# 统计中的计算方法 · 课后作业（5）

梁子龙（15300180026）

2018 年 5 月 24 日

**作业 1** 假设随机变量 $X$ 和 $Y$ 都在 $(0, B)$ 上取值. 假设 $f(x|y) = C(y)\mathrm{e}^{-xy}$, $0 < x < B$; $f(y|x) = C(x)\mathrm{e}^{-xy}$, $0 < y < B$. 给出一种方法来模拟 $X, Y$, 并用模拟的方法来估计 $\mathrm{E}(X)$ 和 $\mathrm{E}(XY)$.

**答.** 模拟高维分布的方式主要有 *Gibbs* 采样法及 *Metropolis-Hastings* 算法，均为 MCMC 的特定形式. 本题中 $X$ 与 $Y$ 的条件概率函数均已给定，很容易地想到用 Gibbs 采样来模拟这一分布. 首先，我们随机生成第一个样本 $(X^{(0)}, Y^{(0)})$，而后对第 $i$ 步，进行如下模拟：

$$\text{选取 } X \sim f(x|Y^{(i-1)}),$$
$$\text{选取 } Y \sim f(y|X^{(i)}).$$

而后选取后一部分样本，即可得到满足所需联合分布的样本集.

数值试验中，我们选取 $B = 10$，并用上述 Gibbs 采样法得到 3000 个样本，实验结果如图 1 所示. 根据题中所述，条件分布均为指数分布，可以看出，模拟出的样本显示出了明显的指数特性. 根据样本对期望进行模拟，得到

$$\mathrm{E}(X) \approx \frac{1}{n}\sum_{i=1}^{n} X_i = 1.879,$$
$$\mathrm{E}(XY) \approx \frac{1}{n}\sum_{i=1}^{n} X_iY_i = 0.817.$$

**作业 2** 假设 $X_i, i = 1, 2, 3$ 相互独立且服从均值为 1 的指数分布. 设计一种模拟方法来估计:

1. $E(X_1 + 2X_2 + 3X_3|x_1 + 2X_2 + 3X_3 > 15)$;
2. $E(X_1 + 2X_2 + 3X_3|x_1 + 2X_2 + 3X_3 < 1)$.

**答.** 由于 $X_i$ 间相互独立，那么直接按照一维分布的生成方式分别生成三组样本即可. 数值试验中，

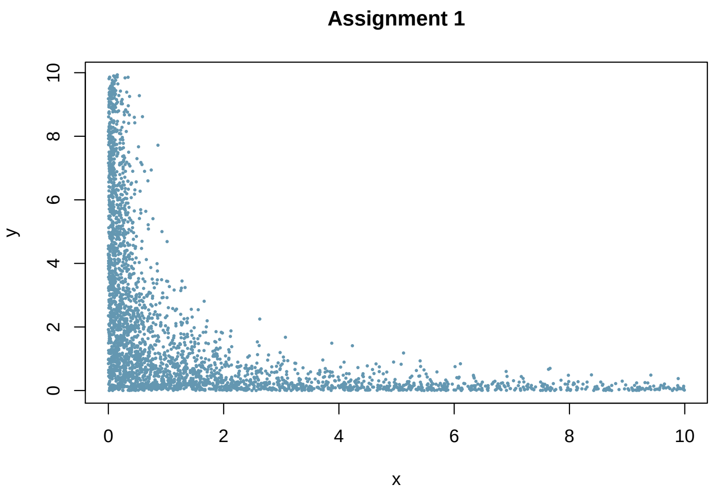


图 1: 作业 1 实验结果图

我们直接根据指数分布得到 3000 个样本，计算有

$$\mathrm{E}(\theta|\theta > 15) \approx \frac{1}{k_1}\sum_{i_1=1}^{k_1}\theta_{i_1} = 18.19,$$

$$\mathrm{E}(\theta|\theta < 1) \approx \frac{1}{k_2}\sum_{i_2=1}^{k_2}\theta_{i_2} = 0.71.$$

其中，我们记 $\theta = X_1 + 2X_2 + 3X_3$，而 $k_1$, $k_2$ 分别为样本中满足 $\theta > 15$ 和 $\theta < 1$ 的数目. □

**作业 3** 假设 $X, Y, Z$ 的联合概率密度为

$$f(x, y, z) = C\mathrm{e}^{-(x+y+z+axy+bxz+cyz)}, \quad x, y, z > 0,$$

其中 $a, b, c$ 为非负常数，$C$ 的取值与 $a, b, c$ 无关. 如何估计 $X, Y, Z$? 给定 $a = b = c = 1$，如何估计 $E(XYZ)$? 给出方法并实现.

**答.** Metropolis-Hastings 算法可以用来模拟这一分布. 特别地，我们这里选取使用随机游走来更新样本的 Metropolis 算法: 首先取定初始值 $\theta^{(0)} = (X^{(0)}, Y^{(0)}, Z^{(0)})$, 而后对第 $i$ 步, 取 $\theta^* = \theta^{(i-1)} + Z$, 其中 $Z$ 服从标准正态分布. 接下来计算如下概率:

$$\alpha(\theta^{(i-1)}, \theta^*) = \min\left(1, \frac{f(\theta^*)}{f(\theta^{(i-1)})}\right),$$

并以概率 $\alpha$ 取 $\theta^{(i)} = \theta^*$，否则 $\theta^{(i)} = \theta^{(i-1)}$ 即可.

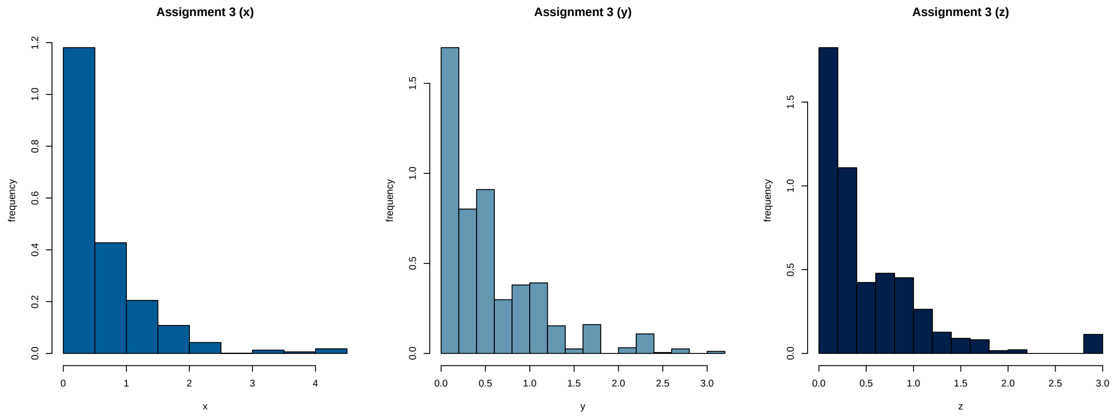


图 2: 作业 3 实验结果图：三幅图从左至右分别代表 $X, Y, Z$ 三个变量的频率分布. 可以看出，它们均呈现出类似与指数分布的特征.

应用数值试验得到 3000 个样本之后，我们可以得到 $\mathrm{E}(XYZ)$ 的估计式：

$$\mathrm{E}(XYZ) \approx \frac{1}{n}\sum_{i=1}^{n} X_i Y_i Z_i = 0.091.$$

图 2 展示了这些样本中 $X, Y, Z$ 三个变量分别的频率分布. 根据概率密度函数所具有的指数特性可以看出，这些频率分布均呈现出与指数分布类似的特征.

**作业 4** 假设 $X, Y, N$ 的联合分布为

$$\mathrm{P}(X = i, y \leq Y \leq y + \mathrm{d}y, N = n) \approx C\binom{n}{i} y^{i+\alpha-1}(1-y)^{n-i+\beta-1}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\,\mathrm{d}y,$$

其中, $i = 0, \ldots, n$, $n > 0$, $y >= 0$. 当 $\alpha = 2$, $\beta = 3$, $\lambda = 4$ 时, 用模拟方法估计 $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{E}(N)$.

**答.** 题中所述联合分布对 $X$ 和 $N$ 来说是离散分布，对于 $Y$ 是连续分布. 这里同样可以使用 Metropolis-Hastings 算法. 首先取定初始值 $\theta^{(0)} = (X^{(0)}, Y^{(0)}, Z^{(0)})$，而后对第 $i$ 步，取

$$\begin{aligned} X^* &= X^{(i-1)} + R_X, \quad R_X \sim \{-1, 1\} \text{ 上的均匀分布;} \\ Y^* &= Y^{(i-1)} + R_Y, \quad R_Y \sim U(-1/5, 1/5); \\ Z^* &= Z^{(i-1)} + R_Z, \quad R_Z \sim \{-1, 1\} \text{ 上的均匀分布.} \end{aligned}$$

那么，这里的建议分布（*proposal distribution*）为常数

$$q(X, Y, Z, X', Y', Z') = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{5}{2}.$$

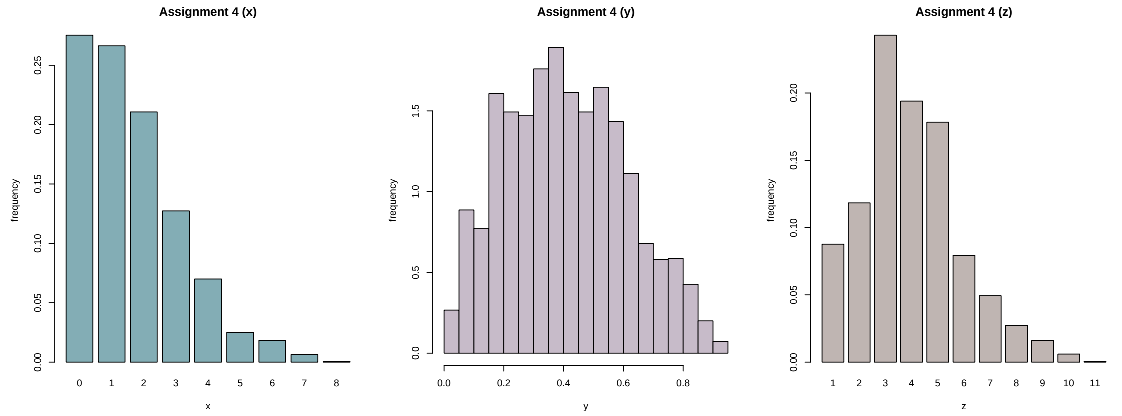


图 3: 作业 4 实验结果图：三幅图从左至右分别代表 $X, Y, Z$ 三个变量的频率分布．在原联合分布中，$X, Y, Z$ 的分布分别具有二项分布、Beta 分布和 Poisson 分布的特点，可以看出，对该联合分布的模拟体现出了这些特性．

得到建议分布后，便可以计算如下概率：

$$\alpha(\theta^{(i-1)}, \theta^*) = \min\left(1, \frac{\pi(\theta^*)q(\theta^*, \theta^{(i-1)})}{\pi(\theta^{(i-1)})q(\theta^{(i-1)}, \theta^*)}\right) = \min\left(1, \frac{\pi(\theta^*)}{\pi(\theta^{(i-1)})}\right).$$

这里，$\theta = (X, Y, Z)$，且 $\pi(\cdot)$ 为题中所述分布的概率函数．我们在这一步以概率 $\alpha$ 取 $\theta^{(i)} = \theta^*$，否则 $\theta^{(i)} = \theta^{(i-1)}$ 即可．

数值试验中，我们取 $\alpha = 2$, $\beta = 3$, $\lambda = 4$，根据上述方法取样 3000 个点，得到期望的估计值：

$$\mathrm{E}(X) \approx 1.361, \quad \mathrm{E}(Y) \approx 0.389, \quad \mathrm{E}(Z) \approx 3.804.$$

图 3 展示了这些样本中 $X, Y, Z$ 三个变量分别的频率分布．在原联合分布中，$X, Y, Z$ 的分布分别具有二项分布、Beta 分布和 Poisson 分布的特点，可以看出，对该联合分布的模拟体现出了这些特性．

**作业 5** 生成两个二维正态分布生成的混合正态分布．两个二维正态分布的均值和协方差为

$$\mu_1 = (1, 4), \quad \mu_2 = (-2, -1), \quad \varSigma_1 = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 2 \end{pmatrix}, \quad \varSigma_2 = \begin{pmatrix} 6 & 0.9 \\ 0.9 & 1 \end{pmatrix},$$

且两个分布中的随机变量产生的概率分别为 0.5，0.5.

**答.** 不难根据高维正态分布的概率密度函数，写出题中所述混合分布的概率密度函数为：

$$f(\theta) = \frac{1}{4\pi\sqrt{|\varSigma_1|}} \exp\left(-\frac{1}{2}(\theta - \mu_1)^{\mathrm{T}} \varSigma_1^{-1} (\theta - \mu_1)\right) + \frac{1}{4\pi\sqrt{|\varSigma_2|}} \exp\left(-\frac{1}{2}(\theta - \mu_2)^{\mathrm{T}} \varSigma_2^{-1} (\theta - \mu_2)\right).$$

那么可以完全根据作业 3 中所述的 Metropolis-Hastings 方法来模拟这一分布．进行数值试验后，

得到如图 4 的样本集合．可以看出，模拟出的样本较好地符合了混合正态分布的概率密度．

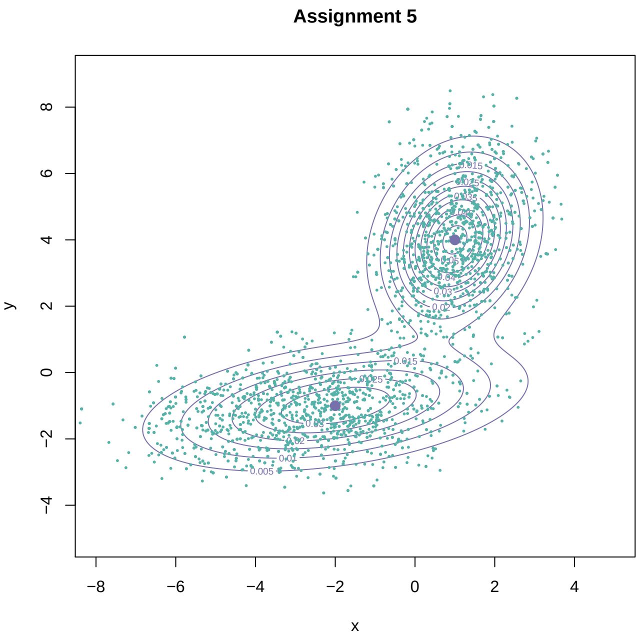


图 4: 作业 5 实验结果图