第 5 章　减少方差的统计方法

5.1　估计定积分例子

5.1.1　随机投点法

设 $0 \le f(x) \le 1$,积分值 I 就是曲线 $f(x)$ 下方 $x = 0$ 和 $x = 1$ 之间的曲边梯形的面积.为求积分值 I, 我们首先构造随机投点概型,即向正方形 $\{0 \le x \le 1, 0 \le y \le 1\}$ 内随机投点$\{\xi_i, \eta_i\}(i = 1, 2, \cdots)$,其中 $\xi_i \sim U(0,1), \eta_i \sim U(0,1)$，且相互独立.

若第 i 个点落入曲边梯形内,即满足条件 $\eta_i \le f(\xi_i)$,则称第 i 次试验成功.随机投点试验成功的概率 p 为:

$$p = P\{\eta \le f(\xi)\} = \int_0^1 \int_0^{f(x)} dydx = \int_0^1 f(x)dx = I$$

因此 $I = \int_0^1 f(x)dx = p$.在随机投点试验的概型下,定积分值 I 就是试验成功的概率 p.

重复进行随机投点试验,记录试验次数 N 和成功次数 M,用频率 M/N 作为概率 p 的估计值,即可得出定积分值 I 的近似解:

$$I \approx M/N \tag{3.2}$$

记 $\theta_1 = M/N$,它是成功概率 p 的估计量.在 N 次试验中,成功次数 M 服从二项分布 $B(N,p)$, 故有

$$E(M) = Np,\ Var(M) = Np(1-p);$$

因此

$$E(\theta_1) = p,\ Var(\theta_1) = \frac{1}{N}p(1-p). \tag{3.3}$$

可见 θ_1 是 p 的无偏估计量.

算法3.1(随机投点法)

①赋初值:试验次数 $n=0$,成功次数 $m=0$;规定投点试验的总次数 N;

②产生两个相互独立的均匀随机数 ξ,η; 置 $n=n+1$;

③判断 $n\le N$ 是否成立,若成立转 ④, 否则停止试验,转 ⑤;

④判断条件 $\eta\le f(\xi)$ 是否成立,若成立置 $m=m+1$,然后转 ②;否则转 ②;

⑤计算 $\theta_1=m/N$,则 θ_1 为 I 的近似解.

对一般区间 $[a,b]$ 上定积分 $I=\int_a^b f(x)dx$ 的计算,只需做线性变换,即可化为 $[0,1]$ 区间上的积分.设 $f(x)$ 在 $[a,b]$ 上有界: $c\le f(x)\le d$,为了化一般区间上的积分为 $[0,1]$ 区间上的积分,且被积函数值在 $[0,1]$ 之间,令 $x=(b-a)u+a$,则有

$$\begin{aligned}\int_a^b f(x)dx &= \int_0^1[f(a+(b-a)u)-c+c](b-a)du\\ &= \int_0^1(b-a)(d-c)\frac{f(a+(b-a)u)-c}{d-c}du+c(b-a)\\ &= S_0\int_0^1\varphi(u)du+c(b-a),\end{aligned}$$

其中 $\varphi(u)=\frac{1}{d-c}[f(a+(b-a)u)-c]$,$S_0=(b-a)(d-c)$. 因 $c\le f(x)\le d$,故有 $0\le\varphi(u)\le 1$(当 $u\in(0,1)$ 时).故只需求解 $[0,1]$ 区间上的标准定积分值,即可得到一般区间 $[a,b]$ 上的定积分值.或者直接构造向长方形区域 $\{a\le x\le b,c\le y\le d\}$ 随机投点的试验模型.这时定积分值 I 为

$$I=\int_a^b f(x)dx=pS_0+c(b-a),$$

其中 $S_0=(b-a)(d-c)$ 是长方形的面积.

5.1.2 平均值估计法

(2)平均值估计法

设随机变量 $R\sim U(0,1)$,则 $Y=f(R)$ 的数学期望为

$$E(f(R))=\int_0^1 f(x)dx=I \tag{3.4}$$

这表明积分值 I 是随机变量 $Y=f(R)$ 的数学期望.可以用数学期望的估计值作为 I 的近似解.这时只需产生均匀随机数 $r_1, r_2, \cdots, \quad r_N$,则

$$I=E(f(R))\approx\frac{1}{N}\sum_{i=1}^{N}f(r_i)\triangleq\theta_2, \tag{3.5}$$

θ_2 是数学期望 $E(f(R))$ 的无偏估计量.随机变量 $Y=f(R)$ 的方差为 $Var(Y)=\int_0^1(f(x)-I)^2dx$, 故而

$$Var(\theta_2)=Var(\overline{Y})=\frac{1}{N}Var(Y).$$

可以证明:$Var(\theta_2)\leq Var(\theta_1)$.

事实上,由于有 $Var(\theta_1)=\frac{1}{N}p(1-p)=\frac{1}{N}I(1-I)$,而且有 $Var(\theta_2)=\frac{1}{N}\int_0^1(f(x)-I)^2dx=\frac{1}{N}[\int_0^1 f^2(x)dx-I^2]$, 因此 $Var(\theta_2)-\quad Var(\theta_1)=\frac{1}{N}[\int_0^1 f^2(x)dx-I]$. 因为已假设 $0\leq f(x)\leq 1$,且 $\int_0^1 f(x)\quad dx=I$, 所以 $I\geq\int_0^1 f^2(x)dx$, 可见

$$Var(\theta_2)-Var(\theta_1)\leq 0.$$

算法3.2(平均值估计法)

①产生 $[0,1]$ 区间的均匀随机数 $r_1, r_2, \cdots, r_N$;

②计算 $f(r_i)(i=1,2,\cdots,N)$;

③令 $\theta_2=\frac{1}{N}\sum\limits_{i=1}^{N}f(r_i)$, 则 θ_2 为积分值 I 的近似解.

5.2 对偶变量的应用

首先给出对偶变量的定义。

定义5.2.1 （对偶变量）对于变量X，如果变量$Y=g(X)$与变量X均值和方差分别相同，也就是$E(X)=E(Y)$, $\mathrm{Var}(X)=\mathrm{Var}(Y)$，且$\mathrm{Cov}(X,Y)\leq 0$, 则$Y$ 称为$E(X)$通过$0.5\hat{E}(X+Y)$估计的对偶变量。

1、对偶变量估计$E(X)$的思想

如果$X_1, X_2, \cdots, X_n$为样本数据，则$\mu=E(X)$的估计为$\hat{\mu}_1=n^{-1}\sum_{i=1}^{n}X_i$。对偶变量估计的思想为：假设存在$X$的对偶变量$Y=g(X)$，则$E(X)=0.5E(X+Y)$。此时$E(X)$的估计为

$$\hat{\mu}_2=\frac{1}{2n}\sum_{i=1}^{n}(X_i+Y_i)=\frac{1}{2n}\sum_{i=1}^{n}(X_i+g(X_i)).$$

对估计$\hat{\mu}_1$和$\hat{\mu}_2$, 不难得到$\mathrm{Var}(\hat{\mu}_1) = n^{-1}\mathrm{Var}(X)$和$\mathrm{Var}(\hat{\mu}_2) = (2n)^{-1}\mathrm{Var}(X + Y) = (2n)^{-1}\mathrm{Var}(X + g(X))$. 因此只要$2^{-1}\mathrm{Var}(X + g(X)) \le \mathrm{Var}(X)$, 则$\hat{\mu}_2$的方差没有$\hat{\mu}_1$的方差大。注意到$Y = g(X)$和$X$的方差相同，可得

$$\begin{aligned} & 2^{-1}\mathrm{Var}(X + g(X)) - \mathrm{Var}(X) \\ = \; & 2^{-1}(\mathrm{Var}(X) + \mathrm{Var}(g(X)) + 2\mathrm{Cov}(X, g(X)) - 2\mathrm{Var}(X)) = \mathrm{Cov}(X, g(X)). \end{aligned}$$

只要$\mathrm{Cov}(X, g(X)) \le 0$, 基于对偶变量得到的估计$\hat{\mu}_2$比平均值估计$\hat{\mu}_1$方差小。基于对偶变量$Y$得到的估计$\hat{\mu}_2$称为对偶变量法。

下面的定理说明如果X是$U(0,1)$均匀分布，则很容易找到符合条件的对偶变量$g(X)$.

定理5.2.1 如果X是$U(0,1)$均匀分布, 则$g(X) = 1 - X$和X同分布，且满足$\mathrm{Cov}(X, g(X)) < 0$。

证明： 由随机变量变换法可以求得随机变量$g(X)$的密度函数为

$$f(x) = 1 \cdot \mathbb{I}_{\{0 \le 1-x \le 1\}} = 1 \cdot \mathbb{I}_{\{0 \le x \le 1\}},$$

也就是$U(0,1)$均匀分布的密度函数。直接计算可得

$$\begin{aligned} \mathrm{Cov}(X, g(X)) &= \mathrm{E}[Xg(X)] - \mathrm{E}(X)\mathrm{E}(g(X)) \\ &= \int_0^1 x(1-x)dx - \frac{1}{2^2} \\ &= \frac{1}{2} - \frac{1}{3} - \frac{1}{4} < 0. \end{aligned}$$

模拟验证1，根据$U(0,1$上的独立样本$u_1, u_2, \cdots, u_n$, 采用对偶变量的方法，估计EU^2, 这里U是$(0,1)$ 上的均匀分布。这里$n = 100$.

```
n=100
hat_mu1<-NULL
hat_mu2<-NULL
for (i in 1:1000){
data=runif(n,0,1)
X=data^2
hat_mu1[i]=mean(X)
```

```
Y=(1-data)^2
hat_mu2[i]=0.5*mean(X+Y)
}
c(mean(hat_mu1),mean(hat_mu2),var(hat_mu1),var(hat_mu2))
```

模拟验证2，假设X服从正态分布$N(\mu,\sigma^2)$，基于样本$x_1,x_2,\cdots,x_n$, 采用对偶变量的方法估计$E(X)$. 这里$n=100$, $\mu=5$, $\sigma=2$.

```
library(MASS)
n=100
hat_mu1<-NULL
hat_mu2<-NULL
Mu=c(5,5)
Sigma=matrix(c(2,0.8,0.8,2),ncol=2,nrow=2,byrow=T)
for (i in 1:1000){
Data=mvrnorm(n,Mu,Sigma)
X=Data[,1]
hat_mu1[i]=mean(X)
Y=Data[,2]
hat_mu2[i]=0.5*mean(X+Y);
}
c(mean(hat_mu1),mean(hat_mu2),var(hat_mu1),var(hat_mu2))
```

2、对偶变量估计$Ef(U)$

假设$U_1,U_2,\cdots,U_n$为来自$U(0,1)$的均匀分布样本，则$\mu=Ef(U)$的估计为$\hat\mu_1=(2n)^{-1}\sum_{i=1}^{2n}f(U_i)$。如果$f(\cdot)$是单调函数，则对偶变量$f(1-U)$与$f(U)$同分布且$\mathrm{Cov}(f(1-U),f(U))\le 0$(具体证明见定理5.2.2). 则对偶估计

$$\hat\mu_2=\frac{1}{2n}\sum_{i=1}^{n}(f(U_i)+f(1-U_i))$$

比估计$\hat\mu_1$更好。

定理5.2.2 设f为单调函数，$U \sim U(0,1)$, 则$\mathrm{Cov}(f(U), f(1-U)) \leq 0$。

证明：不妨设f为单调增函数，令$g(x) = -f(1-x)$，则g为单调增函数，现在证明定理等价于证明$\mathrm{Cov}(f(U), g(U)) \geq 0$. 因为$f$和$g$都是单调增函数，则对任意的$x$和$y$有

$$[f(x) - f(y)][g(x) - g(y)] \geq 0.$$

对任何随机变量X和Y, $[f(X) - f(Y)][g(X) - g(Y)] \geq 0$, 取期望得

$$\mathrm{E}\{[f(X) - f(Y)][g(X) - g(Y)]\} \geq 0$$

或者等价于

$$\mathrm{E}[f(X)g(X)] + \mathrm{E}[f(Y)g(Y)] \geq \mathrm{E}[f(X)g(Y)] + \mathrm{E}[f(Y)g(X)]. \tag{5.2.1}$$

(5.2.1)式对于任意的随机变量X, Y以及任意的单调函数$f(x)$和$g(x) = -f(1-x)$均成立。特别地，当随机变量X和Y独立同分布时有

$$\begin{aligned} \mathrm{E}[f(X)g(X)] &= \mathrm{E}[f(Y)g(Y)] \\ \mathrm{E}[f(X)g(Y)] &= \mathrm{E}[f(Y)g(X)] = \mathrm{E}[f(X)]\mathrm{E}[g(Y)], \end{aligned}$$

此时(5.2.1)式化为

$$2\mathrm{E}[f(X)g(X)] \geq 2\mathrm{E}[f(X)]\mathrm{E}[g(Y)] = 2\mathrm{E}[f(X)]\mathrm{E}[g(X)]$$

则$\mathrm{Cov}(f(U), g(U)) = \mathrm{E}[f(U)g(U)] - \mathrm{E}[f(U)]\mathrm{E}[g(U)] \geq 0$, 定理证毕。

为了将定理5.2.2推广到多变量的情形，下面给出一个具有启发性的定理，而将定理5.2.2在多变量情形下的推广作为其重要推论给出。

定理5.2.3 设随机变量$X_1, X_2, \ldots, X_n$相互独立，函数$f(X) = f(X_1, X_2, \ldots, X_n)$和$g(X) = g(X_1, X_2, \ldots, X_n)$关于每个自变量$X_i, i = 1, 2, \cdots, n$单调增，则

$$\mathrm{E}[f(X)g(X)] \geq \mathrm{E}[f(X)]\mathrm{E}[g(X)].$$

证明： 对变量个数n做归纳法。由定理5.2.2的证明过程知, 当$n = 1$时定理成立。假设定理对$n-1$个变量的情况下成立，则对于n个变量的情况，可得

$$\begin{aligned} &\mathrm{E}[f(X)g(X)|X_n = x_n] \\ &= \mathrm{E}[f(X_1, \ldots, X_{n-1}, x_n)g(X_1, \ldots, X_{n-1}, x_n)|X_n = x_n] \\ &= \mathrm{E}[f(X_1, \ldots, X_{n-1}, x_n)g(X_1, \ldots, X_{n-1}, x_n)] \quad \text{根据独立性} \\ &\geq \mathrm{E}[f(X_1, \ldots, X_{n-1}, x_n)]\mathrm{E}[g(X_1, \ldots, X_{n-1}, x_n)] \quad \text{根据归纳假设} \\ &= \mathrm{E}[f(X)|X_n = x_n]\mathrm{E}[g(X)|X_n = x_n] \end{aligned}$$

由于上面的论述对于任意x_n都成立，因此

$$E[f(X)g(X)|X_n] \geq E[f(X)|X_n]E[g(X)|X_n],$$

两端取期望，由重期望公式得

$$\begin{aligned} E[f(X)g(X)] &\geq E\{E[f(X)|X_n]E[g(X)|X_n]\} \\ &\geq E[f(X)]E[g(X)]. \end{aligned}$$

后一个不等式成立是因为$E[f(X)|X_n]$和$E[g(X)|X_n]$都是X_n的单调增函数，根据$n=1$时的结论可得

$$\begin{aligned} E\{E[f(X)|X_n]E[g(X)|X_n]\} &\geq E\{E[f(X)|X_n]\}E\{E[g(X)|X_n]\} \\ &= E[f(X)]E[g(X)]. \end{aligned}$$

故定理在n个变量的情况下也成立，由归纳原理定理证毕。

定理5.2.4 设$h(x_1, x_2, \ldots, x_n)$是关于每个自变量单调的函数，$U_1, U_2, \ldots, U_n$相互独立，则$\mathrm{Cov}(h(U_1, U_2, \ldots, U_n), h(1-U_1, 1-U_2, \ldots, 1-U_n)) \leq 0$。

证明： 不失一般性，由于$U_1, U_2, \ldots, U_n$之间独立同分布，通过重新定义h,可以假定h是其前r个自变量的单调增函数，后$n-r$个自变量的单调减函数。因此，令

$$\begin{aligned} f(x_1, \ldots, x_n) &= h(x_1, \ldots, x_r, 1-x_{r+1}, \ldots, 1-x_n) \\ g(x_1, \ldots, x_n) &= -h(1-x_1, \ldots, 1-x_r, x_{r+1}, \ldots, x_n) \end{aligned}$$

由此可得f和g都是关于每个自变量的单调增函数。根据定理5.2.3 可得

$$\mathrm{Cov}(f(U_1, \ldots, U_n), g(U_1, \ldots, U_n)) \geq 0$$

这等价于

$$\begin{aligned} &\mathrm{Cov}(h(U_1, \ldots, U_r, 1-U_{r+1}, \ldots, 1-U_n,), \\ &\qquad h(1-U_1, \ldots, 1-U_r, U_{r+1}, \ldots, U_n)) \leq 0. \end{aligned} \tag{5.2.2}$$

而随机向量$(h(U_1, \ldots U_n), h(1-U_1, \ldots, 1-U_n))$的联合分布与随机向量

$$(h(U_1, \ldots, U_r, 1-U_{r+1}, \ldots, 1-U_n,), h(1-U_1, \ldots, 1-U_r, U_{r+1}, \ldots, U_n))$$

的联合分布相同，由(5.2.2),定理证毕。

事实上，定理5.2.4的结果可以推广。如果$(X_i, Y_i), i = 1, 2, \ldots, n$独立，且$X_i$和$Y_i$是对偶变量且负相关，$h(x_1, x_2, \ldots, x_n)$关于每个自变量单调不减，则$\mathrm{Cov}(h(X_1, X_2, \ldots, X_n),\ h(Y_1, Y_2, \ldots, Y_n)) \le 0$。对于期望$Eh(X_1, X_2, \ldots, X_n)$，可以引入对偶变量$h(Y_1, Y_2, \ldots, Y_n)$, 相对于平均值估计法，降低估计的方差。例如，要估计$\theta = Eh(X_1, X_2, \ldots, X_n)$, 如果$X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$相互独立，相应的对偶变量可以选为$Y_i = 2\mu_1 - X_i$。因此可以根据对偶变量$h(Y_1, Y_2, \ldots, Y_n)$降低估计$\hat{\theta}$的方差。

引理5.2.1 如果X的分布函数$F(x)$逆变换存在，则对$E(X) = E(F^{-1}(U))$的估计，可以通过引入对偶变量$F^{-1}(1-U)$, 根据$E(X) = 0.5E(F^{-1}(U) + F^{-1}(1-U))$估计降低估计方差。

根据引理5.2.1, 对随机样本$U_1, U_2, \cdots, U_n$, 利用对偶变量对$\mu = E(X)$的估计$\hat{\mu}_2 = (2n)^{-1}\sum_{i=1}^n \{F^{-1}(U_i) + F^{-1}(1-U_i)\}$。$\mu = E(X)$的方差比估计$\hat{\mu}_1 = n^{-1}\sum_{i=1}^n F^{-1}(U_i)$的方差小。

上述所提出的对偶变量估计，X和Y的权重都是0.5，一个很自然的问题是通过$E(X) = E(\alpha X + (1-\alpha)Y)$, $\alpha \in [0,1]$估计$E(X)$, 也就是$\hat{\mu}_\alpha = n^{-1}\sum_{i=1}^n (\alpha X_i + (1-\alpha)Y_i)$。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

$$\hat{I} = \frac{1}{2n}\sum_{i=1}^n e^{U_i} + e^{1-U_i}, \tag{5.2.3}$$

方差为

$$\mathrm{Var}(\hat{I}) = \frac{1}{2n}\mathrm{Var}(e^U) + \mathrm{Cov}(e^U, e^{1-U}) = \frac{1}{n}(-\frac{3}{4}e^2 + \frac{5}{2}e - \frac{5}{4}) \approx \frac{0.00391}{n}. \tag{5.2.4}$$

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$E(X)$，