统计中的计算方法 · 课后作业（4）

梁子龙（15300180026）

2018 年 5 月 23 日

作业 1 用两种方法产生以下分布，并进行数值试验：

$$\mathrm{P}(i) = \frac{\mathrm{e}^{-\lambda}\lambda^i/i!}{\sum_{j=0}^{k} \mathrm{e}^{-\lambda}\lambda^j/j!}, \quad i = 0, \dots, k.$$

答. 生成服从特定分布的随机数可以使用随机数法和接受-拒绝法. 根据反函数法，我们首先需要生成服从 $[0,1]$ 上均匀分布的随机变量 Y，而后可以根据如下公式生成离散分布的随机数：

$$X = i, \quad 若 \sum_{j=0}^{i-1} \mathrm{P}(j) \leq Y < \sum_{j=0}^{i} \mathrm{P}(j).$$

接下来叙述利用接受-拒绝方法生成离散分布随机数的流程. 首先，我们可以得到根据概率质量函数 Q 生成的随机数 Y，其定义域需与 P 相同；而后生成一个服从 $[0,1]$ 上均匀分布的随机数 U，那么可以取

$$X = Y, \quad 若\ U < \frac{P(Y)}{cQ(Y)},$$

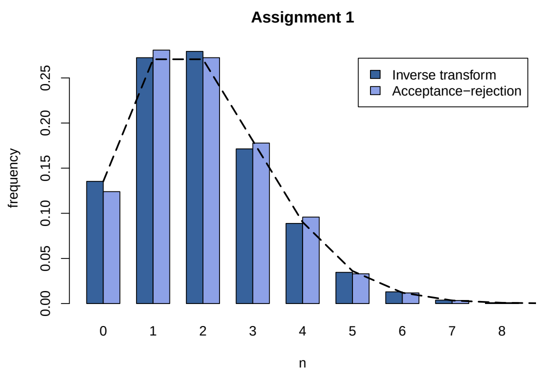
否则拒绝这一次选择，继续生成. 这里 $c = \max_j P(j)/Q(j)$.

我们选取 $k = 15$，$\lambda = 2$ 作为原分布的参数，生成 5000 个点. 经过对上述两种方法进行程序实现和数值试验后，得到的结果如图 1a 所示，图中黑色曲线为原概率质量函数 P，柱状图为两种模拟方式所生成数据的频率. 可以看出，两种方式均良好地模拟出了原分布. □

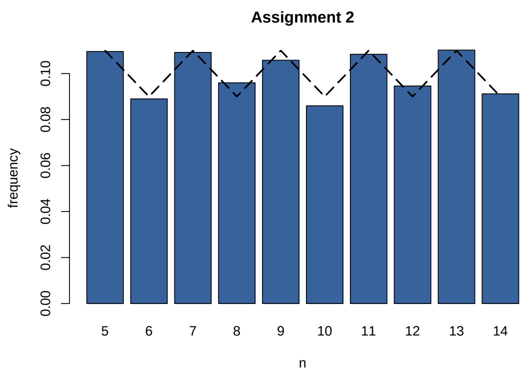
作业 2 给出模拟下述分布的方法，并进行数值试验：

$$p_j = \begin{cases} 0.11 & j\ 为奇数且\ 5 \leqslant j \leqslant 13; \\ 0.09 & j\ 为偶数且\ 6 \leqslant j \leqslant 14. \end{cases}$$

答. 同样根据上题的反函数法便可以实现上述分布的模拟. 生成 5000 个点后，实验结果如图 1b 所示. 可以看出，反函数法对该分布进行了良好的模拟. □



(a) 作业 1 实验结果图．图中黑色曲线为原概率质量函数 P，柱状图为两种模拟方式所生成数据的频率．



(b) 作业 2 实验结果图．图中黑色曲线为原概率质量函数 p，柱状图为利用反函数法所生成数据的频率．

图 1: 作业 1 与作业 2 的实验结果图．

作业 3 给出具有如下概率密度函数的随机变量的产生方法，并进行数值试验：

$$f(x)=\begin{cases}\mathrm{e}^{2x} & -\infty<x<0,\\ \mathrm{e}^{-2x} & 0<x<\infty.\end{cases}$$

答. 这里可以利用接受-拒绝方法来模拟这个分布．对于这样一个连续的分布，我们可以使用标准正态分布生成随机数 Y，而后生成一个服从 $[0,1]$ 上均匀分布的随机数 U，那么可以取

$$X=Y,\quad \text{if } U<\frac{P(Y)}{cQ(Y)},$$

否则拒绝这一次选择，继续生成．这里 $c=\max_x P(x)/Q(x)=\sqrt{2\pi}$．

生成 5000 个点后，实验结果如图 2a 所示．可以看出，接受-拒绝法对该分布进行了良好的模拟． □

作业 4 给出具有如下概率密度函数的随机变量的产生方法，进行数值试验，并讨论方法的运算效率：

$$f(x)=30(x^2-2x^3+x^4),\quad 0\leqslant x\leqslant 1.$$

答. 这里我们利用接受-拒绝方法来模拟这个分布．我们取进行判断的分布 P 为 $[0,1]$ 上的均匀分布，取 $c=f(1/2)=15/8$ 即可．生成 5000 个点后，实验结果如图 2b 所示．可以看出，接受-拒绝法对该分布进行了良好的模拟． □

作业 5 给出产生具有如下概率密度函数的随机变量的接受-拒绝方法：

$$f(x)=\frac{1}{2}x^2\mathrm{e}^{-x},\quad (0\leqslant x<\infty).$$

假设用指数分布来产生此分布，给出最优的参数 λ．

答. 令 $g(x)=\lambda\mathrm{e}^{-\lambda x}$ 为指数分布的概率密度函数，那么

$$\frac{f(x)}{g(x)}=\frac{1}{2\lambda}x^2\mathrm{e}^{-(1-\lambda)x}.$$

要求出 c，必须求出 $f(x)/g(x)$ 的最大值，为此有

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{g(x)}\right)=\frac{1}{2\lambda}\left(2x\mathrm{e}^{-(1-\lambda)x}-(1-\lambda)x^2e^{-(1-\lambda)x}\right),$$

这样就得到，当 $x=\frac{2}{1-\lambda}$ 时，$f(x)/g(x)$ 达到最大值 $c=\frac{2\mathrm{e}^{-2}}{\lambda(1-\lambda)^2}$，这里要求 $\lambda<1$.

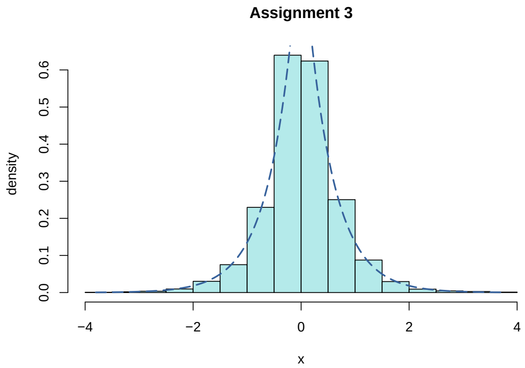
接下来我们要求知最优的参数 λ 是多少．注意到接受-拒绝方法得到一个随机数所需要的平均步数约为 c，因此最优的 λ 将使 c 达到最小．为此，计算

$$\frac{\mathrm{d}c}{\mathrm{d}\lambda}=2\mathrm{e}^{-2}\left(-\frac{1}{\lambda^2(1-\lambda)^2}+\frac{2}{\lambda(1-\lambda)^3}\right),$$

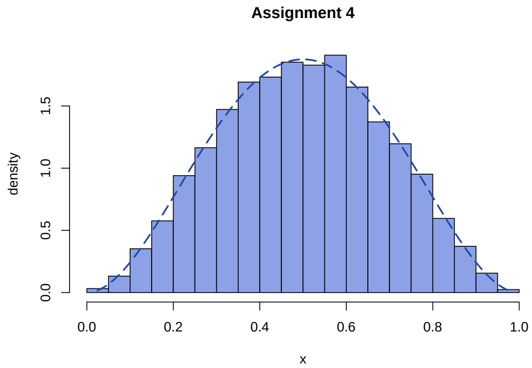
令 $c'(\lambda)=0$，我们得到

$$\lambda=\frac{1}{3},\quad c=\frac{27}{2\mathrm{e}^2}.$$

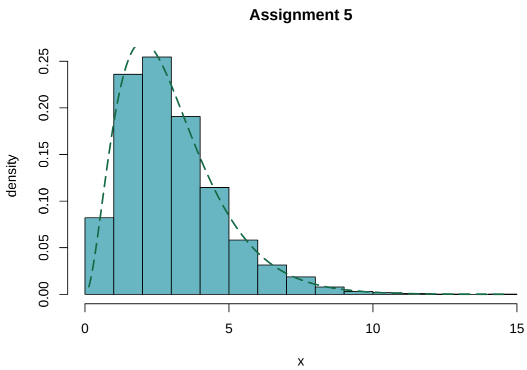
根据上述方法生成 5000 个样本点后，实验结果如图 2c 所示．可以看出，接受-拒绝法对该分布进行了良好的模拟． $\square$



(a) 作业 3 实验结果图. 图中深蓝色曲线为原概率密度函数，直方图为所生成数据的频率.



(b) 作业 4 实验结果图. 图中深蓝色曲线为原概率密度函数，直方图为所生成数据的频率.



(c) 作业 5 实验结果图. 图中深绿色曲线为原概率密度函数，直方图为所生成数据的频率.

图 2: 作业 3、4 和 5 的实验结果图.