统计计算 张楠 2019秋: Variance Reduction: Control Variates

(返回 统计计算 张楠 2019秋)

Control Variates

Monte Carlo 估计中另外一种减少方差的方法是控制变量(Control Variates)的使用. 设要估计的量为$\theta = E[g(X)]$, f为一个函数, 其期望$E[f(X)] = \mu$已知, 且 f和g相关.

对任何常数c, 估计量$\hat{\theta}_c = g(X) + c(f(X) - \mu)$ 为无偏的, 且方差为

$$Var(\hat{\theta}_c) = Var(g(X)) + c^2 Var(f(X)) + 2cCov(g(X), f(X))$$

对c最小化, 则最小值在

$$c^* = -\frac{Cov(g(X), f(X))}{Var(f(X))}$$

处达到, 且最小值为

$$Var(\hat{\theta}_{c^*}) = Var(g(X)) - \frac{[Cov(g(X), f(X))]^2}{Var(f(X))}.$$

随机变量$f(X)$称为$g(X)$的一个控制变量(Control Variate). 显然方差的减少率为

$$100\frac{[Cov(g(X), f(X))]^2}{Var(g(X))Var(f(X))} = 100[Cor(g(X), f(X))]^2.$$

可以看出, 这种方法在f和g强相关时是有优势的, 若f和g不相关, 则不会导致方差减少.

> **例7**: 控制变量方法 使用控制变量方法计算积分
> $\theta = E[e^U] = \int_0^1 e^u du,$

其中$U \sim U(0, 1)$.

此例中,虽然积分值为$\theta = e - 1 \doteq 1.718282$, 我们仍然使用控制变量Monte Carlo方法来计算积分, 用以说明这种方法的使用. 如果使用简单的Monte Carlo积分方法, 则方差为

$$Var(g(U)) = Var(e^U) = \frac{e^2-1}{2} - (e-1)^2 \doteq 0.2420351.$$

重复m次得到的估计的方差为$Var(g(U))/m$.

> 控制变量的自然选择为$U \sim U(0, 1)$, 则$Cov(e^U, U) = 1 - (e-1)/2 \doteq 0.1408591$. 因此
> $c^* = \frac{-Cov(e^U, U)}{Var(U)} = -12 + 6(e-1) \doteq -1.690309.$

而使用控制变量方法得到的估计为$\hat{\theta}_{c^*} = e^U - 1.690309(U - 0.5)$, m次重复后的方差$Var(\hat{\theta}_{c^*})/m$, 其中$Var(\hat{\theta}_{c^*})$为

$$Var(e^U) - \frac{[Cov(e^U,U)]^2}{Var(U)} = \frac{e^2-1}{2} - (e-1)^2 - 12[1 - \frac{e-1}{2}]^2 \doteq 0.003940175.$$

因 此 使 用 控 制 变 量 方 法 导 致 简 单 Monte Carlo 估 计 量 的 方 差 减 少 率 为 $100(1 - 0.003940175/0.2429355) = 98.3781\%$.

下面我们使用控制变量方法来计算其经验的方差减少率.

```
x <- seq(.1, 2.5, length=5)
Phi <- pnorm(x)
set.seed(123)
MC1 <- MC.Phi(x, anti = FALSE)
set.seed(123)
MC2 <- MC.Phi(x)
print(round(rbind(x, MC1, MC2, Phi), 5))
```

方差的减少量可以模拟来比较:

```
m <- 10000
a <- - 12 + 6 * (exp(1) - 1)
U <- runif(m)
T1 <- exp(U)                    #simple MC
T2 <- exp(U) + a * (U - 1/2)    #controlled
mean(T1)
mean(T2)
(var(T1) - var(T2)) / var(T1)
```

例8: 使用控制变量方法的Monte Carlo积分 使用控制变量方法计算积分 $\int_0^1 \frac{e^{-x}}{1+x^2}dx$.

此例中感兴趣的量为$\theta = Eg(X), g(x) = \frac{e^{-x}}{1+x^2}$, 其中$X \sim U(0,1)$. 我们 要寻求一个足够接近$g$的函数$f$且其期望值要已知, 和$g$相关. 比如$f(x) = \frac{e^{-0.5}}{(1+x^2)}$ 是可以的, 若$U \sim U(0,1)$, 则

$$Ef(U) = e^{-0.5}\int_0^1 \frac{1}{1+u^2}du = e^{-0.5}\frac{\pi}{4}.$$

我们也可以估计出$Cor(g(U), f(U)) \approx 0.974$. 因此

```
f <- function(u)   exp(-.5)/(1+u^2)
g <- function(u)   exp(-u)/(1+u^2)
set.seed(510) #needed later
u <- runif(10000)
B <- f(u)
A <- g(u)
a <- -cov(A,B) / var(B)      #est of c*

m <- 100000
u <- runif(m)
T1 <- g(u)
T2 <- T1 + a * (f(u) - exp(-.5)*pi/4)
c(mean(T1), mean(T2))
c(var(T1), var(T2))
(var(T1) - var(T2)) / var(T1)
```

Antithetic variate as control variate

对偶变量方法实际上是控制变量方法的特例. 注意到控制变量方法 是无偏估计的线性组合. 一般地, 若$\hat{\theta}_1$和$\hat{\theta}_2$ 为θ的无偏估计量, 则对任何常数c,有

$$\hat{\theta}_c = c\hat{\theta}_1 + (1-c)\hat{\theta}_2$$

仍为θ的无偏估计. 其方差为

$$Var(\hat{\theta}_2) + c^2 Var(\hat{\theta}_1 - \hat{\theta}_2) + 2cCov(\hat{\theta}_2, \hat{\theta}_1 - \hat{\theta}_2).$$

特别地, 当$\hat{\theta}_1$和$\hat{\theta}_2$同分布, 且$Cor(\hat{\theta}_1, \hat{\theta}_2) = r$. 则$Cov(\hat{\theta}_1, \hat{\theta}_2) = rVar(\hat{\theta}_1)$,此时最优的$c^* = 1/2$. 此时控制变量估计量为

$$\hat{\theta}_{c^*} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}$$

这也是(这种特定选择下)对偶变量方法下的估计量.

Several control variates

将无偏估计量组合起来作为参数θ的估计,以减少方差的方法可以推广到多个控制变量场合:

$$\hat{\theta}_c = g(X) + \sum_{i=1}^k c_i (f_i(X) - \mu_i))$$

其中$\mu_i = Ef_i(X), i = 1, \cdots, k., \sum_{i=1}^k c_i = 1$ 以及

$$E\hat{\theta}_c = E[g(X)] + \sum_{i=1}^k c_i E[f_i(X) - \mu_i] = \theta$$

控制变量方法下的估计量$\hat{\theta}_{\hat{c}}$以及最优的c_i可以通过回归模型来估计.

在 $k=1$ 场合, 考虑二元样本$((g(X_1), f(X_1)), \cdots, (g(X_n), f(X_n)))$, 假设$g(X)$与$f(X)$之间存在线性关系: $g(X) = \beta_0 + \beta_1 f(X) + e$, 且$E[g(X)] = \beta_0 + \beta_1 E[f(X)]$.

β_1的最小二乘估计为

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (g(X_i) - g(\bar{X}))(f(X_i) - f(\bar{X}))}{\sum_{i=1}^n (f(X_i) - f(\bar{X}))^2} = \frac{\hat{Cov}(g(X), f(X))}{\hat{Var}(f(X))} = -c^*.$$

这说明我们可以使用$g(X)$对$f(X)$的回归来估计c:

```
L<-lm(gx~fx)
c.star<--L$coef[2]
```

截距的最小二乘估计为$\hat{\beta}_0 = \overline{g(X)} - (-c^*)\overline{f(X)}$, 因此在$\mu$处的预测值为

$$\hat{\beta}_0 + \hat{\beta}_1 \mu = \overline{g(X)} + \hat{c}^* (\overline{f(X)} - \mu) = \hat{\theta}_{c^*}$$

即控制变量方法下的估计量是预测值.

误差方差的估计为

$$\hat{\sigma}_e^2 = \hat{Var}(g(X) - \hat{g}(X)) = \hat{Var}(g(X) - (\beta_0 + \beta_1 f(X))) = \hat{Var}(g(X) + c^* f(X))$$

控制变量方法下的估计量的方差估计为

$$\hat{Var}(\overline{g(X)} + \hat{c}^* (\overline{f(X)} - \mu)) = \frac{g(X) + \hat{c}^* f(X)}{n} = \frac{\hat{\sigma}_e^2}{n}.$$

因此在R中, 控制变量方法下的估计量的标准差的估计为

```
se.hat<-summary(L)$sigma/sqrt(n)
```

和前面控制变量一节中的结果相同.

对一般的k, 则可以使用回归

$$g(X) = \beta_0 + \sum_{i=1}^{k} \beta_i f(X) + e$$

来估计最优的常数$c^* = (c_1^*, \cdots, c_k^*)$. 则$-c^* = (\hat{\beta}_1, \cdots, \hat{\beta}_k)$, 以及此时控制变量方法下的估计为在$\mu = (\mu_1, \cdots, \mu_k)$处的预测值$\hat{g}(X)$. 估计的方差为$\sigma_e^2/n$.

例9: 控制变量和回归 使用回归方法估计积分

$$g(x) = \int_0^1 \frac{e^{-x}}{1+x^2} dx.$$

这里控制变量取$f(x) = e^{.5}(1+x^2)^{-1}, 0 < x < 1.$, $\mu = E[f(X)] = e^{.5}\pi/4$. 为估计最优常数$c^*$,

```
set.seed(510)
u <- runif(10000)
f <- exp(-.5)/(1+u^2)
g <- exp(-u)/(1+u^2)
L <- lm(g~f)
c.star <-  - L$coeff[2]   # beta[1]
mu <- exp(-.5)*pi/4
c.star
theta.hat <- sum(L$coeff * c(1, mu))  #pred. value at mu
theta.hat
summary(L)$sigma^2
summary(L)$r.squared
```

这里我们使用了和前例中同样的种子, 因此得到同样的c^*估计. 现在$\hat{\theta}_{\hat{c}^*}$是 在$\mu = 0.4763681$处的预测值. 而估计量$\hat{\theta}$ 及其标准差, 方差的减少率都和前例相同.
