K\alpha + m_{:, \cdot}^{\cdot})}. \end{aligned}$$

258
259
260

□

261
262

For our final derivation, let Γ_2 denote the bivariate gamma function, and let $|\cdot|$ denote the determinant.

263
264

Lemma 3. *The conditional probability of the observed locations $\boldsymbol{\ell}$ conditioned on \mathbf{z} is*

265
266
267

$$p(\boldsymbol{\ell} | \mathbf{z}) = \prod_{k=1}^K \frac{\Gamma_2\left(\frac{\hat{v}_k}{2}\right) |\mathbf{\Lambda}_k|^{\frac{\nu}{2}} p_k}{\pi^{m_{:,k}^{\cdot}} \Gamma_2\left(\frac{\nu}{2}\right) \left|\hat{\mathbf{\Lambda}}^k\right|^{\frac{\hat{v}_k}{2}} \hat{p}_k}.$$

268
269

The parameters \hat{v}_k , $\hat{\mathbf{\Lambda}}^k$, and \hat{p}_k are defined in the proof, and the counts $m_{:,k}^{\cdot}$ are defined in the appendix.

270 *Proof.* We will factorize the probability using the conditional independence assumptions made by
 271 the model, and then simplify the resulting probabilities by integrating out the means and covariances
 272 associated with the place clusters:
 273

$$\begin{aligned}
 274 \quad p(\ell | z) &= \prod_{k=1}^K p(\ell_{M^{k,\cdot}} | z) \\
 275 &= \prod_{k=1}^K \int \int p(\ell_{M^{k,\cdot}} | z, \phi^k, \Sigma^k) p(\phi^k, \Sigma^k) d\phi^k d\Sigma^k \\
 276 &= \prod_{k=1}^K \int \int p(\phi_u^k, \Sigma_u^k) \prod_{j \in M^{k,\cdot}} p(\ell_j | z_j, \phi^k, \Sigma^k) d\phi^k d\Sigma^k \\
 277 &= \prod_{k=1}^K \int \int \mathcal{N}\left(\phi^k | \mu, \frac{\Sigma^k}{p_k}\right) IW(\Sigma^k | \Lambda_k, v) \\
 278 &\quad \times \prod_{j \in M^{k,\cdot}} \mathcal{N}(\ell_j | \phi^k, \Sigma^k) d\phi^k d\Sigma^k. \tag{12} \\
 279 & \\
 280 & \\
 281 & \\
 282 & \\
 283 & \\
 284 & \\
 285 & \\
 286 & \\
 287 & \\
 288 &
 \end{aligned}$$

288 We apply Equation 266 from [2], which describes the conjugacy properties of Gaussian distributions,
 289 to reformulate (12) into its final form:
 290

$$\begin{aligned}
 291 \quad p(\ell | z) &= \prod_{k=1}^K \int \int \mathcal{N}\left(\phi^k | \mu, \frac{\Sigma^k}{p_k}\right) IW(\Sigma^k | \Lambda_k, v) \prod_{j \in M^{k,\cdot}} \mathcal{N}(\ell_j | \phi^k, \Sigma^k) d\phi^k d\Sigma^k \\
 292 &= \prod_{k=1}^K \frac{\Gamma_2\left(\frac{\hat{v}_k}{2}\right) |\Lambda_k|^{\frac{v}{2}} p_k}{\pi^{m_{k,\cdot}} \Gamma_2\left(\frac{v}{2}\right) |\hat{\Lambda}^k|^{\frac{\hat{v}_k}{2}} \hat{p}_k}. \\
 293 & \\
 294 & \\
 295 & \\
 296 & \\
 297 &
 \end{aligned}$$

298 The definitions for \hat{v}_k , $\hat{\Lambda}^k$, and \hat{p}_k are provided in (5). □
 299

300 Finally, we combine Lemmas 2 and 3 to provide the log-likelihood of the samples z conditioned on
 301 the observations ℓ .

302 **Lemma 4.** *The log-likelihood of the samples z conditioned on the observations ℓ is*
 303

$$\begin{aligned}
 304 \quad \log p(z | \ell) &= \left(\sum_{k=1}^K \log \Gamma(\alpha + m_{k,\cdot}^{k,\cdot}) \right) \\
 305 &\quad + \left(\sum_{k=1}^K \left(\log \Gamma_2\left(\frac{\hat{v}_k}{2}\right) - m_{k,\cdot}^{k,\cdot} \log \pi - \frac{\hat{v}_k}{2} \log |\hat{\Lambda}^k| - \log \hat{p}_k \right) \right) + C, \\
 306 & \\
 307 & \\
 308 & \\
 309 &
 \end{aligned}$$

310 where C denotes the constant terms.

311 *Proof.* The result follows by multiplying the probabilities stated in Lemmas 2 and 3, and applying
 312 the logarithm function. □
 313

314 2.3 Parameter estimation

315 In Subsection 2.1, we described a collapsed Gibbs sampler for sampling the posteriors of the cate-
 316 gorical random variables. Below, Lemmas 5 and 6 show how these samples, denoted as z , can be
 317 used to approximate the posterior expectations of θ , ϕ , and Σ .
 318

319 **Lemma 5.** *The expectation of θ given the observed geographical coordinates and the posterior*
 320 *samples is*

$$321 \quad \hat{\theta}_k = \mathbb{E}[\theta_k | z, \ell] = \frac{\alpha + m_{k,\cdot}^{k,\cdot}}{K\alpha + m_{k,\cdot}^{k,\cdot}},$$

322 where the counts $m_{k,\cdot}^{k,\cdot}$ and $m_{k,\cdot}^{k,\cdot}$ are defined in the appendix.
 323

324 *Proof.*

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

$$\begin{aligned}
 p(\boldsymbol{\theta} | \mathbf{z}, \ell) &= p(\boldsymbol{\theta} | \mathbf{z}) \\
 &= \frac{p(\mathbf{z} | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{z})} \\
 &= \frac{\prod_j p(z_j | \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{z})} \\
 &\propto \text{Dirichlet}_K(\boldsymbol{\theta} | \alpha) \prod_j \text{Categorical}(z_j | \boldsymbol{\theta}) \\
 &= \text{Dirichlet}_K(\boldsymbol{\theta} | \alpha + m_{:,1}^1, \dots, \alpha + m_{:,K}^K) \\
 \implies \hat{\theta}_k &= \mathbb{E}[\theta_k | \mathbf{z}, \ell] = \frac{\alpha + m_{:,k}^k}{K\alpha + m_{:,k}^k}.
 \end{aligned}$$

□

Lemma 6. *The expectations of ϕ and Σ given the observed geographical coordinates and the posterior samples is*

$$\hat{\phi}_u^k = \mathbb{E}[\phi_u^k | \mathbf{z}, \ell] = \hat{\mu}_k^u$$

and

$$\hat{\Sigma}_u^k = \mathbb{E}[\Sigma_u^k | \mathbf{z}, \ell] = \frac{\hat{\Lambda}_u^k}{\hat{v}_k^u - 3}.$$

Parameters $\hat{\mu}_k^u$, $\hat{\Lambda}_u^k$, and \hat{v}_k^u are defined in the proof of Lemma 1.

378

379 *Proof.*

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

$$\begin{aligned}
p(\phi^k, \Sigma^k | z, \ell) &= p(\phi^k, \Sigma^k | z, \ell) \\
&= p(\phi^k, \Sigma^k | z, \ell_{M_{:,j}^k}) \\
&= \frac{p(\ell_{M_{:,j}^k} | \phi^k, \Sigma^k, z) p(\phi^k, \Sigma^k | z)}{p(\ell_{M_{:,j}^k} | z)} \\
&= \frac{\prod_{j \in M_{:,j}^k} p(\ell_j | \phi^k, \Sigma^k, z) p(\phi^k, \Sigma^k)}{p(\ell_{M_{:,j}^k} | z)} \\
&= \frac{\mathcal{N}(\phi^k | \mu, \frac{\Sigma^k}{p_k}) IW(\Sigma^k | \Lambda_k, v) \prod_{j \in M_{:,j}^k} \mathcal{N}(\ell_j | \phi^k, \Sigma^k)}{p(\ell_{M_{:,j}^k} | z)} \\
&\propto \mathcal{N}(\phi^k | \mu, \frac{\Sigma^k}{p_k}) IW(\Sigma^k | \Lambda_k, v) \prod_{j \in M_{:,j}^k} \mathcal{N}(\ell_j | \phi^k, \Sigma^k) \\
\Rightarrow p(\phi^k, \Sigma^k | z, \ell) &= \mathcal{N}(\phi^k | \hat{\mu}_k, \frac{\Sigma^k}{\hat{p}_k}) IW(\Sigma^k | \hat{\Lambda}^k, \hat{v}_k) \\
\Rightarrow p(\phi^k | z, \ell) &= t_{\hat{v}_k - 1} \left(\phi^k | \hat{\mu}_k, \frac{\hat{\Lambda}^k}{\hat{p}_k (\hat{v}_k - 1)} \right) \\
\Rightarrow \hat{\phi}^k = \mathbb{E}[\phi^k | z, \ell] &= \hat{\mu}_k \\
\Rightarrow p(\Sigma^k | z, \ell) &= IW(\Sigma_u^k | \hat{\Lambda}^k, \hat{v}_k) \\
\Rightarrow \hat{\Sigma}^k = \mathbb{E}[\Sigma^k | z, \ell] &= \frac{\hat{\Lambda}^k}{\hat{v}_k - 3}.
\end{aligned}$$

□

413

414

3 Appendix

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

3.1 Miscellaneous notation

Throughout the paper, we use various notations to represent sets of indices and their cardinalities. Vectors \mathbf{y} and \mathbf{z} denote the component and place assignments in CPM, respectively. Each vector entry is identified by a tuple index (u, w, n) , where $u \in \{1, \dots, U\}$ is a user, $w \in \{1, \dots, W\}$ is a weekhour, and $n \in \{1, \dots, N_{u,w}\}$ is an iteration index.

For the subsequent notations, we assume that the random variables \mathbf{y} and \mathbf{z} are already sampled. We refer to a subset of indices using

$$M_{u_0, w_0}^{k_0, f_0} = \{(\dot{u}, \dot{w}, \dot{n}) \mid z_{\dot{u}, \dot{w}, \dot{n}} = k_0, y_{\dot{u}, \dot{w}, \dot{n}} = f_0, \dot{u} = u_0, \dot{w} = w_0\},$$

where u_0 denotes the user, w_0 denotes the weekhour, k_0 denotes the place, and f_0 denotes the component. If we want the subset of indices to be unrestricted with respect to a category, we use the placeholder “.”. For example,

$$M_{u_0, w_0}^{., f_0} = \{(\dot{u}, \dot{w}, \dot{n}) \mid y_{\dot{u}, \dot{w}, \dot{n}} = f_0, \dot{u} = u_0, \dot{w} = w_0\}$$

has no constraints with respect to places.

Given a subset of indices denoted by M , the lowercase $m = |M|$ denotes its cardinality. For example, given a set of indices

$$M_{u_0, \cdot}^{:,f_0} = \{(\dot{u}, \dot{w}, \dot{n}) \mid z_{\dot{u}, \dot{w}, \dot{n}} = \cdot, y_{\dot{u}, \dot{w}, \dot{n}} = f_0, \dot{u} = u_0, \dot{w} = w_0\},$$

its cardinality is

$$m_{u_0, \cdot}^{:,f_0} = |M_{u_0, \cdot}^{:,f_0}|.$$

For the collapsed Gibbs sampler, the sets of indices and cardinalities used in the derivations exclude the index that will be sampled. We use “ \sim ” to modify sets or cardinalities for this exclusion. Let (u, w, n) denote the index that will be sampled, then given an index set M , let $\tilde{M} = M - \{(u, w, n)\}$ represent the excluding set and let $\tilde{m} = |\tilde{M}|$ represent the corresponding cardinality. For example,

$$\tilde{M}_{u_0, \cdot}^{:,f_0} = M_{u_0, \cdot}^{:,f_0} - \{(u, w, n)\}$$

and

$$\tilde{m}_{u_0, \cdot}^{:,f_0} = |\tilde{M}_{u_0, \cdot}^{:,f_0}|.$$

In the proof of Lemma 1, parameters \tilde{v}_k^u , $\tilde{\mu}_k^u$, $\tilde{\Lambda}_u^k$, and \tilde{p}_k^u are defined using cardinalities that exclude the current index (u, w, n) . Similarly, in the proof of Lemma 6, parameters $\hat{\mu}_k^u$, $\hat{\Lambda}_u^k$, and \hat{v}_k^u are defined like their wiggly versions, but the counts used in their definitions do not exclude the current index.

3.2 Probability distributions

Let Γ_2 denote a bivariate gamma function, defined as

$$\Gamma_2(a) = \pi^{\frac{1}{2}} \prod_{j=1}^2 \Gamma\left(a + \frac{1-j}{2}\right).$$

Let $\nu > 1$ and let $\Lambda \in \mathbb{R}^{2 \times 2}$ be a positive definite scale matrix. The inverse-Wishart distribution, which is the conjugate prior to the multivariate normal distribution, is defined as

$$IW(\Sigma \mid \Lambda, \nu) = \frac{|\Lambda|^{\frac{\nu}{2}}}{2^\nu \Gamma_2\left(\frac{\nu}{2}\right)} |\Sigma|^{-\frac{\nu-3}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Lambda \Sigma^{-1})\right).$$

Let $\Sigma \in \mathbb{R}^{2 \times 2}$ be a positive definite covariance matrix and let $\mu \in \mathbb{R}^2$ denote a mean vector. The multivariate normal distribution is defined as

$$\mathcal{N}(\ell \mid \mu, \Sigma) = (2\pi)^{-1} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\ell - \mu)^T \Sigma^{-1} (\ell - \mu)\right).$$

Let $\nu > 1$ and let $\Sigma \in \mathbb{R}^{2 \times 2}$, then the 2-dimensional t -distribution is defined as

$$t_\nu(x \mid \mu, \Sigma) = \frac{\Gamma\left(\frac{\nu}{2} + 1\right) |\Sigma|^{-\frac{1}{2}}}{\Gamma\left(\frac{\nu}{2}\right) \nu \pi} \left(1 + \frac{1}{\nu} (x - \mu)^T \Sigma^{-1} (x - \mu)\right)^{-\frac{\nu}{2} - 1}.$$

Let $K > 1$ be the number of categories and let $\alpha = (\alpha_1, \dots, \alpha_K)$ be the concentration parameters, where $\alpha_k > 0$ for all $k \in \{1, \dots, K\}$. Then, the K -dimensional Dirichlet distribution, which is the conjugate prior to the categorical distribution, is defined as

$$\text{Dirichlet}_K(\mathbf{x} \mid \alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^K x_k^{\alpha_k - 1},$$

where

$$B(\alpha) = \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^K \alpha_k\right)}.$$

We abuse the Dirichlet notation slightly and use it to define the K -dimensional *symmetric* Dirichlet distribution as well. Let $\beta > 0$ be a scalar concentration parameter. Then, the symmetric Dirichlet distribution is defined as

$$\text{Dirichlet}_K(\mathbf{x} \mid \beta) = \text{Dirichlet}_K(\mathbf{x} \mid \alpha_1, \dots, \alpha_K),$$

where $\beta = \alpha_k$ for all $k \in \{1, \dots, K\}$.

486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539

References

- [1] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5228–5235, April 2004.
- [2] Kevin Murphy. Conjugate bayesian analysis of the gaussian distribution. October 2007.