# 第 13 题作业报告

PB18000341 范玥瑶

## A. 作业题目

用 Metropolis-Hasting 抽样方法计算积分：$I = \int_0^\infty (x-\alpha\beta)^2 f(x)dx = \alpha\beta^2, f(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp(-x/\beta)$.

设积分的权重函数为：$p(x) = f(x)$，给定参数$\alpha$，$\beta$，并用不同的$\gamma$值，分别计算积分，讨论计算精度和效率。

## B. 算法及主要公式

### B.1. $p(x) = f(x)$的 Metropolis-Hasting 抽样

取与初态无关的对称建议分布 T 如下：

$$T_{ij} = T(x \to x') = T(x') = 0.5\exp\{-x'/\gamma\}$$

设$x_0 = 1$，生成$R, Q \sim U(0,1)$，抽样$x' = -\gamma lnR \in (0, +\infty)$，

$$r0 = \frac{p_j T_{ji}}{p_i T_{ij}} = \frac{p(x')T(x_i)}{p(x_i)T(x')} = \frac{f(x')T(x_i)}{f(x_i)T(x')} = \left(\frac{x'}{x_i}\right)^{\alpha-1}\exp(-\frac{x'-x_i}{\beta})\exp(-\frac{x_i-x'}{\gamma})$$

则

$$x_{i+1} = \begin{cases} x', if\ R < \min(1, r0) \\ x_i, if\ R > \min(1, r0) \end{cases}$$

重复上述步骤(n+N)次，除去热化步骤 i=1 到 n 得到 N 个 $x_i$，积分值

$$I = \frac{1}{N}\sum_{i=n+1}^{n+N}(x_i - \alpha\beta)^2$$

算法实现如下：设置全局变量α，β，线性同余法生成的随机数 I，热化步骤数 n，抽样步数 N，输入 n，N，γ，调用子函数 MH 进行计算。

子函数 MH 算法如下：取 $x_0$=1，将积分值变量 Integrate 和有效选取计数 T 置为 0。执行以下步骤（n+N）次：用 Schrage 方法生成（0，1）上均匀分布随机数 R,Q，依照上述方法得到 x'，r0，判断是否满足$Q < \min(1, r0)$即$Q < 1\&\&Q < \mathrm{r0}$。若满足则选取 $x_{i+1}$=x'，反之放弃 $x_{i+1}$=$x_i$。前 n 次为热化步骤，Integrate 不进行操作；后 N 次对$(x_i - \alpha\beta)^2$进行累加存储在 Integrate，当该步被接受时 T 自增。循环结束后得到积分值（程序中由于线性同余法生成的随机数有变量名 I，此处的 I 还是赋给 Integrate）

$$I = \frac{1}{N}Intergrate$$

对应相对误差

$$\delta = \frac{Integrate - \alpha\beta^2}{\alpha\beta^2} * 100\%$$

以及效率η

$$\eta = \frac{T}{N} * 100\%$$

### B.2. 不同的γ值分别计算积分，讨论计算精度和效率

利用 B.1.的算法，输入步骤中改为输入γ的范围，将($\gamma_{min}$,$\gamma_{max}$)进行 100 等分得到 101 个γ值，将各γ下运行结果 Integrate，δ和η输入到 MH.txt 保存。通过选取不同范围的γ对结果作图可以得到合适的γ范围。效率过低的积分不具有代表性，且浪费计算资源；误差过大说明不适用。抽样效率在 50%左右较合理。

## C. 计算结果及具体分析、讨论

### C.1.预实验选取参数

取α=1，β=2，标准结果$\alpha\beta^2=4$。取较小步数试探性运行 B.2.所得程序，种子值较小时要经过比较多步才能行走，可以通过增大热化步数实现，这里节省运行时间取 seed=1069865427。发现大概取γ>0.167 时可以在 15 步之内走出 x0。取γ=0.19，n=100，N=1E3~1E9，运行程序 B.1.得到结果如表 1.从表上得到 N 比较大时积分误差和效率分别随取样数缓慢减小和增大，综合考虑计算资源、误差δ、效率η，不妨取 N=1E6 进行计算。

| N | 积分结果 I | 误差δ | 效率η |
|---|---|---|---|
| 1E3 | 1.468074 | -63.2982% | 23.2000% |
| 1E4 | 0.267489 | -93.3128% | 2.3200% |
| 1E5 | 1.175768 | -70.6058% | 18.5840% |
| 1E6 | 1.424608 | -64.3848% | 22.5238% |
| 1E7 | 1.420514 | -64.4872% | 22.3107% |
| 1E8 | 1.426417 | -64.3396% | 22.1705% |
| 1E9 | 1.453757 | -63.6561% | 21.0698% |

表 1 不同取样步数下积分结果、误差、效率表

### C.2. 不同γ值下计算精度和效率的讨论

取 seed=1069865427，n=100，N=1E6，α=1，β=2，γmin=0.17,γmax=5，进行计算得到结果 Metropolis-Hasting Sampling.txt。导入 Origin 作图得到积分值随γ变化和与精确值对比图如图 1，积分误差、抽样效率随γ变化图（0.167<γ<5）如图 2，3。

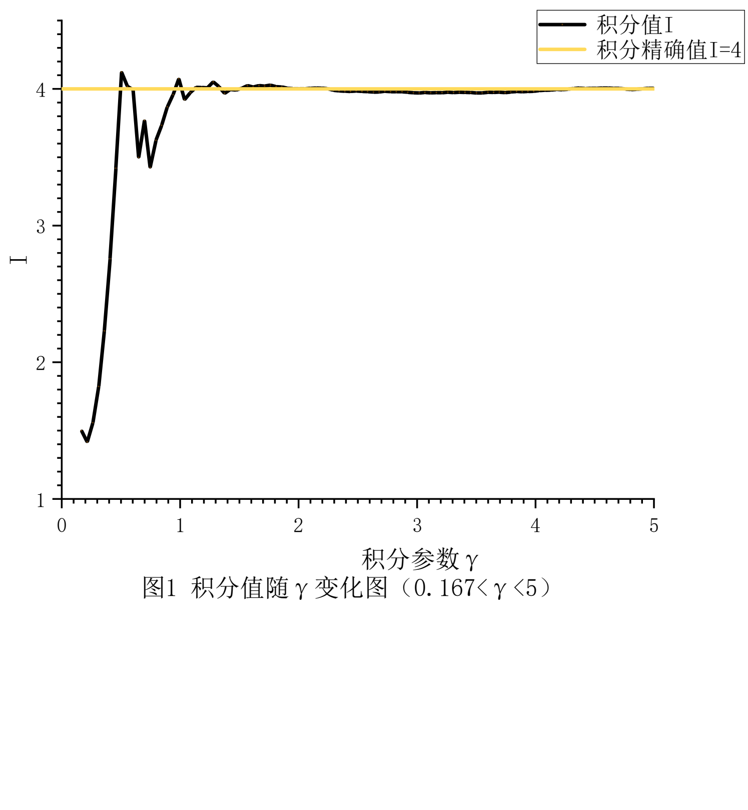


图1 积分值随γ变化图（0.167<γ<5）

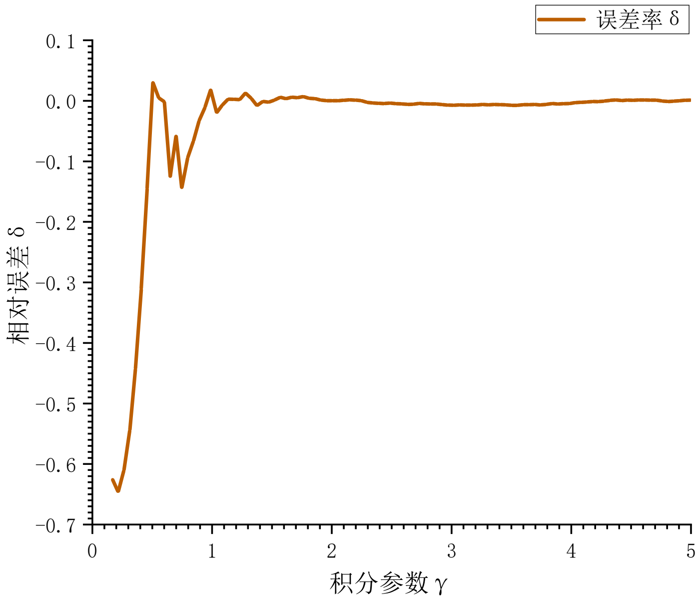

图2 相对误差随 γ 变化图（0.167< γ <5）

由图 1，2 可得γ较小时误差较大且有波动，当γ>1.52024(第 29 个γ值)时积分值在精确值 I=4 附近波动，且幅度较小，误差率<1%。可以通过较大的γ值得到精确度较高的积分也验证了 M-H 方法的有效性。

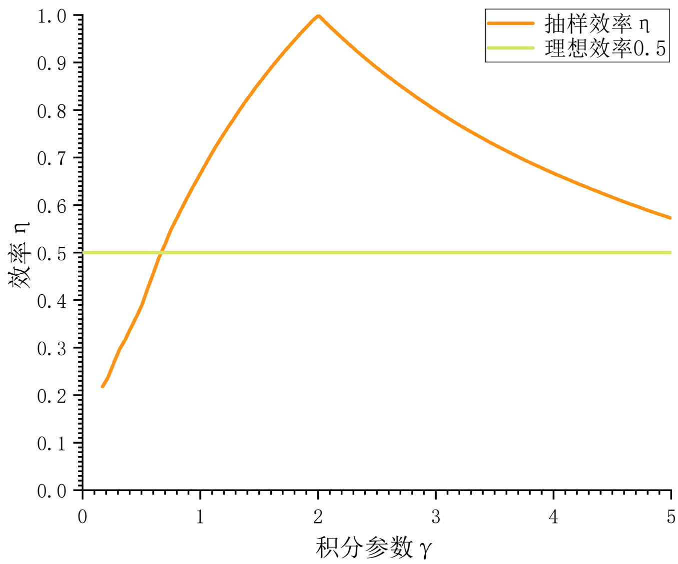

图3 抽样效率随 γ 变化图（0.167< γ <5）

由图 3，抽样效率η随γ变化而先增后减，在γ=2.00354 时取到最高效率 99.9146%，此时γ在γ>1.52024 的，误差<1%的范围内；在γ=0.7071 时取到理想效率 50%，不在γ>1.52024 的，误差<1%的范围内；根据图像趋势推测γ>5 还有一个γ对应η=50%，因此又增加了γ=5~20 的试验。补充实验的数据和第一次试验数据合起来作图如图 4，出现推测中的第二个交点γ=7.72727273，在误差<1%的范围内。这是比较理想的积分计算条件。

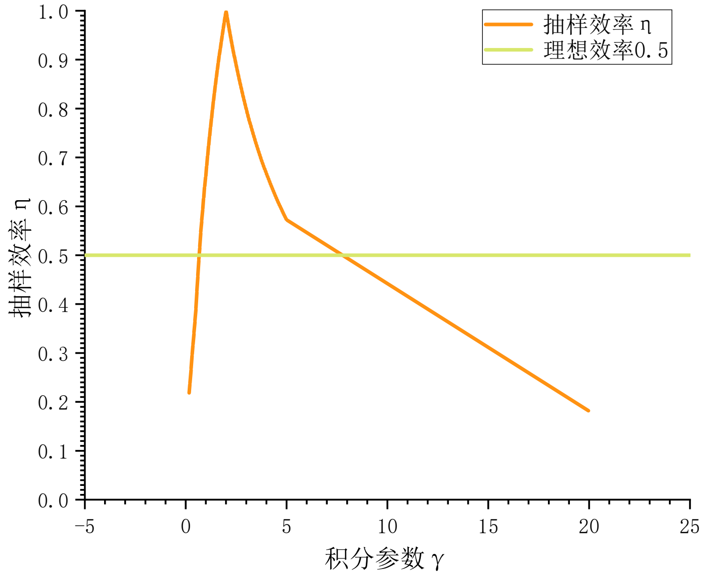


图4 抽样效率随γ变化图（0.167<γ<20）

## D. 总结

本次作业中，学生使用了 Metropolis-Hasting 方法，计算了积分：$I=\int_0^{\infty}(x-\alpha\beta)^2 f(x)dx=\alpha\beta^2$，$f(x)=\frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-x/\beta\right)$。计算中设积分的权重函数为 $p(x)=f(x)=\frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-x/\beta\right)$，给定参数$\alpha=1$，$\beta=2$，并在 0.17 到 5 间等距取了 101 个不同的$\gamma$值，对应取了建议分布$T(x')=0.5\exp\left\{-x'/\gamma\right\}$计算积分，讨论了计算精度和效率。得出γ较小时误差较大且有波动，当γ>1.52024 时积分值在精确值 I=4 附近小幅度波动，且积分结果比较精确，误差在 1%以内，验证了 M-H 方法的有效性。抽样效率η随γ变化而先增后减，在γ=2.00354 时取到最高效率 99.9146%，此时γ在γ>1.52024，误差<1%的范围内。η在γ=0.7071 和γ=7.72727273 时取到理想的接受概率 50%，其中后者在γ>1.52024 的范围内，是比较理想的抽样参数。