

计算物理第13题

PB18000039 徐祺云

一 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 α, β , 并用不同的 γ 值, 分别计算积分, 讨论计算精度和效率。

二 算法及主要公式

用Metropolis-Hasting抽样方法, 选定参数为 $\alpha = 1.5, \beta = 2$ 。

积分的权重函数选为 $p(x) = f(x)$, 即抽样 x 满足分布为 $p(x)$ 作为积分的权重函数。假设步进几率 T 与初态无关, 即

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

选取 $x_0 = 1$, 考虑已经产生了 x_1, x_2, \dots, x_i 个抽样点, 下面考虑如何产生下一个抽样点 x_{i+1} 。构造一个试探解 $x_t = -\gamma \ln R$, R 是 $[0, 1]$ 上的一个随机数, 该点是否被选取取决于比值 $r = \frac{p_t T_{ti}}{p_i T_{it}}$:

$$r = \left(\frac{x_t}{x_i}\right)^{\alpha-1} \exp\{-(x_t - x_i)/\beta\} \exp\{(x_t - x_i)/\gamma\}$$

再在 $[0, 1]$ 上取一个随机数 R , 如果 $R < \min(1, r)$, 则 x_{i+1} 取 x_t , 否

则 x_{i+1} 取 x_i 。如此操作可以得到一个Markov链，由此得到积分结果：

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

可以定义接受效率为 x_{n+1} 取新构造量 x_t 的次数与整个链长 N 的比值，显然效率的高低与 γ 有关，因此可以改变 γ 的值计算积分误差与效率。

三 计算结果与分析

取初值为 $x_0 = 1$ ，链长 $N = 10^6$ ，初始值种子值为1950975620，积分真实值为 $\alpha\beta^2=6$ 。为了先粗略地看一下结果分布，这里取 γ 从0到100步进，步长为5，得到结果如下(数据在`data1.txt`)：

表 1: 不同 γ 下的积分精度与效率

γ	积分值 I	相对误差	效率
2	5.989547	0.174210%	0.707025
4	6.041244	0.687400%	0.803251
6	6.004000	0.066659%	0.649702
8	5.996682	0.055307%	0.536391
10	5.982997	0.283376%	0.454833
12	5.965183	0.580290%	0.394081
14	5.960374	0.660436%	0.347945
16	5.954828	0.752859%	0.310888
18	5.953996	0.766738%	0.281193
20	5.971180	0.480341%	0.256536
22	5.970393	0.493448%	0.236022
24	5.997801	0.036644%	0.218471
26	6.023720	0.395340%	0.203642
28	6.011183	0.186388%	0.190147
30	6.017990	0.299828%	0.178545
32	6.028648	0.477468%	0.168431
34	6.012565	0.209422%	0.159120
36	5.986189	0.230190%	0.150876
38	5.999789	0.003518%	0.143405
40	6.009128	0.152129%	0.136665

粗略地看，随着 γ 的增大，积分值在真实值6附近波动，相对误差

在1%内，效率先升高后降低，可能存在一个峰值，为此可以细化参数取值，这里再取 γ 从 0 到 10 步进，步长为0.1，具体数据结果见文件 *data2.txt*，这里可以得到相关绘图：

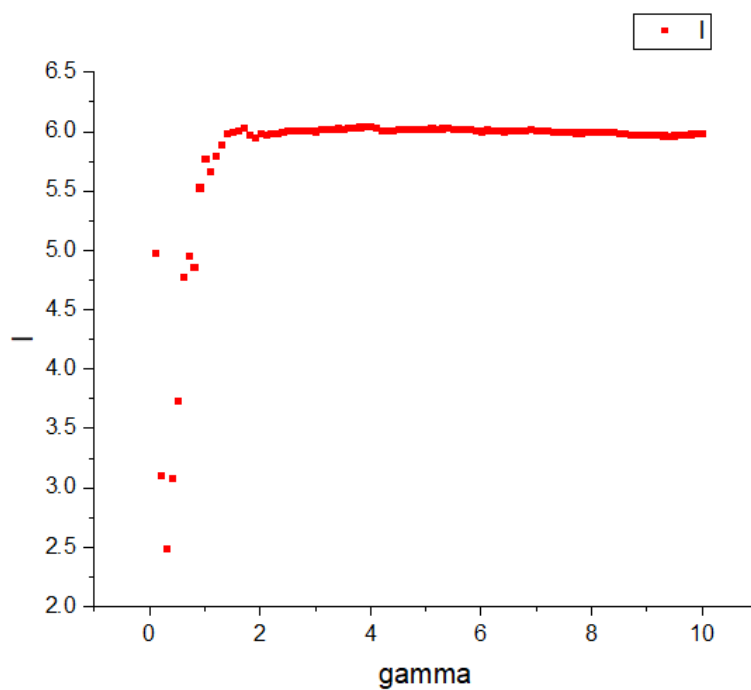


图 1: 不同 γ 值下的积分结果

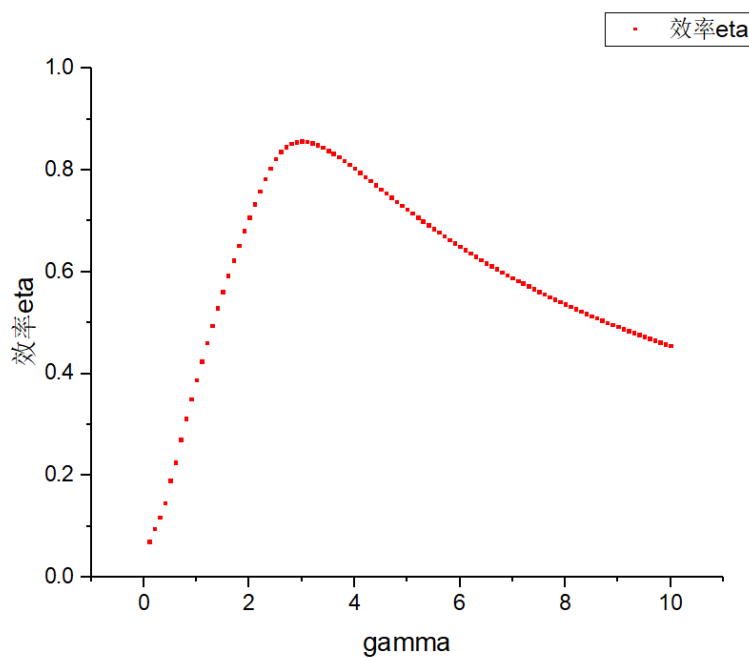


图 2: 不同 γ 值下的效率

可见，当 γ 较小时(约 < 1)，积分结果误差较大，有明显大幅波动，且积分效率很低。随着 γ 的增大，积分结果趋于稳定，在理论值 $\alpha\beta^2$ 附近上下波动，而且约在 $\gamma = \alpha\beta$ 处效率达到最大峰值。随着 γ 的继续增大，积分结果在理论值 $\alpha\beta^2$ 附近的波动略有些许增大，但效率很快降低。

为了更好地说明这一现象，可以取特定的 γ 值观察生成数据的直方图。这里分别取 γ 为 0.3, 3, 300，通过`mian.c`中`x_data_output`函数实现输出，结果如下(N取 10^5 ,初始种子值取1950975620)：

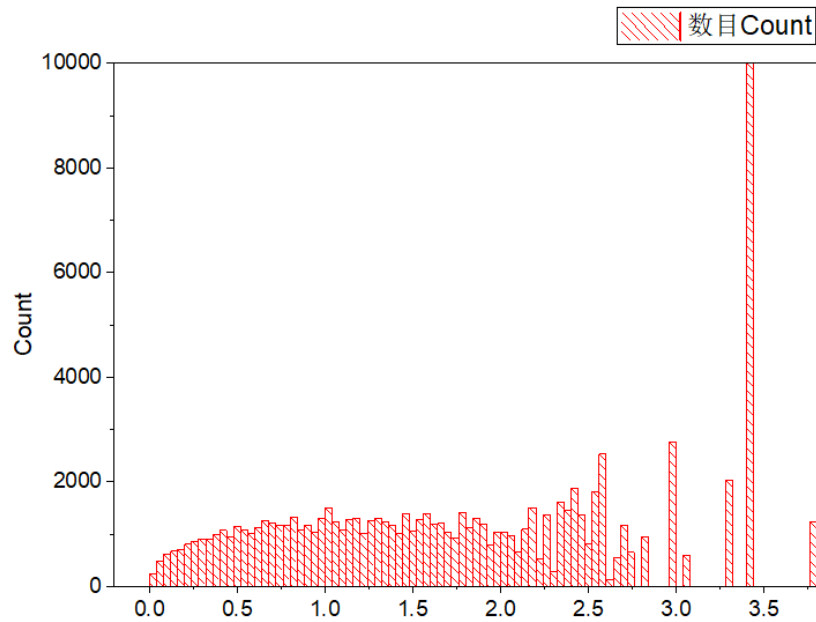


图 3: $\gamma=0.3$ 下的抽样点分布

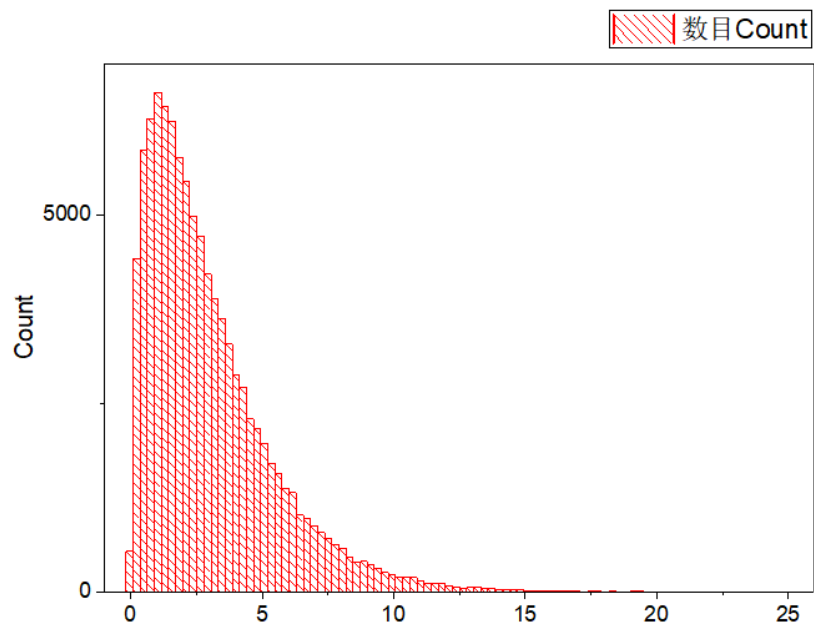


图 4: $\gamma=3$ 下的抽样点分布

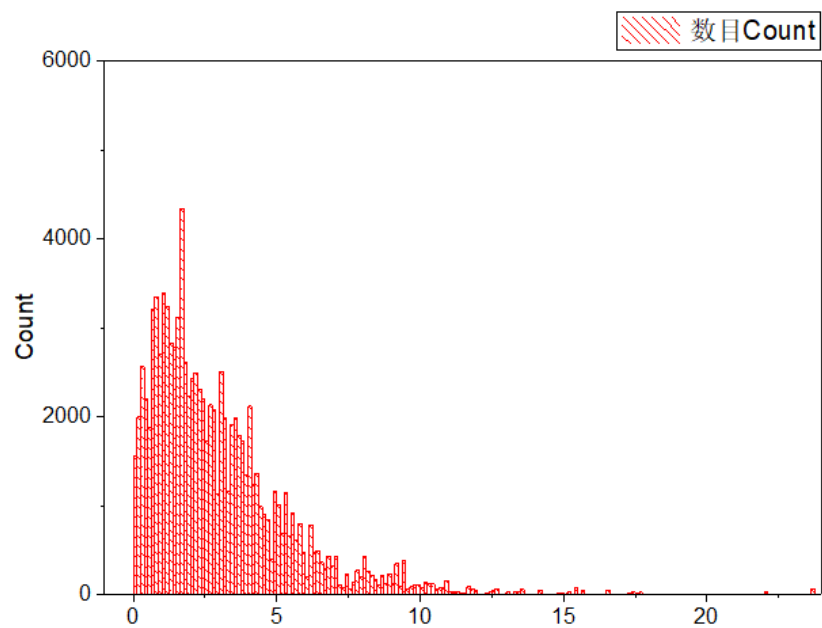


图 5: $\gamma=300$ 下的抽样点分布

理论上我们想得到的是满足 $gamma$ 分布的抽样点，可见参数取 $\gamma = 3$ 时得到的抽样点良好，服从 $gamma$ 分布曲线。而 γ 较小时数据点的强关联性大，分布与预期差别大，抽样结果自然不好； γ 较大时虽然仍保持有 $gamma$ 分布的特征，但与 $\gamma = \alpha\beta$ 时图样相比抽样结果差一些。

四 结论

本次作业用Metropolis-Hasting抽样方法计算了一个 $p(x)$ 为 $gamma$ 分布的积分，理解了MCMC过程并动手实践计算了积分结果。

本次作业先对不同 γ 取值下计算了积分的精度与效率，发现随着 γ 的增大，积分先大幅波动后趋于稳定，在理论值 $\alpha\beta^2$ 附近上下小幅波动，而效率先上升后下降，峰值约在 $\gamma = \alpha\beta$ 处取得。最后分别查看了 $\gamma = 0.3/3/300$ 时的数据点与伽马分布的吻合情况，与预期相符，说明结果良好。