认知建模基础作业3

> **Note**
>
> 1. 本次作业有三道题，总分按100分计，请在**JupyterLab**中使用 *python* 或者 *R* 在本文档中完成本次作业。
> 2. 请在做答时描述你的解题思路，并附上相应代码，我们将会根据结果的正确及清晰程度进行评分。
>   - 请确保文字结果及代码清晰、易读。
>   - 代码应有简要注释以及运行结果，运行结果应有相应的说明。请注意：注释允许使用中文/英文，可以使用co-pilot自动生成的注释，但要求对关键语句进行详细的注释，包括但不限于选择该语句的原因、参数设置的原因等。**代码的注释会纳入评分依据**
>   - 图片结果应当有相应的标题。
> 3. 最终提交时请重命名此ipynb文件。将运行好的结果导出成html，并将相应的文件打包上传，文件命名格式请参照 “作业N_编程语言_本科生/研究生_姓名_学号 ”，使用英文和拼音，例：“homework3_python_undergraduate_zhangsan_2101111111”(本科生)或者 “homework3_python_postgraduate_zhangsan_2101111111”(研究生)。

---

1 Population Coding

上次作业中我们探究了调谐曲线（tuning curve），即神经元对不同刺激的反应模式（称为encoding过程），本题中我们关心已知调谐曲线和神经元的反应后，解码真实刺激的内容（称为decoding内容），接下来我们假设神经元真实反应模式可用冯·米塞斯（von Mises）曲线进行表示：

$$\lambda(\theta) = 100 * \exp(\cos(\theta - \theta_0) - 1)$$

其中 θ 是真实刺激朝向，θ_0 是该神经元偏好朝向，而实际观测到的放电次数 n 服从泊松分布：

$$p(n = k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

其中n是放电次数，而 λ 是神经元真实的反应强度。

1. 假设某个神经元真实反应强度为 $\lambda = \frac{100}{\sqrt{e}}$，其最偏好的朝向为 $\theta_0 = \pi$，假设 θ 的先验分布为 $[0, 2\pi]$之间的均匀分布，请使用贝叶斯公式推导出 θ 的后验分布 $p(\theta|\lambda)$（5分）；
2. 真实反应强度无法直接观测得到，假设对（1）相同的神经元，我们观测到的放电次数为70。请在 θ 的先验分布不变的前提下，请画出概率图模型，使用MCMC采样的方法，画出

θ 的后验分布 $p(\theta|n)$，并说明本例能否使用后验均值作为 θ 的估计值（10分）；

3. 概率群体编码理论认为，真实刺激的朝向是多个神经元共同编码的，`tr.csv`记录了中颞区皮层（MT）中100个神经元对**同一刺激**的反应，`ori`表示该神经元最偏好的朝向（用弧度制表示），`spike`列记录的观测到的放电次数，根据 100 个神经元的放电情况，画出 θ 后验分布 $p(\theta|n)$，并报告后验均值和95%HDI（10分）；
4. 请你根据上一问中 θ 的后验分布进行预测，对于（1）中最偏好朝向为0度的神经元，绘制其真实反应强度和记录到放电次数的后验预测分布，并比较他们的均值。（10分，提示，你可以手动将这个神经元添加到上一问的图模型中获得他们的后验分布）；
5. 你认为参与编码的神经元数量对刺激编码的效果有怎样的影响，尝试用现有数据证明你的猜想（5分）。

---

2 Why We Use MCMC

> **Note**
>
> **完整贝叶斯公式**如下所示：
>
> $$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta')p(\theta')d\theta'} \tag{1}$$
>
> 分母项是一个常数，需要进行积分计算，由于我们并不一定知道其值，上式也常常写成
>
> $$p(\theta|x) \propto p(x|\theta)p(\theta) \tag{2}$$
>
> 这样做的一个好处在于，如果likelihood ($p(x|\theta)$)和先验的形式我们都知道，那么一定情况下后验分布仍然满足某一分布。在张老师给大家的课件`BayesUpdateCoin.ipynb`中就利用了这种特性。在对应代码中，似然函数是`二项分布`, 先验分布是`beta分布`，而`beta分布`恰好是`二项分布`的共轭分布(which means 后验分布也是`beta分布`的形式，证明可参考`lecture4 slides中27页`)。尽管共轭分布为计算后验提供了一种方式，对于一些概率分布我们较难找到其共轭分布，而直接计算分母项在低维空间中则是容易的。，

1. (低维贝叶斯计算) 请你用**完整贝叶斯公式**来计算下面的例子：假设扔一枚硬币，其正面朝上概率为p，不同投掷彼此之间独立。p 的先验概率分布为 $\mathrm{Beta}(2,3)$, 现在投了 10 次，有 7 次正面朝上，请你按照如下步骤，利用贝叶斯公式(2)计算 p 的后验概率并进行可视化（10分）。
   - 定义先验分布 $p(\theta)$ 和似然函数 $p(x|\theta)$；
   - 计算积分 $\int p(x|\theta)p(\theta)d\theta$，可能会用到scipy.integrate（for Python）或integrate（for R）；
   - 可视化先验分布 $p(\theta)$、似然函数 $p(x|\theta)$ 和（归一化后的）后验分布 $p(\theta|x)$。
2. (高维状态下的贝叶斯-MCMC) 在（1）中你会发现你很快就计算成功了数值，其实对于低维数据，低维积分是很快的，但是如果参数维度相对较高，计算机计算误差较大同时计算用时较久。下面我们来看一个包含多个参数的例子：`lr.csv`中有三列，`y`, `x1`, `x2`，我

们希望建模如下的方程：

$$y = k_0 + k_1 x_1 + k_2 x_2 + \epsilon$$

其中先验分布满足：$\epsilon \sim N(0, \sigma^2), \sigma \sim U(0,1)$，$k_i \sim N(0,1), i = 0,1,2$。请你利用MCMC采样估计上述参数、可视化这四个参数的后验均值，同时报告采样所花的时间(请使用print报告)，(10分)；

3. （高维状态下的贝叶斯-数值积分）请你使用**完整贝叶斯公式**完成以下计算进行计算（5分)：
   - 首先定义先验分布 $p(\theta)$ 和似然函数 $p(x|\theta)$；
   - 由于计算复杂性，你可以选择固定 $\sigma = 1$，只对三个k参数进行积分
   - 计算积分 $\int p(x|\theta)p(\theta)d\theta$ 并报告积分用时，需要用到多重积分的函数，积分上下界 $[-2, 2]$；
   - 注意：不需要完成进一步后验分布的计算和报告
4. （为什么MCMC更快）对比多元回归中的两个问题，你会发现好像MCMC的方法的确要比计算积分式求解更快，请结合下面关于算法和复杂度的介绍，说说为什么MCMC相比使用贝叶斯公式进行积分在计算速度上具有优势（5分）

> **Note**
>
> （1）计算复杂度：这里我们只关注大O表示法下的时间复杂度，大概就是表示执行算法所需要的操作次数，可以认为，执行一次操作的复杂度是 $O(1)$ ，如果用for循环重复 n 次计算复杂度为 $O(n)$，d 个for循环的复杂度是 $O(n^d)$，以此类推。
> （2）对于数值积分，如果是一维积分，常常把积分区间划分成很 N 个小份(如下图)，然后用矩形面积（对应位置的函数值乘bins的宽度，可以看作1次O(1)操作），求和后近似积分结果，即$S = \sum_{i=1}^{n} f(x_i)\Delta_x$, 如果是二维则需要划分成 $N \times N$的小份，以此类推d维则需要对d个维都进行划分。
> （3）对于mcmc，以ppt上的Metropolis 算法为例，大概的算法如下：
> - 首先定义参数的初始值 $\theta_{current}$，以及采样次数 M;
> - 其次根据提倡分布（比如$q(\theta_{proposed}|\theta_{current}) = N(\theta_{current}, \sigma^2)$）采样的到 $\theta_{proposed}$，这一步用时可以忽略；
> - 根据公式 $p_{move} = \min(\frac{p(\theta_{proposed})p(x|\theta_{proposed})}{p(\theta_{current}p(x|\theta_{current})}, 1)$，可以看作一次O(1)操作；
> - 以概率 p_{move} 将参数更新为 $\theta_{proposed}$，以概率 $1 - p_{move}$ 参数保留为 $\theta_{current}$，这一步用时可以忽略；
> - 重复2-4的操作 M 次。

---

3 Coffee or Beer

在一项食品研究中，研究人员想要区分两类专业人士的口味偏好：

- **咖啡师**：专长于咖啡的风味、烘焙程度、酸度等；

- **酿酒师**：专长于啤酒的苦度、麦芽风味、酒精发酵工艺等。
  研究人员提供了一系列饮品，其中包括咖啡变种与啤酒变种，参与者品尝每种饮品并给出评分(评分范围0-5，为连续值)。`cb.csv`中储存了不同`rater`对不同`drink`的打分。现有以下两种假设：
- **H1(职业影响口味偏好)**：咖啡师给咖啡评分(α_0)比啤酒分数(β_0)更高，而酿酒师给啤酒评分(α_1)比咖啡分数更高(β_1)。同时咖啡师给分极差更大($\beta_0 < \beta_1 < \alpha_1 < \alpha_0$)。
- **H2(跨界专家的影响)**： 除了H1中给出的咖啡师和酿酒师外，还有一群风味专家混进了评分队伍，他们是研究食品风味的科学家，对不同风味有自己研究，对啤酒和咖啡并无给分高低之分，但由于见多识广,评分(θ)不高不低($\beta_0 < \beta_1 < \theta < \alpha_1 < \alpha_0$)
  假设每次给分时，对于每个人每个饮品噪音σ相同(可参考`two country quiz`的例子进行作答)。**与此同时，研究人员认识0号评分者，知道他是一个酿酒师(hints：想想这个对应着`two country quiz`中pymc模型定义的哪部分, 如果在H2中这个信息如何利用)。**

1. 请为上述参数（包含σ）设置合理的先验分布，并绘制出H1和H2的graphical model（10分）；
2. 在此问中分别运行H1和H2中的模型，请在答案中完成如下要求（10分）
   - 对于H1：分别绘制β_0，β_1，α_0 , α_1，σ的后验分布并标注出95% HDI区间；
   - 对于H2：分别绘制β_0，β_1，α_0 , α_1，σ，θ的后验分布并标注出95% HDI区间；
   - 报告两个模型参数相应的收敛性指标，同时回答对于H2，模型认为哪个评分者是风味专家
3. 请分别从waic和loo来比较两个假设，说明哪个模型更合理。(10分)