统计学习：第十九章

1　用蒙特卡罗积分法求

$$\int_{-\infty}^{\infty} x^2 \exp(-\frac{x^2}{2}) dx$$

答：令 $f(x) = \sqrt{2\pi}x^2$,$p(x)$ 为标准正态分布。$E(\sqrt{2\pi}x^2) = \sqrt{2\pi}$

2　证明如果马尔可夫链是不可约的，且有一个状态是非周期的，则其他所有状态也是非周期的，即这个马尔可夫链是非周期的。

答:若 i 是非周期的,则 $gcd(\{t : P(x_t = i|x_0 = i) > 0\}) = 1$。考虑任意一个状态 j，由于不可约 $\exists t_{ij}, t_{ji}, \text{s.t.} P(x_{t_{ij}} = j|x_0 = i) > 0, P(x_{t_{ji}} = i|x_0 = j) > 0$,则有 $P(x_{t_{ij}+t_{ji}} = j|x_0 = j) > 0$。以及对于任意 $t' \in \{t : P(x_t = i|x_0 = i) > 0\}$，$P(x_{t_{ij}+t_{ji}+t'} = j|x_0 = j) > 0$。则 $gcd(\{t : P(x_t = j|x_0 = j) > 0\}) \leq gcd(\{t : P(x_t = i|x_0 = i) > 0\}) = 1$，故 j 非周期。

3　验证具有以下转移概率矩阵的马尔可夫链是可约的，但是非周期的。

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1 \end{bmatrix}$$

答：$\forall t \in N^*, j \neq 4, P(x_t = j|x_0 = 4) = 0,$ 故可约。$P(x_1 = 1|x_0 = 1) = 1/2,$
$P(x_4 = 1|x_0 = 4) = 1,$
$P(x_2 = 2|x_0 = 2) = 1/2$,$P(x_3 = 2|x_0 = 2) = 1/8,$
$P(x_2 = 3|x_0 = 3) = 1/4$,$P(x_5 = 3|x_0 = 3) = 1/32$
可知，转移概率矩阵非周期。

4　验证具有以下转移概率矩阵的马尔可夫链是不可约的，但是周期性的。

$$P = \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1 \\ 0 & 0 & 1/2 & 0 \end{bmatrix}$$

答：数学归纳法可得 P^{2n}, P^{2n+1} 矩阵的值交替出现零和非零数，故不可约，且周期皆为 2。

5 证明可逆马尔可夫链一定是不可约。
答：不一定。将两个可逆马尔可夫链拼在一起就不可约，例如 I_2。

6 从一般的 Metropolis-Hastings 算法推导出单分量 Metropolis-Hastings 算法。答：
输入：抽样的目标分布的密度函数 $p(x)$，函数 $f(x)$
输出：$p(x)$ 的随机样本 $x_{m+1}, \ldots, x_n$，函数样本均值 f_{mn}
参数：收敛步数 m，迭代步数 n。
1. 任意选择一个初始值为 $x_0 = (x_0^1, \ldots, x_0^k)$
2. 对 $i = 1, \ldots, n$ 开始循环
    设定 $x_i = x_{i-1}$, 对 $j = 1, \ldots, k$ 开始循环
        a. 由建议分布 $q(x_{i-1}^j, x_j | x_i^{-j})$ 抽取候选值 x'^j_i
        b. 计算接受概率

$$\alpha(x_{i-1}^j, x'^j_i) = \min\left\{1, \frac{p(x'^j_i | x_i^{-j}) q(x'^j_i, x_{i-1}^j | x_i^{-j})}{p(x_{i-1}^j | x_i^{-j}) q(x_{i-1}^j, x'^j_i | x_i^{-j})}\right\}$$

        c. 从区间 (0,1) 中按均匀分布随机抽取一个数 u
            若 $u \leq \alpha(x_{i-1}^j, x'^j_i)$,$x_i^j = x'^j_i$, 否则，保持 x_i^j 不变 3. 得到样本集合 $x_{m+1}, \ldots, x_n$, 计算 f_{mn}

7 假设进行伯努利实验，后验概率为 $P(\theta|y)$, 其中变量 $y \in \{0, 1\}$ 表示实验可能的结果，变量 θ 表示结果为 1 的概率。再假设先验概率 $P(\theta)$ 遵循 Beta 分布 $B(\alpha, \beta)$，其中 $\alpha = 1, \beta = 1$，似然函数 $P(y|\theta)$ 遵循二项分布 $\text{Bin}(n, k, \theta)$, 其中 $n = 10, k = 4$, 即实验进行 10 次其中结果为 1 的次数为 4, 试用 Metropolis-Hastings 算法求后验概率分布 $P(\theta|y) \propto P(\theta)P(y|\theta)$ 的均值和方差。
答：后验分布为 $B(5, 7)$，剩余见代码。

8 设某试验可能有五种结果，其出现的概率分别为

$$\frac{\theta}{4} + \frac{1}{8}, \frac{\theta}{4}, \frac{\eta}{4}, \frac{\eta}{4} + \frac{3}{8}, \frac{1}{2}(1 - \theta - \eta)$$

模型含有两个参数 θ 和 η 都介于 0 和 1 之间。现有 22 次试验结果的观测值为

$$y = (y_1, y_2, y_3, y_4, y_5) = (14, 1, 1, 1, 5)$$

其中 y_i 表示 22 次试验中第 i 个结果出现的次数，$i = 1, 2, \ldots, 5$ 试用吉布斯抽样估计参数 θ 和 η 的均值和方差。

答:

$$P(\theta, \eta | y) \propto (\frac{\theta}{4} + \frac{1}{8})^{14} \frac{\theta}{4} \frac{\eta}{4} (\frac{\eta}{4} + \frac{3}{8}) (\frac{1}{2}(1 - \theta - \eta))^5$$

剩余见代码