蒙特卡洛方法

王宁宁

蒙特卡洛方法

很多时候，我们需要计算积分，但是大多数情形下的积分很难给出解析解，我们可以通过蒙特卡洛方法给出数值模拟解。

$$Z_N^{\mathrm{MC}} = \frac{1}{N}\sum_{j=1}^{N} f(X_j),$$

1、蒙特卡洛原理

通常形式的任一个积分都可以表示成某个随机变量期望的形式：

$$\int_0^1 \mathrm{f}(\mathrm{t})\mathrm{dt} = \mathrm{E}(\mathrm{f}(\mathrm{U}))$$

思考：

a）、如果是$\int_a^b f(t)dt$怎么表示成某个随机变量的期望？

b）、任一个概率怎么表示成某个随机变量的期望？

$$P(X \in A) = \mathbb{E}(\mathbb{1}_A(X)) \approx \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_A(X_j)$$

根据（强）大数定律：

$$\mathbb{E}(f(X)) = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} f(X_j)$$

所以理论上只要能够得到 X 的随机数，就可以逼近想要的积分。

$$Z_N^{\mathrm{MC}} = \frac{1}{N}\sum_{j=1}^{N} f(X_j),$$

2、模拟次数

如何根据精度选择模拟的次数？

无偏性：

$$\mathbb{E}\left(Z_N^{\mathrm{MC}}\right)=\mathbb{E}\left(\frac{1}{N}\sum_{j=1}^{N} f(X_j)\right)=\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left(f(X_j)\right)=\mathbb{E}(f(X))$$

由于 X 是独立的：

$$\operatorname{Var}\left(Z_N^{\mathrm{MC}}\right)=\operatorname{Var}\left(\frac{1}{N}\sum_{j=1}^{N} f(X_j)\right)=\frac{1}{N^2}\sum_{j=1}^{N}\operatorname{Var}\left(f(X_j)\right)=\frac{1}{N}\operatorname{Var}(f(X)).$$

$$\operatorname{MSE}\left(Z_N^{\mathrm{MC}}\right)=\frac{\operatorname{Var}(f(X))}{N}\approx\frac{\hat{\sigma}^2}{N},$$

其中

$$\hat{\sigma}^2=\frac{1}{N-1}\sum_{j=1}^{N}\left(f(X_j)-Z_N^{\mathrm{MC}}\right)^2$$

可以这样估计次数：

$$N\geq\frac{\operatorname{Var}(f(X))}{\varepsilon^2}.$$

MSE

$$\begin{aligned}
\operatorname{MSE}(\hat{\theta}) &= \mathbb{E}_\theta\left((\hat{\theta}-\theta)^2\right)\\
&= \mathbb{E}_\theta(\hat{\theta}^2)-2\theta\mathbb{E}_\theta(\hat{\theta})+\theta^2\\
&= \mathbb{E}_\theta(\hat{\theta}^2)-\mathbb{E}_\theta(\hat{\theta})^2+\mathbb{E}_\theta(\hat{\theta})^2-2\theta\mathbb{E}_\theta(\hat{\theta})+\theta^2\\
&= \mathbb{E}_\theta(\hat{\theta}^2)-\mathbb{E}_\theta(\hat{\theta})^2+\left(\mathbb{E}_\theta(\hat{\theta})-\theta\right)^2\\
&= \operatorname{Var}(\hat{\theta})+\operatorname{bias}(\hat{\theta})^2.
\end{aligned}$$

MC 算法的例子

例 1

1、假定 X 来自 Exp(1), 而 Y|X 来自 N(0,X), 现在已知 Y=4，利用 MC 方法求 E(X|Y=4)以及 Var(X|Y=4) （从后验分布里抽取 10000 份样本，计算样本均数和样本方差）。

贝叶斯公式：

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$$

已知：

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi x}} \exp\left(-y^2/2x\right).$$

故：

$$p_Y(y) = \int_0^\infty p_{Y|X}(y|\tilde{x})\, p_X(\tilde{x})\, d\tilde{x} = \int_0^\infty \frac{1}{\sqrt{2\pi \tilde{x}}} \exp\left(-y^2/2\tilde{x} - \tilde{x}\right)\, d\tilde{x}.$$

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)} \\ &= \frac{\frac{1}{\sqrt{2\pi x}} \exp\left(-y^2/2x - x\right)}{\int_0^\infty \frac{1}{\sqrt{2\pi \tilde{x}}} \exp\left(-y^2/2\tilde{x} - \tilde{x}\right)\, d\tilde{x}} \\ &= \frac{1}{Z_f} f(x) \end{aligned}$$

其中：

$$f(x) = \frac{1}{\sqrt{x}} \exp\left(-y^2/2x - x\right)$$

使用筛法对上述 f(x)抽样，选取：

$$g(x) = \exp(-x)$$

以及：

$$c=\frac{1}{|y|}\exp(-1/2)$$

故：

$$\begin{aligned} cg(X)U &\le f(X) \\ \Longleftrightarrow \quad \frac{1}{|y|}\exp(-1/2)\exp(-x)U &\le \frac{1}{\sqrt{x}}\exp\left(-y^2/2x-x\right) \\ \Longleftrightarrow \quad U &\le \frac{|y|}{\sqrt{x}}\exp\left(-y^2/2x+1/2\right). \end{aligned}$$

抽样算法：

(a) Generate $X \sim \text{Exp}(1)$.
(b) Generate $U \sim \mathcal{U}[0,1]$.
(c) Accept X if $U \le \frac{|y|}{\sqrt{x}}\exp\left(-y^2/2x+1/2\right)$.

```
GeneratePosteriorSamples <- function(n, y) {
  res <- c()
  while (length(res) < n) {
      X <- rexp(1)
      U <- runif(1)
      if (U <= abs(y) * exp(-y^2/(2*X) + 0.5) / sqrt(X)) {
        res <- c(res, X)
      }
  }
  return(res)
};
X <- GeneratePosteriorSamples(10000, 4)
mean(X)

## [1] 3.348227

var(X)

## [1] 1.879751
```

MC 算法的精度

例 2

2、假定 X 来自 N(0,1), 利用 MC 算法估计$E(\sin(x)^2)$，具体如下：

1）、抽样 1000000 次，算出均值。

2）、每次抽样 1000 次，算出一个均值，重复 10000 次，画出这 10000 个均值的直方图。

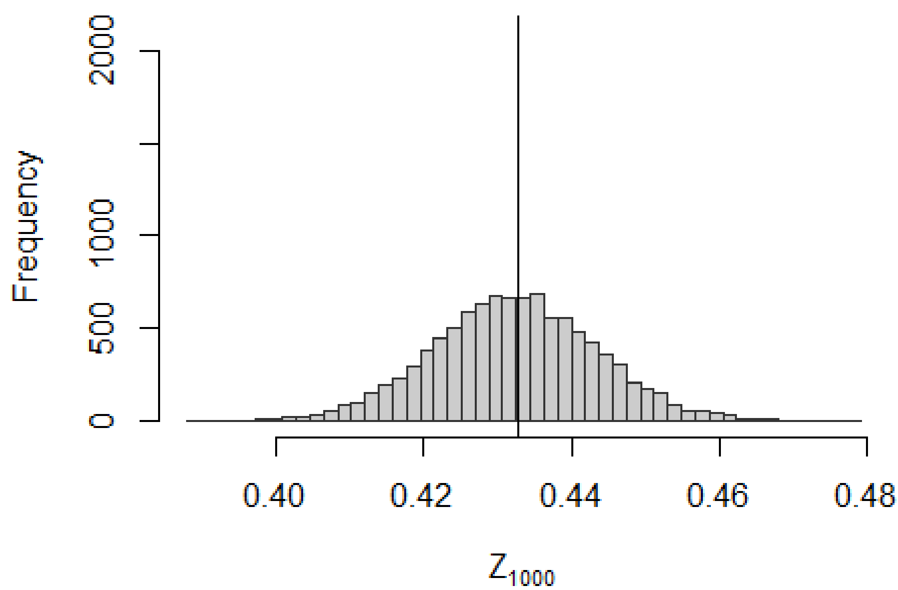
3）、每次抽样 10000 次，算出一个均值，重复 10000 次，画出这 10000 个均值的直方图。

```
GetMCEstimate <- function(N) {
  X <- rnorm(N)
  return(mean(sin(X)^2))
}
GetMCEstimate(1000000)
```

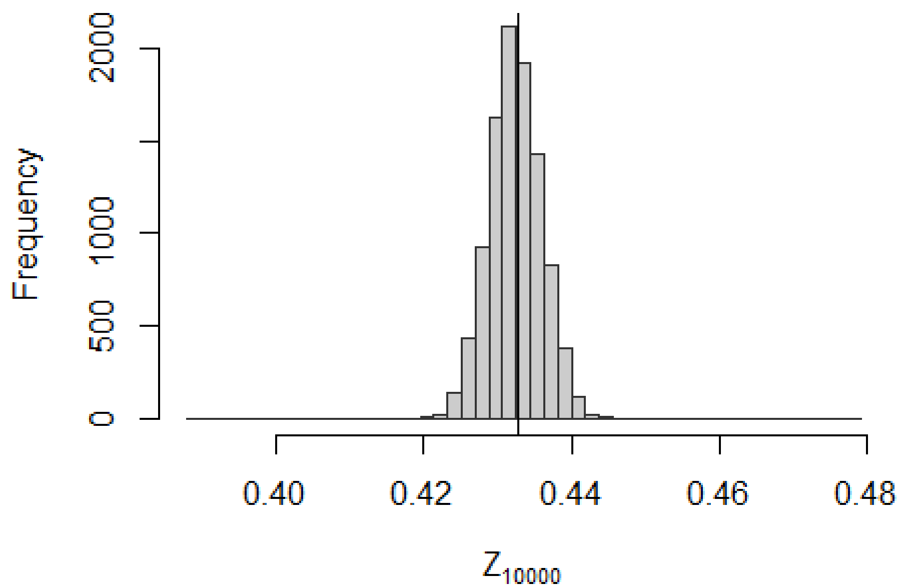
```
## [1] 0.4322745
```

```
estimates <- replicate(10000, GetMCEstimate(1000))
range <- c(min(estimates), max(estimates))
hist(estimates, breaks=seq(range[1], range[2], length.out=50),xlab=expr
ession(Z[1000]), xlim=range, ylim=c(0,2100),main=NULL, col="gray80", bo
rder="gray20")


good.estimate <- GetMCEstimate(1000000)
abline(v=good.estimate)
```



```
estimates2 <- replicate(10000, GetMCEstimate(10000))
hist(estimates2, breaks=seq(range[1], range[2], length.out=50),xlab=exp
ression(Z[10000]), xlim=range, ylim=c(0,2100),main=NULL, col="gray80",
border="gray20")
abline(v=good.estimate)
```



例 3

3、已知假定 X 来自 N(0,1)，利用 MC 算法估计E(cos(x))，如果需要 3 位有效数字，至少需要抽样多少次？（确定数量级即可）

$$\mathrm{MSE}\left(Z_N^{\mathrm{MC}}\right) \leq \frac{\mathrm{Var}\left(\cos(X)\right)}{N} \leq \frac{1}{N}.$$

根据：

$$N \geq \frac{\mathrm{Var}(f(X))}{\varepsilon^2}.$$

需要10^6抽样

```r
N <- 1e6
X <- rnorm(N)
mean(cos(X))
```

```
## [1] 0.6065775
```

```r
var(cos(X))
```

```
## [1] 0.199669
```
