

Report 13

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: $p(x) = f(x)$ 和 $p(x) = (x - \alpha\beta)^2 f(x)$

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

算法及公式

A. 权重 $p(x) = f(x)$

设 $T(x \rightarrow x') = T(x') = \frac{1}{\gamma} e^{-x/\gamma}$

所得简单抽样

$$F(x') = \int_0^{\infty} T(u) du = 1 - e^{-x'/\gamma}$$

抽样 $x' = -\gamma \ln R$

$$r = \frac{p_j T_{ji}}{p_i T_{ij}} = \frac{T(x') \cdot f(x)}{T(x) \cdot f(x')} = \left(\frac{x'}{x}\right)^{\alpha-1} e^{(x'-x)(\frac{1}{\gamma}-\frac{1}{\beta})}$$

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

$$\text{去均值化过程 } I = \frac{1}{N-m} \sum_{i=m}^N (x_i - \alpha\beta)^2$$

B. 权重 $p(x) = (x - \alpha\beta)^2 f(x)$

抽样方法类似 A 中

$$r = \left(\frac{x'}{x}\right)^{\alpha-1} e^{(x'-x)(\frac{1}{\gamma}-\frac{1}{\beta})} \cdot \left(\frac{x'-\alpha\beta}{x-\alpha\beta}\right)^2$$

所得分布为归一化后 $g(x) = \frac{p(x)}{C}$

$$I = \int_0^{\infty} g(x) \cdot \frac{p(x)}{g(x)} dx = \frac{1}{N-m} \sum_{i=m}^N \frac{p(x_i)}{g(x_i)}$$

$g(x)$ 由 x 的样本分布离散估计

结果及讨论

实验中取 $\alpha=3$, $\beta=2$

- A的结果:

gama为2时,误差为0.931791352431528,效率为0.25153251532515325

gama为4时,误差为0.006079842636086852,效率为0.5628756287562876

gama为6时,误差为0.0034340305475062394,效率为0.6375663756637566

gama为8时,误差为0.009723314666597815,效率为0.6097860978609786

- B的结果：

有效抽样点数：94379

gama为0.5时,误差为6.8339755229682195,效率为0.09306093060930609

有效抽样点数：84725

gama为1时,误差为6.644978084262902,效率为0.16950169501695017

有效抽样点数：79242

gama为2时,误差为1.3849481002861292,效率为0.1993719937199372

有效抽样点数：96668

gama为4时,误差为1.0232615971510128,效率为0.34052340523405233

有效抽样点数：97888

gama为6时,误差为0.43116737865270593,效率为0.47403474034740345

有效抽样点数：98016

gama为8时,误差为0.32413900245912686,效率为0.5407954079540795

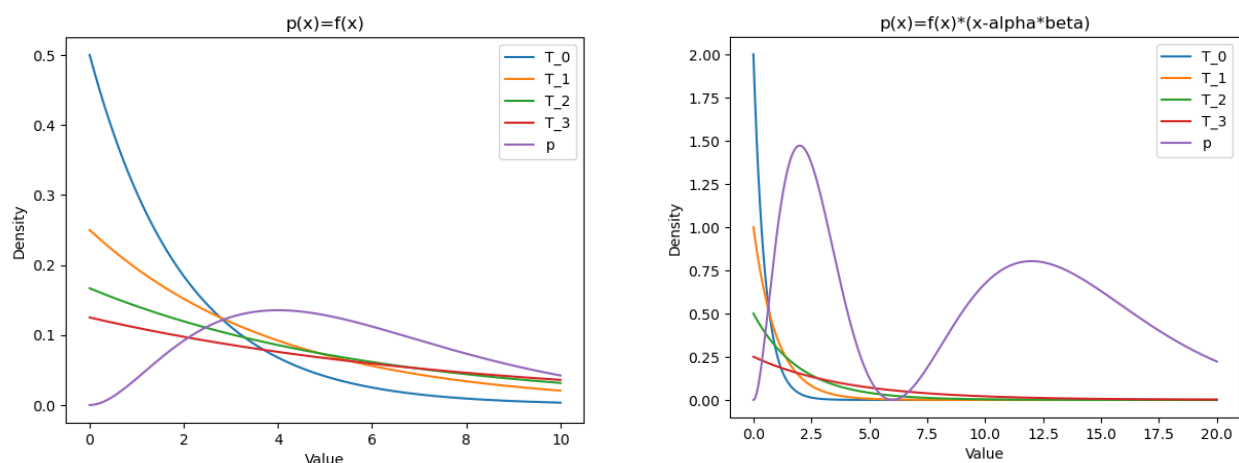
有效抽样点数：98098

gama为10时,误差为0.3055004209647407,效率为0.5659656596565966

- 分析：

理论上来说，当T与p有相接近的分布时，抽样效率会比较高。

对于A，B两种算法，分别取gama为[2, 4, 6, 8] 和[0.5, 1, 2, 4] 看抽样函数的分布情况



由实验可得，A中gama取6时，获得了最高的效率和精度。与理论相符，此时的 $T(x)$ 与 $p(x)$ 较为接近。

B中gama取10时，精度和效率更高。与A相比，B的 $p(x)$ 分布更偏向左侧，所以需要更大的gama。

综上，一个合适的gama值，应该是使得 $T(x)$ 与 $p(x)$ 分布最接近的。

总结

本次实验利用 Metropolis-Hasting 方法，采用了两种分布来计算积分值。并通过改变采样参数的值，讨论了 T 与 p 的分布关系，对采样效率有很大的影响。