统计中的计算方法
第四次作业

于慧倩
14300180118

2017 年 6 月

1 第一题

假设随机变量 X 和 Y 都在 $(0,B)$ 上取值。假设 $f(x|y)=C(y)e^{-xy}, 0<x<Bf(y|x)=C(x)e^{-xy}, 0<y<B$，给出一种方法来近似模拟 X,Y, 并用模拟的方法来估计 $E(X), E(XY)$

1.1 模拟方法

首先通过 $\int_0^B f(x|y)=C(y)e^{-xy}=1$ 计算出,$C(y)=\frac{y}{1-e^{-By}}$.
同理：$C(x)=\frac{x}{1-e^{-Bx}}$.
选择用 Gibbs 采样进行模拟:

- Step 1: 随机给出初始值,$(x_0,y_0)\in(0,B)\times(0,B)$
- Step 2: 重复取样 $x_i\sim f(x|y_{i-1})$,$y_i\sim f(y|x_i)$;

其中生成 $x_i\sim f(x|y_{i-1})$，采用 inverse 方法进行:

- Step 1: 生成 $(0,1)$ 均匀分布的随机数 U
- Step 2:$x_i=-\frac{1}{y}\ln(1-\frac{Uy}{C(y)})$;

生成 y_i 的步骤同理。

1.2 模拟估计

调用 MCMC.Gibbsxy.R 中的函数,生成 10^4 个随机变量且 B 时,MCMC.Gibbsxy(10^4,1): 得到 $E(X)$ 如下表:

E(X)	**E(XY)**
0.4596391	0.2047234
0.4610378	0.2064006
0.4615507	0.2033755
0.4651505	0.2075964
0.468295	0.208894

2 第二题

假设 $Xi, i = 1, 2, 3$，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,$E(X1 + 2X2 + 3X3|X1 + 2X2 + 3X3 > 15), E(X1 + 2X2 + 3X3|X1 + 2X2 + 3X3 < 1)$

分别用 Gibbs 采样与 Metropolis-Hastings 算法生成两种分布：

2.1 $E(X1 + 2X2 + 3X3|X1 + 2X2 + 3X3 > 15)$

运用 Gibbs 采样方法：

- Step 1: 选取初值 (X_1, X_2, X_3)，使得 $X1 + 2X2 + 3X3 > 15)$
- Step 2: 重复操作，从三个变量中选出两个，I, J, 固定另一个不变，对 $I\ J$ 进行变化；

其中假设固定 X_3 不变，意识到有 x_i 相互独立且服从均值为 1 的指数分布且 $X_1 + 2X_2 > 15 - 3X_3$。计算条件概率，

$$f(X_1|X_1{+}2X_2 > 15{-}3X_3 = a) = Ce^{-1\times X_1}\int_{(a-X_1)/2}^{\infty} e^{-1\times X_2}\mathrm{d}X_2 = C'e^{-1\times X_1+1/2X_1}$$

生成参数为 $1 - 1/2$ 的指数分布的随机数，令 X_1 等于这个数，然后再用 inverse 方法计算 X_2。

$$f(X_2|X_2 > 15 - 3X_3 - X_1 = b) = e^{b-X}I_{X\geqslant b}$$

$$F(X_2|X_2 > 15 - 3X_3 - X_1 = b) = (1 - e^{b-X})I_{X\geqslant b} = U$$

运用 inverse 方法生成 X_2：

- Step 1: 生成 (0,1) 上均匀分布的随机数 U
- Step 2: 令 $X_2 = b - \ln(1 - U) = b - \ln(U)$

2.2 模拟估计

见 $H4\_2\_1.R$

选取初值为 (5,5,5)，调用 $H4\_2\_1(10^5)$，生成 10^5 个随机数，去除前 1000 个，运行得到 $E(X_1 + 2X_2 + 3X_3|X_1 + 2X_2 + 3X_3 > 15)$ 如下表：

$E(X_1+2X_2+3X_3\|X_1+2X_2+3X_3>15)$
17.66761
17.65996
17.67592
17.68168
17.66473

2.3 $E(X_1+2X_2+3X_3|X_1+2X_2+3X_3<1)$

用 Metropolis-Hastings 算法生成,Proposal Kernel 采用 Random Walk：

$$\theta' = \theta + Z, Z \sim N(0,\sigma^2)$$

概率分布

$$f(X_1{+}2X_2{+}3X_3|X_1{+}2X_2{+}3X_3<1) = Ce^{-(X_1+X_2+X_3)} \times I_{(X_1+2X2+3X3<1)} \times I_{X_1>0} \times I_{X_2>0} \times I_{X_2>0}$$

单次循环中接受概率：

$$\alpha(X^{(i-1),X*}) = exp(\sum(X^{(i-1)}) - \sum(X*)) \times I_{(X*_1+2X2+3X3<1)} \times I_{X_1>0} \times I_{X_2>0} \times I_{X_2>0}$$

具体算法如下：

- Step 1: 选取适当的初值，保证 $X_1+2X_2+3X_3<1$
- Step 2: 生成 $Z \sim N(0,\sigma^2)$，令 $X^*=X+Z$
- Step 3: 计算单次接受概率 $\alpha(X^{(i-1),X^*})$
- Step 4: 生成 (0,1) 上均匀分布的随机变量 U，若 $U<\alpha(X^{(i-1),X^*})$，令 $X^{(i)}=X^*$，否则保持 $X^{(i)}=X^{(i-1)}$

2.4 模拟估计

见 $H4\_2\_2.R$

选取初始值为 $(0.1,0.1,0.1)$,为保证接受率选取 $\sigma=0.09$,调用 $H4\_2\_2(10^4)$,生成 10^4 个随机数,去除前 1000 个,得到 $E(X_1{+}2X_2{+}3X_3|X_1{+}2X_2{+}3X_3<1)$ 如下表

$E(X_1+2X_2+3X_3\|X_1+2X_2+3X_3<1)$
0.7340738
0.7261952
0.727004
0.7210271
0.7357822

3 第三题

假设 X,Y,Z 的联合概率密度为 $f(x,y,x)=Ce^{-(x+y+z+axy+bxz+cyz)},x>0,y>0,z>0$，其中 a,b,c 为非负常数，C 取值与 a,b,c 无关，估计 X，Y，Z。并估计 E(XYZ)，当 $a=b=c=1$ 时。

3.1 模拟方法

采用 Metropolis-Hastings 算法估计 X，Y，Z：

Proposal Kernel 采用 Random Walk：

$$\theta'=\theta+Z,Z\sim N(0,\sigma^2)$$

概率分布

$$f(x,y,x)=Ce^{-(x+y+z+axy+bxz+cyz)},x>0,y>0,z>0$$

单次循环的接受概率:

$$\begin{aligned}&\alpha(X^{(i-1)},X*)\\=\ &exp((X+Y+Z+aXY+bXZ+cYZ)-(X^*+Y^*+Z^*+aX^*Y^*+bX^*Z^*+cY^*Z^*))\\&\times I_{X*>0}\times I_{Y*>0}\times I_{Z*>0}\end{aligned}$$

具体算法如下：

- Step 1: 选取适当的初值，保证 $X_1>0,X_2>0,X_3>0$
- Step 2: 生成 $Z\sim N(0,\sigma^2)$，令 $X^*=X+Z$
- Step 3: 计算单次接受概率 $\alpha(X^{(i-1),X^*})$
- Step 4: 生成 (0,1) 上均匀分布的随机变量 U，若 $U<\alpha(X^{(i-1),X^*})$，令 $X^{(i)}=X^*$，否则保持 $X^{(i)}=X^{(i-1)}$

3.2 模拟估计

见 MCMC.Metroxyz.R

调用 MCMC.Metroxyz.R(1,1,1), 选取初始值为 $(1,1,1)$，为保证接受率选取 $\sigma = 0.3$，生成 10^4 个随机数，去除前 1000 个，得到 $E(X_1X_2X_3)$ 如下表所示

$E(X_1X_2X_3)$
0.0840624
0.08302989
0.09016883
0.09280316
0.1034397

4 第四题

4.1 模拟方法

假设 X,Y,N 的联合分布为

$$\begin{aligned} & P(X = i, y \leqslant Y \leqslant y + dy, N = n) \\ \approx\ & C\binom{n}{i} y^{i+\alpha-1}(1-y)^{n-i+\beta-1} e^{-\lambda}\frac{\lambda^n}{n!}dy \end{aligned}$$

其中，i= 0, ... ,n;n = 0,1, ... ;y 0. 当 α=2，β=3，λ=4 时，用模拟方法估计 $E(X), E(Y), E(N)$

首先有

$$F(y+dy) - F(y) = \int_y^{y+dy} f(x,t,n)dt = C\binom{n}{i} y^{i+1}(1-y)^{n-i+2} e^{-\lambda}\frac{\lambda^n}{n!}dy$$

所以有

$$\frac{F(y+dy) - F(y)}{dy} = C\binom{n}{i} y^{i+1}(1-y)^{n-i+2} e^{-\lambda}\frac{\lambda^n}{n!}$$

令 $dy \to 0$ 得到

$$P(X = i, Y = y, N = n)C\binom{n}{i} y^{i+1}(1-y)^{n-i+2} e^{-\lambda}\frac{\lambda^n}{n!}$$

再求 X, Y, N 分别的条件概率

$$P(X=i|Y=y,N=n)=\frac{\binom{n}{i}y^i(1-y)^{-i}}{\sum_{i=0}^{i=n}\binom{n}{i}y^i(1-y)^{-i}}$$

$$P(Y=y|X=i,N=n)=\frac{y^{i+1}(1-y)^{n-i+2}}{\Gamma(i+2)\Gamma(n-i+2)/\Gamma(n+5)}$$

$\lambda=4(1-y)$:

$$P(N=n|X=i,Y=y)=\frac{e^{-4(1-y)}[4(1-y)]^{n-i}}{(n-i)!}$$

可以看出 P(X=i|Y=y,N=n) 是二项分布，P(Y=y|X=i,N=n) 是 Beta 分布，P(N=n|X=i,Y=y) 看作是 $(n-i)$ 的泊松分布,

选择用 Gibbs 采样进行模拟:

- Step 1: 随机给出初始值,(x_0, y_0, n_0) 并且满足条件
- Step 2: 选出随机数∈(1,2,3),对应进行采样。取样方式为:$x_i \sim f(x|y_{i-1}, n_{i-1})$,$y_i \sim f(y|x_i, n_{i-1})$,$n_i \sim f(n|x_i, y_i)$，并且保证 $x=0,\ldots,n; 0\leqslant y\leqslant 1; n=0,\ldots,;$

4.2 模拟估计

调用函数 $H4\_4(10000, 2, 3, 4)$ 五次得到的均值如下:

E(X)	**E(Y)**	**E(N)**
1.5511	0.3952789	3.9191
1.5537	0.3960138	3.9497
1.6074	0.4058824	3.9492
1.5895	0.3958827	4.0418
1.5077	0.385433	3.9543

5 第五题

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矩阵为 $(1,4),(-2,-1)$;, 两个分布中随机变量产生的概率分别为 0.5,0.5。

5.1 模拟方法

采用 Metropolis-Hastings 算法得到两个单独的二维正态分布链，然后再进行随机混合，使得两个分布中随机变量产生的概率分别为 0.5,0.5。

概率分布

$$f(x_j, y_j) = \frac{1}{2\pi\sqrt{(\Sigma_j)}} exp(-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j))$$

单次循环的接受概率:

$$\alpha_j(X^{(i-1)}, X*) = exp(\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j) - \frac{1}{2}(x*-\mu_j)^T \Sigma_j^{-1}(x*-\mu_j))$$

具体算法如下:

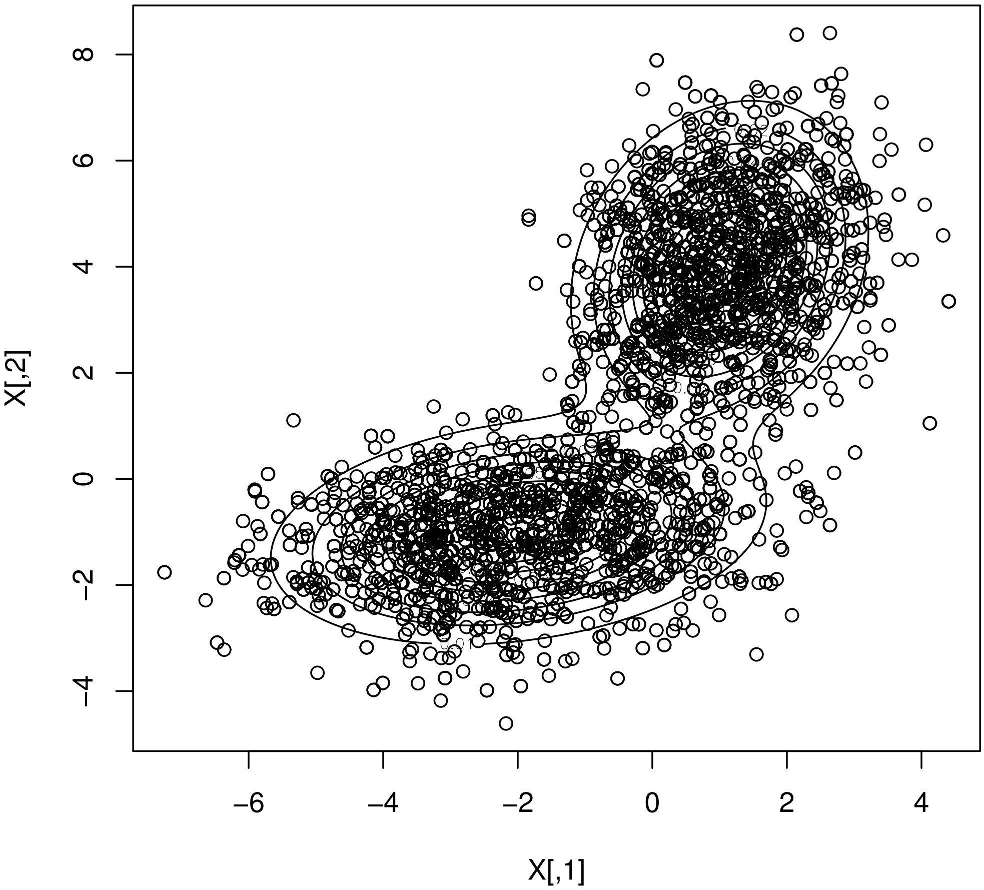
- Step 1: 选取适当的初值，保证 $X_j > 0, Y_j > 0$
- Step 2: 生成 $Z \sim N(0, \sigma_j^2)$，令 $X_j^* = X_j + Z$
- Step 3: 计算单次接受概率 $\alpha_j(X^{(i-1), X^*})$
- Step 4: 生成 (0,1) 上均匀分布的随机变量 U，若 $U < \alpha_j(X^{(i-1), X^*})$，令 $X_j^{(i)} = X_j^*$，否则保持 $X_j^{(i)} = X_j^{(i-1)}$

如此分别得到两个二维正态分布的随机数，在随机生成 $(0,1)$ 上均匀分布的随机数 U, 若 $U < 1/2$，令 $X = X_1$，否则 $X = X_2$。这样得到两个二维正态分布的混合分布。

5.2 模拟估计

见 $H4\_5.R$

这里选取初值为二维正态分布的均值，共计生成 5000 个随机数，去掉前 1000 个，得到数据如图:


X[,2]
X[,1]
8
6
4
2
0
−2
−4
−6
−4
−2
0
2
4