

蒙特卡罗方法

续本达

清华大学 工程物理系

2022-10-28

- ① 引言
- ② 复习与概念引入
- ③ 舍选法
- ④ 马尔可夫链法
- ⑤ 总结

蒙特卡罗方法

续本达

引言

复习与概念
引入

舍选法

马尔可夫链法

总结

引言

《马兰花开》两弹的研发



《马兰花开》剧照、脚本

钱学森：咱们搞科学的，容不得半点马虎啊。

....

理论组：老邓，第八次计算结果与第九次完全吻合！

邓稼先：中国的第一颗原子弹，就从这里开始吧！

- 九次计算耗时半年，最结果一致
- 计算尺、手摇计算器、电子管计算机用于计算和仿真

《马兰花开》两弹的研发



《马兰花开》剧照、脚本

钱学森：咱们搞科学的，容不得半点马虎啊。

....

理论组：老邓，第八次计算结果与第九次完全吻合！

邓稼先：中国的第一颗原子弹，就从这里开始吧！

- 九次计算耗时半年，最结果一致
- 计算尺、手摇计算器、电子管计算机用于计算和仿真

问题

- 核物理中的“仿真”是如何实现的？ 蒙特卡罗方法

蒙特卡罗方法

续本达

引言

复习与概念
引入

舍选法

马尔可夫链法

总结

复习与概念引入

算术平均值依概率收敛于数学期望

设 $X_1, X_2, \dots, X_n$ 是独立同分布的随机变量序列, 那么

$$\underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i)}_{\text{算术平均值}} \xrightarrow[n \rightarrow \infty]{P} \underbrace{\int g(x) f_X(x) dx}_{g(X) \text{ 的数学期望}}$$

$f_X(x)$ 是 X_i 的概率密度函数, $g(x)$ 是 x 的任意函数。

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow[n \rightarrow \infty]{P} \int g(x) f_X(x) dx \quad \text{做个试验}$$

取

$$\left. \begin{array}{l} g(x) = \sin x, \quad x \in [0, \pi] \\ f_X(x) = \frac{1}{\pi}, \quad \text{均匀分布} \end{array} \right\} \implies \int_0^\pi g(x) f_X(x) dx = \frac{2}{\pi} = 0.64$$

生成 $[0, 1]$ 上均匀分布的随机数 X_i , 计算 $\frac{1}{n} \sum_{i=1}^n g(X_i)$

- 算术平均值依概率收敛于数学期望

$$\frac{1}{n} \sum_{i=1}^n \sin X_i \xrightarrow[n \rightarrow \infty]{P} \int_0^{\pi} \sin x \underbrace{f_X(x)}_{1/\pi} dx = \frac{2}{\pi}$$

- 反向应用：能否把待求量构造成 数学期望，使用 算术平均 近似？

$$\pi \approx \frac{2n}{\sum_{i=1}^n \sin X_i}$$

定义 (蒙特卡罗方法)

使用 **随机数** 来解决 **计算问题** 方法的统称。

$$\pi \approx \frac{2n}{\sum_{i=1}^n \sin X_i}, X_i \in [0, \pi] \text{均匀}$$

- 估算 π 需要事先知道 $\sin(\cdot)$ 函数的计算方法

$$\pi \approx \frac{2n}{\sum_{i=1}^n \sin X_i}, X_i \in [0, \pi] \text{均匀}$$

- 估算 π 需要事先知道 $\sin(\cdot)$ 函数的计算方法



$$\pi \approx \frac{2n}{\sum_{i=1}^n \sin X_i}, X_i \in [0, \pi] \text{均匀}$$

- 估算 π 需要事先知道 $\sin(\cdot)$ 函数的计算方法



蒙特卡罗方法

续本达

引言

复习与概念
引入

舍选法

马尔可夫链法

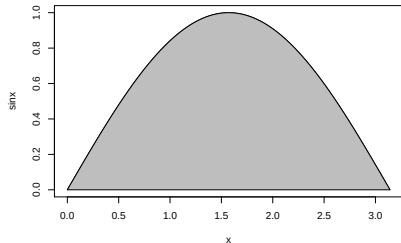
总结

舍选法

$$\pi \approx \frac{2n}{\sum_{i=1}^n \sin X_i}, X_i \in [0, \pi] \text{均匀}$$

- 计算 $\sin X_i$ 不易，如何简化问题？

- $\int_0^{\pi} \sin x dx$ 正是 $\sin x$ 与 x 轴围成的面积



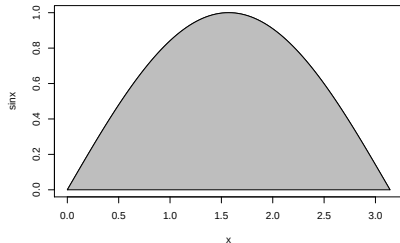
$$\pi \approx \frac{2n}{\sum_{i=1}^n \sin X_i}, X_i \in [0, \pi] \text{ 均匀}$$

- 计算 $\sin X_i$ 不易，如何简化问题？

- $\int_0^\pi \sin x dx$ 正是 $\sin x$ 与 x 轴围成的面积
- 每次均匀生成一对随机数 (X_i, Y_i)

$$X_i \in [0, \pi], Y_i \in [0, 1]$$

- 定义函数 $h(\cdot, \cdot)$ ，如果它落在阴影中，则记 $h(X_i, Y_i) = 1$ ，否则记 $h(X_i, Y_i) = 0$ 。



$$\frac{1}{n} \sum_{i=1}^n h(X_i, Y_i) \approx \frac{2}{\pi}$$

什么叫做“如果 (X_i, Y_i) 落在阴影中”？

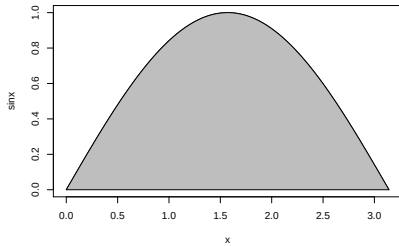
引言

复习与概念
引入

舍选法

马尔可夫链法

总结



什么叫做“如果 (X_i, Y_i) 落在阴影中”？

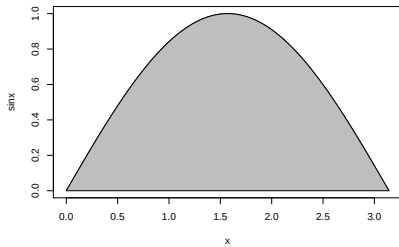
引言

复习与概念
引入

舍选法

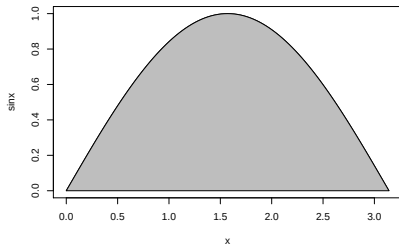
马尔可夫链法

总结



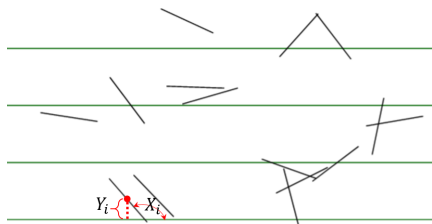
$$h(X_i, Y_i) = \begin{cases} 1 & \sin X_i \leq Y_i \\ 0 & \sin X_i > Y_i \end{cases}$$

还需要算 $\sin$ 函数吗？

什么叫做“如果 (X_i, Y_i) 落在阴影中”？

$$h(X_i, Y_i) = \begin{cases} 1 & \sin X_i \leq Y_i \\ 0 & \sin X_i > Y_i \end{cases}$$

还需要算 $\sin$ 函数吗？



- X_i 为针与平行线的夹角
- Y_i 为针的中点与最近平行线的距离
- 针长与平行线间隔相等

那么

$$\begin{cases} \sin X_i \leq Y_i & \Leftrightarrow \text{针与直线相交} \\ \sin X_i > Y_i & \Leftrightarrow \text{两者不相交} \end{cases}$$

蒲丰（1707–1788 法国）投针方案 $\pi \approx \frac{2n}{\sum_{i=1}^n h_i}$

引言

复习与概念
引入

舍选法

马尔可夫链法

总结

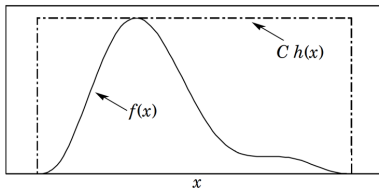
试验者	时间	针长	投掷次数	相交次数	π 的近似值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218	3.1554
De Morgan	1860	1.0	600	382	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

- 对比：祖冲之（429-500，南北朝），割圆术，3.1415927

思考

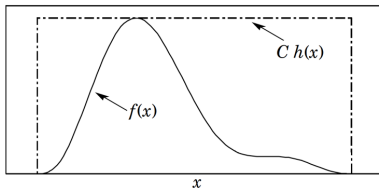
- 投回形针，效果如何？

1950 年代 von Neumann 在领导团队计算核弹引爆机制，在最早一代计算机 (MANIAC) 上运行



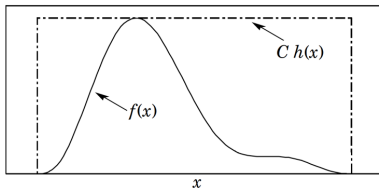
- ① 选取容易抽样的分布 $h(x)$ ，使得曲线 $Ch(x)$ 可以覆盖待抽样的分布 $f(x)$

1950 年代 von Neumann 在领导团队计算核弹引爆机制，在最早一代计算机 (MANIAC) 上运行



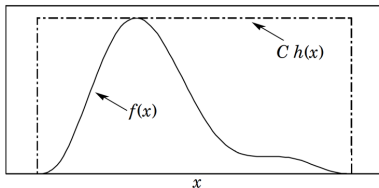
- ① 选取容易抽样的分布 $h(x)$ ，使得曲线 $Ch(x)$ 可以覆盖待抽样的分布 $f(x)$
- ② 从 $h(x)$ 抽样得 x_i

1950 年代 von Neumann 在领导团队计算核弹引爆机制，在最早一代计算机 (MANIAC) 上运行



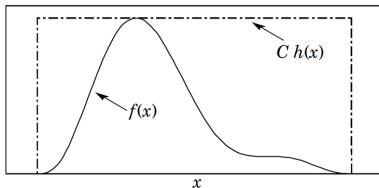
- ① 选取容易抽样的分布 $h(x)$ ，使得曲线 $Ch(x)$ 可以覆盖待抽样的分布 $f(x)$
- ② 从 $h(x)$ 抽样得 x_i
- ③ 从 $[0, 1]$ 均匀分布抽样得 r_i

1950 年代 von Neumann 在领导团队计算核弹引爆机制，在最早一代计算机 (MANIAC) 上运行



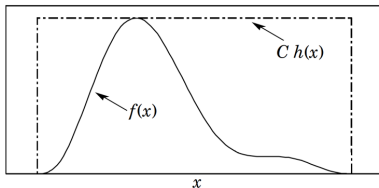
- ① 选取容易抽样的分布 $h(x)$ ，使得曲线 $Ch(x)$ 可以覆盖待抽样的分布 $f(x)$
- ② 从 $h(x)$ 抽样得 x_i
- ③ 从 $[0, 1]$ 均匀分布抽样得 r_i
- ④ 如果 $r_i Ch(x_i) \leq f(x_i)$ 则保留 x_i 否则舍弃

1950 年代 von Neumann 在领导团队计算核弹引爆机制，在最早一代计算机 (MANIAC) 上运行



- ① 选取容易抽样的分布 $h(x)$ ，使得曲线 $Ch(x)$ 可以覆盖待抽样的分布 $f(x)$
- ② 从 $h(x)$ 抽样得 x_i
- ③ 从 $[0, 1]$ 均匀分布抽样得 r_i
- ④ 如果 $r_i Ch(x_i) \leq f(x_i)$ 则保留 x_i 否则舍弃
- ⑤ 被接受的 x_i 序列，服从以 $f(x)$ 为概率密度的分布

1950 年代 von Neumann 在领导团队计算核弹引爆机制，在最早一代计算机 (MANIAC) 上运行



- ① 选取容易抽样的分布 $h(x)$ ，使得曲线 $Ch(x)$ 可以覆盖待抽样的分布 $f(x)$
- ② 从 $h(x)$ 抽样得 x_i
- ③ 从 $[0, 1]$ 均匀分布抽样得 r_i
- ④ 如果 $r_i Ch(x_i) \leq f(x_i)$ 则保留 x_i 否则舍弃
- ⑤ 被接受的 x_i 序列，服从以 $f(x)$ 为概率密度的分布
- ⑥ 使用 x_i 序列的算术平均，估算数学期望

蒙特卡罗方法

续本达

引言

复习与概念
引入

舍选法

马尔可夫链法

总结

马尔可夫链法

舍选法的关键步骤

如果 $r_i Ch(x_i) \leq f(x_i)$ 则保留 x_i ，否则舍弃。

- 困难：在高维问题中，无法知道 $f(x)$ 的绝对数值，只知道相对值 $\tilde{f}(x)$

$$f(x_i) = \frac{\tilde{f}(x_i)}{\int \tilde{f}(x) dx}$$

分母上的积分无法有效算出

舍选法的关键步骤

如果 $r_i Ch(x_i) \leq f(x_i)$ 则保留 x_i ，否则舍弃。

- 困难：在高维问题中，无法知道 $f(x)$ 的绝对数值，只知道相对值 $\tilde{f}(x)$

$$f(x_i) = \frac{\tilde{f}(x_i)}{\int \tilde{f}(x) dx}$$

分母上的积分无法有效算出

Metropolis (1915–1999, 希腊) 改进方案

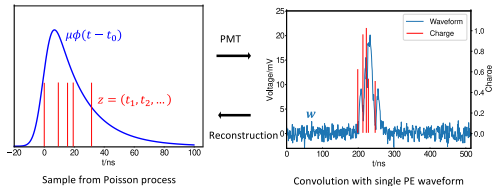
- 把 $f(x_i)$ 与上一个采样点 $f(y_{i-1})$ 比较，决定是否保留。
- 以概率 $\min \left[1, \frac{f(x_i)}{f(x_{i-1})} \right]$ 接受 x_i ，否则令 $x_i = x_{i-1}$

例子：辐射能量的测量

设 PMT 接收到的光强曲线为 $\mu\phi(t - t_0)$ ，光电子 (photoelectron, PE) 数为 N ，各光电子时间为 $\vec{z} := (t_1, t_2, \dots, t_N) \in T^N$ ， $T \subseteq \mathbb{R}$ ，波形 $\vec{w}$ 是时间序列向量。

$$p(\vec{w}|t_0, \mu) = \underbrace{p(\vec{w}|\emptyset)p(\emptyset|t_0, \mu)}_{\text{空白解0重积分}} + \underbrace{\sum_{N=1}^{\infty} \int_{\vec{z} \in T^N} p(\vec{w}|\vec{z})p(\vec{z}|t_0, \mu) d\vec{z}}_{\text{1至任意重积分的相加}}.$$

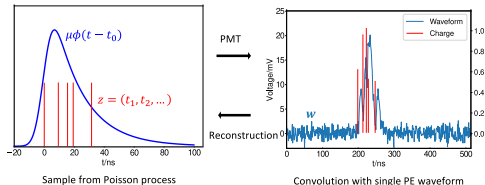
是一个极其复杂的，维度不确定的多重积分。使用马尔可夫链法得到完满解决。



设 PMT 接收到的光强曲线为 $\mu\phi(t - t_0)$ ，光电子 (photoelectron, PE) 数为 N ，各光电子时间为 $\vec{z} := (t_1, t_2, \dots, t_N) \in T^N$ ， $T \subseteq \mathbb{R}$ ，波形 $\vec{w}$ 是时间序列向量。

$$p(\vec{w}|t_0, \mu) = \underbrace{p(\vec{w}|\emptyset)p(\emptyset|t_0, \mu)}_{\text{空白解0重积分}} + \underbrace{\sum_{N=1}^{\infty} \int_{\vec{z} \in T^N} p(\vec{w}|\vec{z})p(\vec{z}|t_0, \mu) d\vec{z}}_{\text{1至任意重积分的相加}}$$

是一个极其复杂的，维度不确定的多重积分。使用马尔可夫链法得到完满解决。



Journal of Instrumentation

PAPER

Towards the ultimate PMT waveform analysis for neutrino and dark matter experiments

D.C. Xu^{1,2,3}, B.D. Xu^{1,2,3,4}, E.J. Bao^{5,6}, Y.Y. Wu^{1,2,3}, A.Q. Zhang^{1,2,3}, Y.Y. Wang^{1,2,3}, G.L. Zhang⁷, Y. Xu⁸, Z.Y. Guo^{1,2,3}, J.H. Pei⁹, H.Y. Mao¹⁰, J.S. Liu¹⁰, Z. Wang^{1,2,3} and S.M. Chen^{1,2,3} [Hide full author list](#)

Published 30 June 2022 • © 2022 IOP Publishing Ltd and Sissa Medialab

[Journal of Instrumentation, Volume 17, June 2022](#)

Citation D.C. Xu et al 2022 *JINST* 17 P06040

蒙特卡罗方法

续本达

引言

复习与概念
引入

舍选法

马尔可夫链法

总结

总结

- ↓ 已知的信息越来越少，适用范围越来越广

	方法	已知的分布信息	维度
下周	逆函数法	完全已知	低
✓	舍选法	与标准分布的比值	低
✓	马尔可夫链法	与自身的比值	高
下周	对抗生成网络法*	无解析形式，无法计算比值	高

- ↓ 已知的信息越来越少，适用范围越来越广

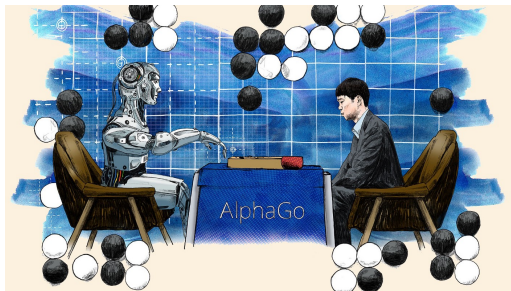
	方法	已知的分布信息	维度
下周	逆函数法	完全已知	低
✓	舍选法	与标准分布的比值	低
✓	马尔可夫链法	与自身的比值	高
下周	对抗生成网络法*	无解析形式，无法计算比值	高

延伸阅读：为什么蒙特卡罗方法有效？——

- 姚（期智）准则 (Yao's principle)：给算法合理加入随机因素，会有效提升效率。（参考文献 4）
 - 蒙特卡罗方法为物理学中 **确定性的困难问题** 提供了 **仅有的可操作** 解法。

- 蒙特卡罗方法在物理之外的几乎所有学科中都有应用（延伸阅读 1）

The results (马尔可夫链法) are really used in every aspect of scientific inquiry... I believe you can take any area of science, from hard to social, and find a burgeoning MCMC (马尔可夫链法) literature specifically tailored to that area.

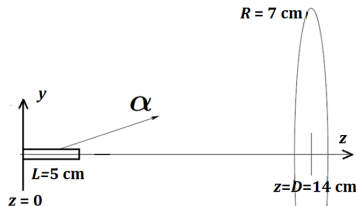


① 证明

若随机变量 X 的分布函数 $F_X(x)$ 为严格单调增的连续函数，其反函数 $F_X^{-1}(y)$ 存在，则随机变量 $Y \equiv F_X(x)$ 服从 $(0, 1)$ 上的均匀分布，即 $Y \sim U(0, 1)$ 。

② 计算

有一个长度为 $L = 5 \text{ cm}$ 的细长棒状的 α 放射源，棒的左端位于原点，棒体与 z 轴平行。放射源在棒中均匀分布，发射的 α 粒子在空间中的角分布各向同性。在 z 轴正向距离原点 $D = 14 \text{ cm}$ 处放置了一个圆盘状的探测器用来记录 α 粒子，圆盘半径 $R = 7 \text{ cm}$ ，轴线与 z 轴重合。求探测器的几何接受效率，即 α 粒子有多大概率可以入射到探测器中。（忽略细棒的半径。）



- ① Diaconis, P., 2009. The Markov chain Monte Carlo revolution. Bull. Amer. Math. Soc. 46, 179 – 205.
<https://doi.org/10.1090/S0273-0979-08-01238-X>
- ② Dellaportas, P., Roberts, G.O., 2003. An Introduction to MCMC, in: Møller, J. (Ed.), Spatial Statistics and Computational Methods, Lecture Notes in Statistics. Springer, New York, NY, pp. 1 – 41.
https://doi.org/10.1007/978-0-387-21811-3_1
- ③ MacKay, D.J.C., MacKay, vid J.C., Kay, D.J.C.M., 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press. 第 29 章, Monte Carlo Methods
- ④ Galison, P.L., 1997. Image and Logic: A Material Culture of Microphysics, 1st edition. ed. University of Chicago Press, Chicago. 第 6 章, The Electronic Image: Iconoclasm and the New Icons

- ① Xu, D.C., Xu, B.D., Bao, E.J., Wu, Y.Y., Zhang, A.Q., Wang, Y.Y., Zhang, G.L., Xu, Y., Guo, Z.Y., Pei, J.H., Mao, H.Y., Liu, J.S., Wang, Z., Chen, S.M., 2022. Towards the ultimate PMT waveform analysis for neutrino and dark matter experiments. J. Inst. 17, P06040.
<https://doi.org/10.1088/1748-0221/17/06/P06040>
- ② Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. 21, 1087 – 1092. <https://doi.org/10.1063/1.1699114>
- ③ Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97 – 109.
<https://doi.org/10.1093/biomet/57.1.97>
- ④ Yao, A.C.-C., 1977. Probabilistic computations: Toward a unified measure of complexity, in: 18th Annual Symposium on Foundations of Computer Science (Sfcs 1977). Presented at the 18th Annual Symposium on Foundations of Computer Science (sfcs 1977), pp. 222 – 227.
<https://doi.org/10.1109/SFCS.1977.24>

考虑均匀分布的概率密度函数

$$u(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \text{其它} \end{cases}$$

改写

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow{P} \int_0^1 f(x) dx \implies \frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow{P} \int_0^1 f(x) u(x) dx$$

- 如果能高效取到使 $f(X_i)$ 值更大的 X_i ，效率就会提升。

均匀分布法计算积分

X_i 在 $[0, 1]$ 区间上均匀分布,

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow{P} \int_0^1 f(x) dx$$

困难

被积函数 $f(x)$ 的值很小的地方，加和的效率低下

复习：变量替换

设随机变量 Y 的分布函数是 $G_Y(y) = P(Y < y)$ ，那么 $F_Y(Y)$ 服从 $[0, 1]$ 区间上的均匀分布。

设 Y 概率密度是 $g_Y(y) = G'_Y(y)$ ，把 x 替换为 $F_Y(y)$

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow{P} \int_0^1 f(x) dx = \int_{-\infty}^{+\infty} f[F_Y(y)] g_Y(y) dy$$

$$(\text{转移 } g_Y) \implies \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g_Y(Y_i)} \xrightarrow{P} \int_{-\infty}^{+\infty} f[F_Y(y)] dy = \int_0^1 f(x) dx$$

$$(Y_i = F_Y^{-1}(X_i)) \implies \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g_Y[F_Y^{-1}(X_i)]} \xrightarrow{P} \int_0^1 f(x) dx$$

逆函数法

$$\frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g_Y[F_Y^{-1}(X_i)]} \xrightarrow{P} \int_0^1 f(x) dx$$

- $f(X_i)$ 越接近于与 $g_Y[F_Y^{-1}(X_i)]$ 成正比，运算效率越高。

问题转化

- ① 如何 **找到** 一个随机变量 Y ，满足这个要求？
 - 非常难，大多数时间都找不到有解析形式分布函数 G_Y 的 Y

逆函数法

$$\frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g_Y[F_Y^{-1}(X_i)]} \xrightarrow{P} \int_0^1 f(x) dx$$

- $f(X_i)$ 越接近于与 $g_Y[F_Y^{-1}(X_i)]$ 正比，运算效率越高。

问题转化

- ① 如何 **找到** 一个随机变量 Y ，满足这个要求？
 - 非常难，大多数时间都找不到有解析形式分布函数 G_Y 的 Y
- ① 如何 **构造** 一个随机变量 Y ，满足这个要求？
 - von Neumann 舍选法
 - Metropolis-Hastings 马尔可夫链法
 - GAN (generative adversarial network) 对抗生成网络法

马尔可夫链法的困难

- 在更复杂的高维层级问题中, $g(y)$ 只能由一个生成过程描述, 无法有效计算相对值

解决思想

- 在问题的指引下, 有效选择以什么条件来执行采样 y_i 的生成过程
Generative Adversarial Network 对抗生成网络