统计计算与软件包第二次作业

姓名：王韦力　　学号：2019302010009

2022 年 3 月 27 日

5.3

分析过程 一般地，对函数 $g(x)$，要估计 $\theta=\int_a^b g(x)dx$ 的值，我们可以选择服从 $U(a,b)$ 的随机样本 X_1，$\cdots$，X_m. 由强大数定律，$\overline{g_m(X)}=\frac{1}{m}\sum_{i=1}^m g(X_i)$ 依概率 1 收敛于 $E\left[g(X)\right]=\int_a^b g(x)\frac{1}{b-a}dx=\frac{\theta}{b-a}$. 所以此时 θ 的蒙特卡罗估计量为

$$\hat{\theta}=(b-a)\overline{g_m(X)}.$$

本题中 $a=0$，$b=0.5$，函数 $g(x)=e^{-x}$. 所以

$$\hat{\theta}=\frac{1}{2m}\sum_{i=1}^m e^{-X_i}.$$

从指数分布抽样即采用 hit-or-miss 方法，选取 $f(x)=e^{-x}(x\geq 0)$ 为随机变量 X 的密度函数，$F(x)=\int_0^x e^{-t}dt$，本题中取 $x=0.5$. 生成服从 X 的分布的随机样本 X_1，$\cdots$，X_n，则 $\theta^*=\widehat{F(x)}=\frac{1}{n}\sum_{i=1}^n I(X_i\leq x)$.

下面比较 $\hat{\theta}$ 和 θ^* 的方差. 事实上，

$$\mathrm{Var}(\hat{\theta})=\mathrm{Var}\left((b-a)\overline{g_m(X)}\right)=\frac{(b-a)^2}{m}\mathrm{Var}\left(g(X)\right). \tag{1}$$

本题中 $a=0$，$b=0.5$，$g(x)=e^{-x}(x\geq 0)$ 且 $X\sim U(0,0.5)$，所以

$$\begin{aligned}\mathrm{Var}\left(g(X)\right)=&\mathrm{Var}\left(e^{-X}\right)\\=&E\left(e^{-2X}\right)-\left[E(e^{-X})\right]^2\\=&\int_0^{0.5}e^{-2x}\cdot 2dx-\left(\int_0^{0.5}e^{-x}\cdot 2dx\right)^2\\=&(1-e^{-1})-4\left(1-e^{-\frac{1}{2}}\right)^2\\=&0.01284807.\end{aligned} \tag{2}$$

把 (2) 式的结果代回 (1) 式，得

$$m\mathrm{Var}(\hat{\theta})=0.003212018.$$

依课本 P113 中 hit-or-miss 方式的内容，$\mathrm{Var}\left(\widehat{F(x)}\right)=F(x)\left[1-F(x)\right]/n$，本题中 $x=0.5$，所以

$$n\mathrm{Var}\left(\theta^*\right)=n\mathrm{Var}\left(\widehat{F(x)}\right)=e^{-\frac{1}{2}}-e^{-1}=0.2386512.$$

所以当 $m=n$ 时，$\mathrm{Var}(\hat{\theta})$ 远小于 $\mathrm{Var}\left(\theta^*\right)$. 我们将在 Rcode 部分验证这个结论的正确性.

Rcode

```
# Uniform(0,0.5)
m <- 10000
x <- runif(m,0,0.5)
T1 <- exp(-x)
theta.hat <- 0.5*mean(T1)
var.theta.hat.est <- 0.5^2*var(T1)/m
# exp
n <- 10000
y <- rexp(n,1)
theta.star <- length(which(y>=0 & y<0.5))/n
var.theta.star.est <- theta.star*(1-theta.star)/n
c(theta.hat,theta.star)
c(var.theta.hat.est,var.theta.star.est)
```

运行结果 Rcode 中最后两行的运行结果如下：

```
> c(theta.hat,theta.star)
[1] 0.392676 0.394600
> c(var.theta.hat.est,var.theta.star.est)
[1] 3.197965e-07 2.388908e-05
```

结论 由运行结果可见，$\hat{\theta}$ 和 θ^* 都很接近 θ 的真实值 0.393469，且当 $m = n = 10000$ 时，$\mathrm{Var}(\hat{\theta}) = 3.197965 \times 10^{-7} \ll 2.388908 \times 10^{-5} = \mathrm{Var}\left(\theta^*\right)$，这与我们在分析过程部分给出的结论吻合.

5.6, 5.7

分析过程 因为 $U \sim \mathrm{Uniform}(0,1)$，所以 U 和 $1-U$ 同分布，有

$$\begin{aligned}\mathrm{Cov}(e^U, e^{1-U}) =& E(e) - E(e^U)E(e^{1-U}) \\ =& e - \left(\int_0^1 e^u \cdot 1 du\right)^2 \\ =& e - (e-1)^2 \\ =& -0.2342106.\end{aligned}$$

因为 $\mathrm{Var}(e^U) = E(e^{2U}) - \left(E(e^U)\right)^2 = \frac{1}{2}(e^2-1) - (e-1)^2$，所以

$$\begin{aligned}\mathrm{Var}(e^U + e^{1-U}) =& \mathrm{Var}(e^U) + \mathrm{Var}(e^{1-U}) + 2\mathrm{Cov}(e^U, e^{1-U}) \\ =& 2\left[\frac{1}{2}(e^2-1) - (e-1)^2\right] + 2\left[e - (e-1)^2\right] \\ =& -3e^2 + 10e - 5 \\ =& 0.01564999.\end{aligned}$$

记简单蒙特卡罗方法和对偶变量法的估计量分别为 $\hat{\theta}$ 和 θ^*，则 $\mathrm{Var}(\hat{\theta}) = \dfrac{\mathrm{Var}(e^U)}{m}$. 又

$$
\begin{aligned}
\mathrm{Var}(\theta^*) =&\mathrm{Var}\left[\frac{1}{m}\left(\sum_{i=1}^{m/2} e^{U_i} + \sum_{i=1}^{m/2} e^{1-U_i}\right)\right] \\
=&\frac{1}{m^2}\left[\frac{m}{2}\mathrm{Var}(e^U) + \frac{m}{2}\mathrm{Var}(e^{1-U}) + m\mathrm{Cov}(e^U, e^{1-U})\right] \\
=&\frac{1}{2m}\left[\mathrm{Var}(e^U) + \mathrm{Var}(e^{1-U}) + 2\mathrm{Cov}(e^U, e^{1-U})\right] \\
=&\frac{1}{2m}\mathrm{Var}(e^U + e^{1-U}).
\end{aligned}
$$

所以使用对偶变量法方差缩减百分比为 $\dfrac{\left[\mathrm{Var}(\hat{\theta}) - \mathrm{Var}(\theta^*)\right]}{\mathrm{Var}(\hat{\theta})} \approx 96.7670\%$. 我们在 RCode 部分验证这个结论.

Rcode

```
# Simple M-C, Uniform(0,1)
m <- 10000
x <- runif(m)
T1 <- exp(x)
theta.hat <- mean(T1)
var.theta.hat.est <- var(T1)/m
# Antithetic Variables
n <- 10000
R <- 1000
theta.star <- numeric(R)
for (i in 1:R) {
u <- runif(n/2)
v <- 1-u
u <- c(u,v)
T2 <- exp(u)
Tv <- exp(v)
Tu <- exp(1-v)
theta.star[i] <- mean(T2)
}
var.theta.star.est <- var(theta.star)
theta.star <- mean(theta.star)
c(theta.hat,theta.star)
c(var.theta.hat.est,var.theta.star.est)
(var.theta.hat.est-var.theta.star.est)/var.theta.hat.est
```

运行结果 Rcode 最后一行的结果为

```
> (var.theta.hat.est-var.theta.star.est)/var.theta.hat.est
[1] 0.9678651
```

结论 由运行结果可见，模拟的方差缩减百分比与理论推导的很接近.

5.11

$\hat{\theta}_2 + c(\hat{\theta}_1 - \hat{\theta}_2)$

分析过程 $\hat{\theta}_c = c\hat{\theta}_1 + (1-c)\hat{\theta}_2$ 的方差为

$$\mathrm{Var}(\hat{\theta}_c) = \mathrm{Var}(\hat{\theta}_2) + c^2\mathrm{Var}(\hat{\theta}_1 - \hat{\theta}_2) + 2c\mathrm{Cov}(\hat{\theta}_2, \hat{\theta}_1 - \hat{\theta}_2).$$

$\mathrm{Var}(\hat{\theta}_c)$ 可看成关于 c 的二次函数 $V(c)$，于是当

$$c = c^* = \frac{2\mathrm{Cov}(\hat{\theta}_2, \hat{\theta}_1 - \hat{\theta}_2)}{-2\mathrm{Var}(\hat{\theta}_1 - \hat{\theta}_2)} = \frac{\mathrm{Var}(\hat{\theta}_2) - \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1) + \mathrm{Var}(\hat{\theta}_2) - 2\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)}$$

时，$\mathrm{Var}(\hat{\theta}_c)$ 最小.

结论 当 $c = c^* = \dfrac{\mathrm{Var}(\hat{\theta}_2) - \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1) + \mathrm{Var}(\hat{\theta}_2) - 2\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)}$ 时，$\hat{\theta}_c$ 的方差最小.

5.13

分析过程 我们取函数 $f_1(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{(x-\sqrt{2})^2}{2}}$（即正态分布 $\mathrm{N}(\sqrt{2},1)$ 的密度函数，$-\infty < x < +\infty$），$f_2(x) = e^{-x}$（$0 < x < +\infty$）. 它们都支撑在 $(0, +\infty)$ 上且比较接近于 $g(x)$. 用 R 生成函数图像并计算重要抽样法的估计值和方差，再做出判断.

Rcode

```
g <- function(x) { # g(x)
(x^2*exp(-x^2/2)/sqrt(2*pi))*(x>1) # note that x>1
}
x <- seq(1.05, 5, 0.05)
y <- g(x)
par(mfrow=c(1,2))
par(pin=c(2,1))
plot(x,y,type = 'l',ylim = c(0,0.5),ylab="")
lines(x,dnorm(x,mean = sqrt(2),1),lty=2)
lines(x,exp(-(x)),lty=3)
legend(x='topright', legend = c(expression(g), expression(f_1), ex
pression(f_2)),
lty = c(1,2,3),bty = 'n',cex = 0.8)
plot(x,y/dnorm(x,mean = sqrt(2),1),type = 'l',lty=2,ylim = c(0,2),ylab="")
lines(x,g(x)/exp(-(x)),lty=3)
```

```
legend(x='topright', legend = c(expression(g/f_1), expression(g/f_2)),
lty = c(2,3),bty = 'n',cex = 0.8)
par(mfrow=c(1,1))
m <- 10000
theta.hat <- se <- numeric(2)
x <- rnorm(m,mean = sqrt(2),1) # f_1
i <- c(which(x>1),which(x<0))
x[i] <- 2 # to catch overflow errors in g(x)
g.f <- g(x)/dnorm(x,mean = sqrt(2), 1) # g/f
theta.hat[1] <- mean(g.f)
se[1] <- sd(g.f)
x <- rexp(m,1) # f_2
g.f <- g(x)/exp(-x) # g/f
theta.hat[2] <- mean(g.f)
se[2] <- sd(g.f)
```

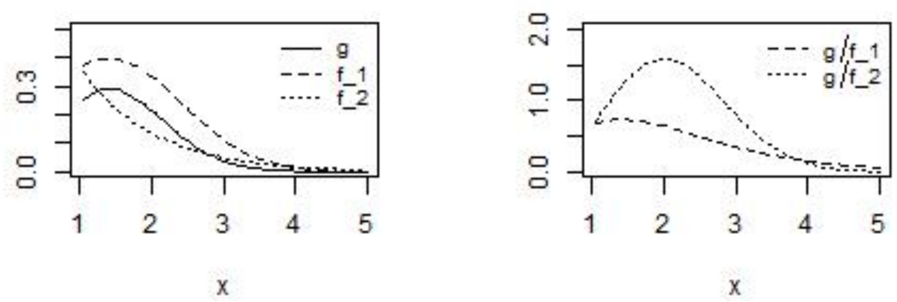

图 1: 左：练习 5.13 中的重要函数 $f_1(x)$，$f_2(x)$ 和函数 $g(x)$ 的图像；右：比例 $g(x)/f_i(x)$，$i = 1, 2$

运行结果 R 输出的图像如图 1所示，并且 theta.hat 和 se 的值如下：

```
> theta.hat
[1] 0.4722297 0.4017991
> se
[1] 0.2837117 0.5869486
```

结论 从图 1可以明显看出，$g(x)/f_1(x)$ 更接近于一个常数，但 $f_1(x)$ 的支撑集为整个实轴，有相当一部分的比例 $g(x)/f_1(x)$ 为 0，所以 R 给出的结果是 f_2 的方差要更小一些.

5.15

分析过程 课本中给出的密度函数似乎有误，因为 $\int_{\frac{j-1}{5}}^{\frac{j}{5}} \frac{5e^{-x}}{1-e^{-1}}dx \neq 1, j=1,\ldots,5$. 因此考虑新的密度函数

$$\frac{e^{-x}}{e^{-\frac{j-1}{5}} - e^{-\frac{j}{5}}}, \frac{j-1}{5} < x < \frac{j}{5}.$$

若记 $F_j(x) = \int_{\frac{j-1}{5}}^{x} f_j(s)ds$，则容易验证 $F_j(\frac{j}{5}) = 1$ 对每个 j 都成立. 此外，计算可得 $F_j^{-1}(x) = \frac{j}{5} - \log[e^{\frac{1}{5}}(1-x)+x]$，从而可以用逆变换法在每个 $(\frac{j-1}{5}, \frac{j}{5})$ 区间上生成密度函数为 f_i 的随机数，进而用重要抽样法算出每个区间上的积分值，求和即可算出 $(0,1)$ 区间上的积分值.

Rcode

```
g <- function(x) { # g(x)
exp(-x)/(1+x^2)
}
n <- 5
f <- function(x,n,j) { # density function f_j(x)
exp(-x)/(exp((1-j)/n)-exp(-j/n))
}
m <- 10000
theta.hat <- se <- numeric(n)
stra.est <- function(m,n,j) {
u <- runif(m)
x <- j/n-log(u+exp(1/n)*(1-u))
g.f <- g(x)/f(x,n,j)
theta.hat <- mean(g.f)
se <- sd(g.f)
return(c(theta.hat,se))
}
stra.est1 <- function(j) {
stra.est(m,n,j)
}
j <- seq(1,n)
stra.result <- as.data.frame(lapply(j,stra.est1),
row.names = c("theta_i.est","sd_i"),
col.names = seq(1,5))
apply(stra.result,1,sum)
```

运行结果 Rcode 最后一行的运行结果为

```
> apply(stra.result,1,sum)
theta_i.est      sd_i
 0.52482989  0.01700146
```

结论 运行结果部分的 theta_i.est 与例 5.10 中的结果很接近，而且标准差 sd_i 要更小.