MC 方法的改进——方差压缩技术(重要性抽样)

王宁宁

MC 方法 回顾

$$Z_N^{MC} = E(f(X)) = \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

MC 的精度：

$$\mathrm{MSE}(Z_N^{MC}) = \frac{\mathrm{var}(f(X))}{N} \equiv \mathcal{O}\left(\frac{1}{N}\right)$$

而：

$$\mathrm{RMSE}(Z_N^{MC}) = \frac{\mathrm{stdev}(f(X))}{\sqrt{N}} \equiv \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

这意味着精度如果想增加一位有效数字，样本量需要增加 100 倍。

下面考虑一种方差压缩技术：

设$X \sim \varphi(x)$，$Y \sim \psi(x)$，考虑$E(f(X))$的抽样估计

$$\begin{aligned} E(f(X)) = & \int f(x)\varphi(x)\,dx \\ = & \int f(x)\frac{\varphi(x)}{\psi(x)}\psi(x)\,dx \\ = & E\left(f(Y)\frac{\varphi(Y)}{\psi(Y)}\right) \\ \equiv & Z_N^{IS} \end{aligned}$$

算法如下：

$s \leftarrow 0$
for $j = 1, 2, \ldots, N$ **do**
    generate $Y_j \sim \psi$
    $s \leftarrow s + f(Y_j)\varphi(Y_j)/\psi(Y_j)$
end for
return s/N

上述抽样方法称为**重要性抽样**。

重要性抽样的性质

$$\mathrm{bias}(Z_N^{IS}) = 0$$

$$\begin{aligned}\mathrm{MSE}(Z_N^{IS}) = & \frac{1}{N}\mathrm{Var}\left((f(X))\frac{\varphi(Y)}{\psi(Y)}\right) \\ = & \frac{1}{N}\left(\mathrm{Var}(f(X)) - E\left(f(X)^2(1-\frac{\varphi(X)}{\psi(X)})\right)\right)\end{aligned}$$

ψ 的选取：

$$c_\psi = E(f(X)^2(1-\frac{\varphi(X)}{\psi(X)})) > 0$$

- 上述 c_ψ 尽可能得大。
- $\mathrm{Var}\left((f(Y))\frac{\varphi(Y)}{\psi(Y)}\right)$ 尽可能的小。

重要性抽样的样本容量选择

$$\mathrm{MSE}(Z_N^{IS}) = \frac{1}{N}\mathrm{Var}\left((f(Y))\frac{\varphi(Y)}{\psi(Y)}\right) \approx \frac{\hat{\sigma}^2}{N}$$

例子

设$X \sim N(0,1)$且$A = [3,4]$，现在要使用重要性抽样估计$P(X \in A)$，使用以下的辅助密度函数:

a、$Y \sim N(1,1)$

b、$Y \sim N(2,1)$

c、$Y \sim N(3.5,1)$

d、$Y \sim \mathrm{Exp}(1) + 3$

1、设计程序完成上述四种重要性抽样，各抽样 10000 次，比较四种方法的抽样误差。想一下哪一种抽样的精度最好？

2、假如想要$P(X \in A)$的精度达到 0.01，每种方法各需要抽样多少次？