$n^{-1}\sum_{i=1}^{n}(\alpha X_i+(1-\alpha)Y_i)$。容易验证当$\alpha=\mathrm{Cov}(Y,Y-X)/\mathrm{Var}(Y-X)$时$\mathrm{Var}(\hat{\mu}_\alpha)$最小。（验证一下最后一句话是否正确？？）。

例5.2.1 利用对偶变量法，根据$[0,1]$均匀随机样本$U_1,U_2,\cdots,U_n$，估计$I=\int_0^1 e^t\,dt=E(e^U)$, 并计算估计量的方差。

上例中$X=e^U$, 选对偶变量$Y=e^{1-U}$, 则用对偶变量法估计I 为

$$\hat{I}=\frac{1}{2n}\sum_{i=1}^{n}(e^{U_i}+e^{1-U_i}), \tag{5.2.4}$$

方差为

$$\mathrm{Var}(\hat{I})=\frac{1}{2n}\mathrm{Var}(e^U)+\mathrm{Cov}(e^U,e^{1-U})=\frac{1}{n}(-\frac{3}{4}e^2+\frac{5}{2}e-\frac{5}{4})\approx\frac{0.00391}{n}. \tag{5.2.5}$$

样本量$n=20$的情况下，用对偶变量法估计$E(e^U)$的程序如下：

```
n=20
hat_mu1<-NULL
hat_mu2<-NULL
for (i in 1:1000){
U=runif(n,0,1)
X=exp(U)
hat_mu1[i]=mean(X)
Y=exp(1-U)
hat_mu2[i]=0.5*mean(X+Y);
}
c(mean(hat_mu1),mean(hat_mu2),var(hat_mu1),var(hat_mu2))
```

例5.2.2 利用对偶变量法，估计$I=Ee^{X_1+X_2}$, 这里$X_1\sim N(\mu_1,\sigma_1^2)$和$X_2\sim N(\mu_2,\sigma_2^2)$独立。)

作业：用对偶变量法给出参数I的估计。

5.3 控制变量法的应用

首先给出控制变量的定义。

定义5.3.1 （控制变量）对于变量X，如果变量Y的期望为0, 且$\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)\le 0$, 则称$Y$为$E(X)$通过$E(X+Y)$估计的控制变量。

本节通过控制变量Y降低期望$\mu = EX$的估计。从X中抽取n个独立样本值$X_1, X_2, \ldots, X_n$, 则平均值估计为$\hat{\mu}_1 = n^{-1}\sum_{i=1}^n X_i$。注意到$E(Y) = 0$, 则$\mu = E(X+Y)$。基于样本$X_1, X_2, \ldots, X_n$, 如果相应的可得$Y_1, Y_2, \cdots, Y_n$，则用控制变量估计$\mu = EX$为$\hat{\mu}_2 = n^{-1}\sum_{i=1}^n (X_i + Y_i)$. 简单计算可得两个估计的方差分别为

$$\mathrm{Var}(\hat{\mu}_1) = n^{-1}\mathrm{Var}(X), \quad \mathrm{Var}(\hat{\mu}_2) = n^{-1}(\mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)).$$

只要$\mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) \le 0$, 则$\mathrm{Var}(\hat{\mu}_2) \le \mathrm{Var}(\hat{\mu}_1)$. 因此控制变量$Y$可以降低$E(X)$估计的方差。

注意到$E(X) = E(X+\beta Y)$, 可以通过$\hat{\mu}_\beta = n^{-1}\sum_{i=1}^n (X_i+\beta Y_i)$估计$E(X)$，相应可得

$$\mathrm{Var}(\hat{\mu}_\beta) = n^{-1}(\mathrm{Var}(X) + \beta^2\mathrm{Var}(Y) + 2\beta\mathrm{Cov}(X,Y)).$$

通过关于β最小化$\mathrm{Var}(\hat{\mu}_\beta)$, 得到使得方差最小的$\beta_0$. 也就是关于$\beta$最小化$\beta^2\mathrm{Var}(Y)+2\beta\mathrm{Cov}(X,Y)$. 不难得到

$$\beta_0 = -\mathrm{Cov}(X,Y)/\mathrm{Var}(Y)$$

此时

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_{\beta 0}) &= n^{-1}(\mathrm{Var}(X) + \beta^2\mathrm{Var}(Y) + 2\beta\mathrm{Cov}(X,Y)) \\
&= n^{-1}(\mathrm{Var}(X) - \mathrm{Cov}(X,Y)^2/\mathrm{Var}(Y)) \le n^{-1}\mathrm{Var}(X) = \mathrm{Var}(\hat{\mu}_1)
\end{aligned}$$

根据上式, Y和X 的相关性越强，降低估计的方差幅度越大。这种减小估计方差的方法叫做控制变量法。

在实际问题中，由于$\mathrm{Cov}(X,Y)$和$\mathrm{Var}(Y)$未知，通过选取一部分样本数据估计，从而得到β_0的估计

$$\hat{\beta}_0 = -\widehat{\mathrm{Cov}}(X,Y)/\widehat{\mathrm{Var}}(Y)$$

如果$EY \ne 0$但$EY = \mu_Y$ 已知，只要用$Y - \mu_Y$ 代替Y，上述控制变量方法做相应的修改即可。

例5.3.1 基于正态分布$N(\mu,\sigma^2)$随机样本$x_1,x_2,\cdots,x_n$，用控制变量法估计$E(X)$。

模拟设置，样本$n=100$，$X\sim N(5,2)$，产生正态分布$Y\sim N(0,1)$且X和Y的相关系数为-0.8，则$\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)\le 0$.

```
library(MASS)
n=100
hat_mu1<-NULL
hat_mu2<-NULL
Mu=c(5,0)
Sigma2=matrix(c(2,-0.8,-0.8,1),ncol=2,nrow=2,byrow=T)
for (i in 1:1000){
  Data=mvrnorm(n,Mu,Sigma2)
  X=Data[,1]
hat_mu1[i]=mean(X);
Y=Data[,2]
hat_mu2[i]=mean(X+Y);
}
c(mean(hat_mu1),mean(hat_mu2),var(hat_mu1),var(hat_mu2))
```

例5.3.2 基于$[0,1]$均匀分布随机样本$U_1,U_2,\cdots,U_n$，用控制变量法估计$I=\int_0^1 e^t\,dt=Ee^U$。当然，可以得到积分真值为$e-1$。

设$U\sim \mathrm{U}(0,1)$，$X=e^U$，则$I=Ee^U=EX$，可以用平均值法估计I为

$$\hat{I}_1=\frac{1}{N}\sum_{i=1}^{N}e^{U_i}. \tag{5.3.6}$$

其方差为

$$\mathrm{Var}(\hat{I}_1)=\frac{1}{N}\mathrm{Var}(e^U)=\frac{1}{N}(-\frac{1}{2}e^2+2e-\frac{3}{2})\approx\frac{0.2420}{N}. \tag{5.3.7}$$

令$Y=U-\frac{1}{2}$，则$EY=0$，X与Y正相关，可以计算出$\mathrm{Cov}(X,Y)\approx 0.14086$，$\mathrm{Var}(Y)=1/12$(更复杂的问题中可能需要从一个小的随机抽样中近似估

计)，于是$b = -\mathrm{Cov}(X,Y)/\mathrm{Var}(Y) = -1.690$, 对$Z(b) = e^U - 1.690(U - \frac{1}{2})$有$\mathrm{Var}(Z(b)) = [1-\rho^2_{X,Y}]\mathrm{Var}(X) = (1-0.9919^2)\mathrm{Var}(X) = 0.016\mathrm{Var}(X) = 0.0039$, 用控制变量法估计I为

$$\hat{I}_2 = \frac{1}{N}\sum_{i=1}^{N}\left[e^{U_i} - 1.690(U_i - \frac{1}{2})\right].$$

$\hat{I}_1$的方差比控制变量法$\hat{I}_2$的方差大60倍以上。

5.4 条件期望法

本节通过条件期望法降低期望$\mu = EX$的估计。从X中抽取n个独立样本值$X_1, X_2, \ldots, X_n$, 则平均值估计为$\hat{\mu}_1 = n^{-1}\sum_{i=1}^n X_i$。注意到$E(X) = E\{E(X|Y)\} = E(g(Y))$, 基于样本$X_1, X_2, \ldots, X_n$, 如果相应的可得$Y_1, Y_2, \cdots, Y_n$，则用条件期望法估计$\mu = EX$为$\hat{\mu}_2 = n^{-1}\sum_{i=1}^n g(Y_i)$. 相应的两个估计方差分别为

$$\mathrm{Var}(\hat{\mu}_1) = n^{-1}\mathrm{Var}(X), \quad \mathrm{Var}(\hat{\mu}_2) = n^{-1}\mathrm{Var}(g(Y)).$$

根据Rao-Blackwell不等式:

$$\mathrm{Var}(g(Y)) = \mathrm{Var}\left\{\mathrm{E}(X|Y)\right\} \le \mathrm{Var}(X),$$

可得$\mathrm{Var}(\hat{\mu}_2) \le \mathrm{Var}(\hat{\mu}_1)$, 也就是通过条件期望估计$E(X)$比平均值估计的方差要小。这种降低估计方差的方法叫做条件期望法，或Rao-Blackwell 方法。

例5.4.1 用条件期望法估计EX, 这里X是均值为μ_X, 方差为σ_X^2的正态分布。

注意到$(X,Y)^\tau \sim N(\mu, \Sigma)$, 这里$\mu = (\mu_X, \mu_Y)^\tau$, $\Sigma = (\sigma_X^2, \rho_{XY}; \rho_{XY}, \sigma_Y^2)$, 则$E(X|Y) = N(\mu_{X|Y}, \Sigma_{X|Y})$, 其中$\mu_{X|Y} = \mu_X + \rho_{XY}\sigma_Y^{-2}(Y - \mu_Y)$, $\Sigma_{X|Y} = \sigma_X^2 - \rho_{XY}^2\sigma_Y^{-2}$.

例5.4.2 作业：用条件期望估计的例子

增加相应的模拟

5.5 重要抽样法

对于EX的估计，假设X的密度函数是$f(x)$，则$\mu = EX$ 可以写为

$$\mu = \int xf(x)dx = \int g(x)\frac{xf(x)}{g(x)}dx = E\frac{Yf(Y)}{g(Y)},$$

其中$g(x)$是某个随机变量Y的密度函数, 也就是

$$g(x) \geq 0; \int g(x)dx = 1.$$

则如果Y的随机数$Y_1, Y_2, \cdots, Y_n$很容易产生，$\mu = EX$可以通过下式估计

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i f(Y_i)}{g(Y_i)}. \tag{5.5.8}$$

通过计算不难得到

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_2) &= \frac{1}{n}\mathrm{Var}\frac{Yf(Y)}{g(Y)} \\ &= \frac{1}{n}\int (\frac{xf(x)}{g(x)} - \mu)^2 g(x)dx = \int \frac{x^2 f^2(x)}{g(x)}dx - \mu^2.\end{aligned}$$

根据上式，只要$x^2 f^2(x)/g^2(x) = \mu^2$, 可得$\mathrm{Var}(\hat{\mu}_2) = 0$. 如果$f(x)$的非零取值部分对应$x > 0$, 则$xf(x)/g(x) = \mu$, 也就是$g(x) = xf(x)/\mu$. 但实际上$\mu$是未知量，无法选取$g(x)$，使得$\mathrm{Var}(\hat{\mu}_2) = 0$。但是我们可以通过选择和曲线$xf(x)$走势一致的密度函数做为$g(x)$.

根据(5.5.8)式得到的估计就是重要抽样法估计。(5.5.8)式还可写成

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i f(Y_i)\omega(Y_i),$$

其中$\omega(Y_i) = 1/g(Y_i)$称为重要抽样的权因子。

例5.5.1 用重要抽样法计算积分 $\mu = \int_0^1 xf(x)dx = \int_0^1 e^x dx$ 的估计值.

解 由于 $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$, 取 $g(x) = 2(1+x)/3$, $x \in [0,1]$ (用展开式的前二项 $1+x$ 作为 e^x 的近似式,(1+x) 前面的系数是归一化常数). 则μ的重要抽样法估计为

$$\hat{\mu}_2 = \frac{3}{2n}\sum_{i=1}^{n}\frac{e^{x_i}}{1+x_i},$$

上式中$x_1, x_2, \cdots, x_n$是密度为 $g(x)$ 的随机数.

重要抽样法步骤如下：

①产生均匀随机数 $r_i(i = 1, 2, \cdots, N)$;

②用逆变换法产生 $g(x)$ 随机数,即由 r_i 计算 $x_i = \sqrt{3r_i + 1} - 1$,则 $x_i \sim g(x)$;

③计算 $\theta_3 = \frac{3}{2N} \sum\limits_{i=1}^{N} \frac{e^{x_i}}{1+x_i}$, 则 θ_3 是 I 的估计量.

下面我们计算估计量 $\hat{\mu}_2$ 的方差 $Var(\hat{\mu}_2)$.显然 $\mu = \int_0^1 e^x dx = \quad e - 1$ 是积分的真值.而

$$\int_0^1 \frac{x^2 f^2(x)}{g(x)} dx = \frac{3}{2} \int_0^1 \frac{e^{2x}}{1+x} dx = \frac{3}{2e^2} \int_0^1 \frac{e^{2(x+1)}}{1+x} dx = \frac{3}{2e^2} \int_2^4 \frac{e^t}{t} dt.$$

记 $E(x) = \int_0^z \frac{e^t}{t} dt$,故有

$$Var(\hat{\mu}_2) = \frac{1}{N}[\frac{3}{2e^2}(E(4) - E(2)) - (e-1)^2] = \frac{0.0269}{N}$$

如果用平均值估计法,估计量的方差为

$$Var(\hat{\mu}_1) = \frac{1}{N}[\int_0^1 e^{2x} dx - I^2] = \frac{0.2420}{N}.$$

显然有 $Var(\hat{\mu}_2) < Var(\hat{\mu}_1)$.

5.6 分层抽样法

分层抽样法的基本想法与重要抽样法相似,它们都是使得对积分值贡献大的抽样更多的出现.但分层抽样法的作法与重要抽样法不同,它并不改变原来的概率分布,而是将抽样区间分成一些小区间,在各小区间内的抽样点数根据贡献大小决定,使得对积分值贡献大的抽样更多地出现,以便提高抽样效率.

考虑积分 $\mu = \int_0^1 f(x)dx$,将积分区间 $[0, 1]$ 用分点 $a_i(i = 0, \quad 1, \cdots, m)$ 分成 m 个互不相交的子区间,其长度分别为

$$\mu_i = a_i - a_{i-1} \quad (i = 1, 2, \cdots, m; a_0 = 0, a_m = 1),$$

于是 $\mu = \int_0^1 f(x)dx = \sum\limits_{i=1}^{m} \int_{a_{i-1}}^{a_i} f(x)dx = \sum\limits_{i=1}^{m} \mu_i$.

用平均值估计法求每个小区间 $[a_{i-1}, a_i]$ 的积分值 $\mu_i(i = 1, 2, \quad \cdots, m)$.具体做法是首先产生 n_i 个 $[a_{i-1}, a_i]$ 区间上的均匀随机数:$r_j^{(i)} = a_{i-1} + l_i r_j (j =$

$1,2,\cdots,n_i;r_j\sim U(0,1))$,于是有

$$\mu_i=\int_{a_{i-1}}^{a_i}f(x)dx=l_i\int_{a_{i-1}}^{a_i}\frac{f(x)}{l_i}dx\approx\frac{l_i}{n_i}\sum_{j=1}^{n_i}f(r_j^{(i)})\triangleq\hat{\mu}^{(i)};$$

$$\mu\approx\sum_{i=1}^{m}\theta^{(i)}=\sum_{i=1}^{m}\frac{l_i}{n_i}\sum_{j=1}^{n_i}f(r_j^{(i)})\triangleq\hat{\mu}.$$

显然容易验证:$E(\hat{\mu})=\sum\limits_{i=1}^{m}\frac{l_i}{n_i}\sum\limits_{j=1}^{n_i}E(f(r_j^{(i)}))=\sum\limits_{i=1}^{m}\frac{l_i}{n_i}\sum\limits_{j=1}^{n_i}\frac{I_i}{l_i}=\sum\limits_{i=1}^{m}I_i=I$. 这表明 $\hat{\mu}$ 是积分值 I 的无偏估计量,且估计量 $\hat{\mu}$ 的方差为

$$\begin{aligned}Var(\hat{\mu})&=Var(\sum_{i=1}^{m}\frac{l_i}{n_i}\sum_{j=1}^{n_i}f(r_j^{(i)}))=\sum_{i=1}^{m}\frac{l_i^2}{n_i^2}\sum_{j=1}^{n_i}Var(f(r_j^{(i)}))\\&=\sum_{i=1}^{m}\frac{l_i^2}{n_i}Var(f(r^{(i)}))\quad(r^{(i)}\sim U(a_{i-1},a_i)).\end{aligned}$$

其中

$$\begin{aligned}Var(f(r^{(i)}))&=E[f^2(r^{(i)})]-[E(f(r^{(i)}))]^2\\&=\int_{a_{i-1}}^{a_i}\frac{1}{l_i}f^2(x)dx-(\frac{I_i}{l_i})^2.\end{aligned}$$

例5.6.1 用分层抽样法求 $I=\int_0^1 e^x dx$.

解 由 e^x 在 $[0,1]$ 上的图形可知,靠近 1 的区域对积分值的贡献大,而靠近 0 的区域对积分值的贡献小.我们将积分区间 $[0,1]$ 等分成两个子区间.在 $[0,0.5]$ 上抽样 4 次,得 $r_k^{(1)}=0.5r_j(r_j\sim U(0,1);j=1,2,3,4)$.在 $[0.5,1]$ 上抽样 6 次,得 $r_k^{(2)}=0.5+0.5r_k(r_k\sim U(0,1);k=1,2,\cdots,6)$, 共抽样 $n=10$ 次,由分层抽样公式得

$$\hat{\mu}=\frac{1}{8}\sum_{j=1}^{4}e^{r_j^{(1)}}+\frac{1}{12}\sum_{j=1}^{6}e^{r_j^{(2)}}.$$

估计量 $\hat{\mu}$ 的方差为

$$Var(\hat{\mu})=\frac{1}{16}Var(e^{r^{(1)}})+\frac{1}{24}Var(e^{r^{(2)}}),$$

其中

$$\begin{aligned} Var(e^{r^{(1)}}) =& E(e^{2r^{(1)}}) - [E(e^{r^{(1)}})]^2 \quad (r^{(1)} \sim U(0, 0.5)) \\ =& \int_0^{0.5} 2 \cdot e^{2r} dr - [\int_0^{0.5} 2 \cdot e^r dr]^2 \\ =& (e-1) - 4(\sqrt{e}-1)^2 = 0.03492; \end{aligned}$$

类似可得: $Var(e^{r^{(2)}}) = 0.09493$.

因此

$$Var(\hat{\mu}) = \frac{1}{16} \times 0.03492 + \frac{1}{24} \times 0.09493 = 0.006138.$$

5.7 随机数重复使用

在统计研究中，经常需要比较若干种统计方法的性能，如偏差、方差、覆盖率等。除了努力获取理论结果以外，可以用随机模拟方法进行比较：重复生成N组随机样本，对每个样本同时用不同统计方法计算结果，最后从N组结果比较不同方法的性能。这样比较时，并没有对每种方法单独生成N组样本，而是每个样本同时应用所有要比较的方法。这样不仅减少了计算量，而且在比较时具有更高的精度。

例5.7.1 对正态分布总体$X \sim N(\mu, \sigma^2)$, 如果有样本$X_1, X_2, \dots, X_n$, 估计σ^2有两种不同的公式:

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \hat{\sigma}_2^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

$$\begin{aligned} b_1 =& E\hat{\sigma}_1^2 - \sigma^2, \quad b_2 = E\hat{\sigma}_2^2 - \sigma^2, \\ s_1 =& E(\hat{\sigma}_1^2 - \sigma^2)^2, \quad s_2 = E(\hat{\sigma}_2^2 - \sigma^2)^2. \end{aligned}$$

当然，这个问题很简单，可以得到偏差和均方误差的理论值：

$$\begin{aligned} b_1 =& 0, \quad b_2 = -\frac{1}{n}\sigma^2, \\ s_1 =& \frac{2\sigma^4}{n-1}, \quad s_2 = \frac{(2n-1)\sigma^4}{n^2}, \quad s_1 - s_2 = \frac{(3n-1)\sigma^4}{n^2(n-1)}. \end{aligned}$$

我们用随机模拟来作比较。重复地生成N组样本$(X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)}), j = 1, 2, \dots, N$。对每组样本分别计算$\hat{\sigma}_{1j}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^{(j)} - \bar{X}^{(j)})^2$和 $hat\sigma_{2j}^2 =$

$\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)} - \bar{X}^{(j)})^2$，得到偏差和均方误差的估计

$$\hat{b}_1 = \frac{1}{N}\sum_{j=1}^{N}\hat{\sigma}_{1j}^2 - \sigma^2, \quad \hat{b}_2 = \frac{1}{N}\sum_{j=1}^{N}\hat{\sigma}_{2j}^2 - \sigma^2,$$
$$\hat{s}_1 = \frac{1}{N}\sum_{j=1}^{N}(\hat{\sigma}_{1j}^2 - \sigma^2)^2, \quad \hat{s}_2 = \frac{1}{N}\sum_{j=1}^{N}(\hat{\sigma}_{2j}^2 - \sigma^2)^2.$$

容易看出，两种方法使用相同的模拟样本得到的偏差、均方误差的估计精度与每种方法单独生成模拟样本得到的估计精度相同。

但是，如果要估计$\Delta s = s_1 - s_2$，利用相同的样本的估计精度更好。一般地，

$$\text{Var}(\hat{s}_1 - \hat{s}_2) = \text{Var}(\hat{s}_1) + \text{Var}(\hat{s}_2) - 2\text{Cov}(\hat{s}_1, \hat{s}_2). \tag{14.1}$$

如果每种方法使用不同的样本，则(14.1)变成

$$\text{Var}(\hat{s}_1 - \hat{s}_2) = \frac{1}{N}\left[\text{Var}((\hat{\sigma}_1^2 - \sigma^2)^2) + \text{Var}((\hat{\sigma}_2^2 - \sigma^2)^2)\right]. \tag{14.2}$$

如果两种方法利用相同的样本计算，则(14.1)变成

$$\text{Var}(\hat{s}_1 - \hat{s}_2) = \frac{1}{N}\left[\text{Var}((\hat{\sigma}_1^2 - \sigma^2)^2) + \text{Var}((\hat{\sigma}_2^2 - \sigma^2)^2) - 2\text{Cov}((\hat{\sigma}_1^2 - \sigma^2)^2, (\hat{\sigma}_2^2 - \sigma^2)^2)\right]. \tag{14.3}$$

而$(\hat{\sigma}_1^2 - \sigma^2)^2$与$(\hat{\sigma}_2^2 - \sigma^2)^2$ 明显具有强正相关，所以两种方法针对相同样本计算时$\hat{s}_1 - \hat{s}_2$的方差比两种方法使用单独样本时的方差要小得多。这说明重复利用相同的随机数或样本往往可以提高比较的精度。可以计算出(14.2)约为$16n^{-2}/N$, (14.2)约为$104n^{-3}/N$。