

统计学习：第二十章

1 推导狄利克雷分布数学期望公式。

答：

$$\begin{aligned} E(\theta_i) &= \int_0^1 \int_{\sum_{j \neq i} \theta_j = 1 - \theta_i} \frac{1}{B(\alpha)} \theta_i^{\alpha_i} \prod_{j=1}^k \theta_j^{\alpha_j - 1} d\theta \\ &= \int_{\sum_j \theta_j = 1} \frac{1}{B(\alpha)} \theta_i^{\alpha_i} \prod_{j=1}^k \theta_j^{\alpha_j - 1} d\theta \\ &= \frac{\Gamma(\alpha_i + 1) \Gamma(\sum_i \alpha_i)}{\Gamma(\alpha_i) \Gamma(\sum_i \alpha_i + 1)} \\ &= \frac{\alpha_i}{\sum_i \alpha_i} \end{aligned}$$

2 针对 17.2.2 的文本例子，使用 LDA 模型进行话题分析。

答：见代码

3 找出 LDA 的吉布斯抽样算法、变分 EM 算法中利用到狄利克雷分布的部分，思考 LDA 中使用狄利克雷分布的重要性。

答：吉布斯抽样算法中话题关于文本的分布 $p(z|\mathbf{w}, \alpha, \beta)$ 和后验计算；变分 EM 算法中导数和期望的计算。

由于狄利克雷分布是多项分布的共轭分布，简化了计算，提供了解析计算公式，以及通过先验概率控制了数据的稀松问题。

4 给出 LDA 的吉布斯抽样算法和变分 EM 算法的算法复杂度。

答：吉布斯抽样的算法复杂度为 $O(mnkv)$ ，变分 EM 算法的算法复杂度为 $O(\max(n^2 k^2, k^3 + mk + mn))$

5 证明变分 EM 算法收敛。

答：由于迭代过程中 ELBO 单调递增，且存在上界 $\log p(x)$ ，由单调有界定理得到，算法收敛。