Lecture 9-10

§1 Sampling (抽样)

1. Usage

1° 在已知 model / X的distribution的情况下，计算 $E[g(x)]$

2° 用于推测数据的 model / distribution

2. 两种用 sampling 计算 $E[g(x)]$ 的情况

1° simple X distribution, complicated g

2° simple g, complicated X distribution

§2 Simple X, complicated g

1. 原理

Relation between samples and models

- 对于 N 的 observations: $X_1, X_2, X_3, \cdots\cdots X_N$

  $Q(w) = \frac{\text{Number of } (X=w)}{N}$ (proportion)

- 当 $N \to \infty$ 时，有 $Q(w) \to P(w)$

  $Q(w)$ 为某一指定 outcome w 的 sample proportion

  $P(w)$ 为某一指定 outcome w 的 true probability

例：摇一个骰子足够多次，摇出六点的 sample proportion 趋近于 $\frac{1}{6}$

2. Approximation

Target value $E[g(x)] = \sum_{w \in \Omega} g(w) \cdot P(w)$

$= \sum_w g(w) Q(w)$

$= \sum$ 每一次的 outcome $\cdot \frac{1}{\text{总实验次数}}$

例：Calculate $E[e^{\sin(x)} x^2]$

- 取 X 的 N 个 samples (N 尽可能大)
- 计算 $\frac{\sum_i e^{\sin x_i} (x_i)^2}{N}$

3. 用 sampling 估算定积分

1° 本质：将定积分转化为某一连续型概率分布的数学期望形式

2° 步骤：

① 根据积分上下限与被积函数的形式决定该转化为哪种随机变量

② 配凑/拆项，构造 $\int_a^b g(x) \cdot f(x)\,dx$ 的形式，满足

- $f(x)$ 为某种 probability distribution 的 pdf
- $\int_a^b f(x)\,dx = 1$

则 $\int_a^b g(x) f(x)\,dx$ 可转化为 $E[g(x)]$

③ 在 sampling 时，X 的取值要服从 $f(x)$ 的分布

例: Calculate $\int_0^2 e^{x+\cos(x)}\,dx$

$$\int_0^2 e^{x+\cos(x)}\,dx = \int_0^2 \frac{2e^{x+\cos(x)}}{2}\,dx$$
$$= \int_0^2 2e^{x+\cos(x)} f(x)\,dx$$
$$= E[2e^{x+\cos(x)}]$$

$pdf = f(x) = \frac{1}{2}$ (check $\int_0^2 f(x)dx = 1$)

$X \sim unif[0,2]$

例: Calculate $\int_0^\infty e^{-x+\cos x}\,dx$

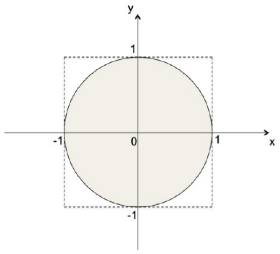
$$\int_0^\infty e^{-x+\cos x}\,dx = \int_0^\infty e^{\cos x}\cdot e^{-x}\,dx$$
$$= \int_0^\infty e^{\cos x}\cdot f(x)\,dx$$
$$= E[e^{\cos x}]$$

$pdf = f(x) = e^{-x}$ (check $\int_0^\infty e^{-x}dx = 1$)

$X \sim exp(1)$

Revisit: Estimation the value of π

- π is the ratio of the circumference of a circle to its diameter
- π is also the area of a unit disc
- Can write it as an integral: $\pi = \int_{-1}^{1}\int_{-1}^{1} 1\{x^2+y^2 \le 1\}dxdy$



- Question: if I pick a random point in the dotted square, what is the probability that it lands in the disc?
- Answer: the ratio of the areas
- More formally, a random point (X, Y) with X; Y ~Unif(-1; 1) and independent
- We are interested in $P(X^2 + Y^2 \le 1)$

§3 Simple g, complicated X : Case 1

know how data is generated, but PDF/CDF is hard to compute

1. For common R.V.

directly use built-in functions in python

2. For uncommon R.V.

Repeat many times, and take the average.

Example

- You are playing a game. For each round, you win with a probability p, and you lose with a probability q=1-p.
- Suppose that you will not play once there are N **consecutive** losses.
- How many rounds will you play the game on average?

Code

Case 1: Know how data is generated -> then generate.

- Let list = [] (the history of your paly)

One experiment:

1. Draw a Bernoulli RV X
   - if X = 1, add an element 1 at the end of the list
   - If X = 0, add an element 0 at the end of the list
2. Check:
   - If all the last N elements in list are 1, stop -> let the length of list be one sample denoting how many rounds you have played. Then stop.
   - Otherwise -> go back to step 1.

§4 Simple g, complicated X : Case 2

only know a complicated CDF : The inverse transform method (ITM) (逆转换法)

1. Principle

若 $F(x)$ 是某个变量的分布函数，记 $Y=F(x)$，则有 $Y\sim Unif(0,1)$

证明：$F_Y(y)=P(Y\le y)$

$=P(F(x)\le y)$

因为分布函数 $F(x)$ 不减，因此

$=P(X\le F^{-1}(y))$

$=F(x)\Big|_{-\infty}^{F^{-1}(y)}$

$=F(F^{-1}(y))$

$=y$

对等号两边求导，得 $P(Y)=1$

故 $Y=F(X)\sim U(0,1)$

也就是说，若已知分布函数 $F(x)$，将 $F(x)$ 记作 Y，则 $X=F^{-1}(Y)$ 且 $Y\sim Unif(0,1)$

若将 $F^{-1}(Y)$ 看作 $g(Y)$，则 $E(X)=E(g(Y))$，$Y\sim Unif(0,1)$，sampling 时对 Y 在 $(0,1)$ 上抽取随机值即可.

ITM: basic idea

- First, generate a uniform RV, U
- Then, let **X = h(U)** be a new RV, where where h= F^{-1}

U follow a uniform distribution [0,1]

$$P(U\le u)=\begin{cases}u & 0\le u\le 1\\ 0 & otherwise\end{cases}$$

X will be what we need.

X has a CDF F.

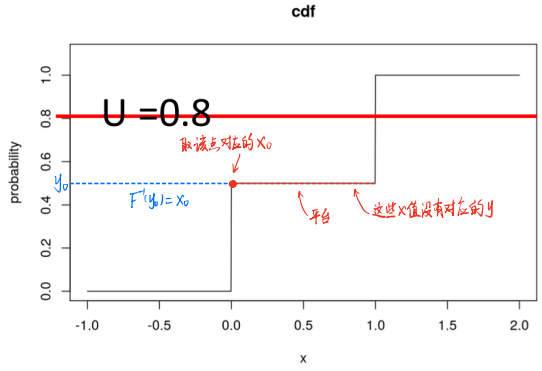
Question: how to set h function?

2. $F^{-1}(y)$ 的设定

1° 问题一：出现一个 y_0 对应多个 x 时（$F(x)$ 图象出现平台）

$F^{-1}(y_0)$ 取满足 $F(x)=y_0$ 的最小的 x

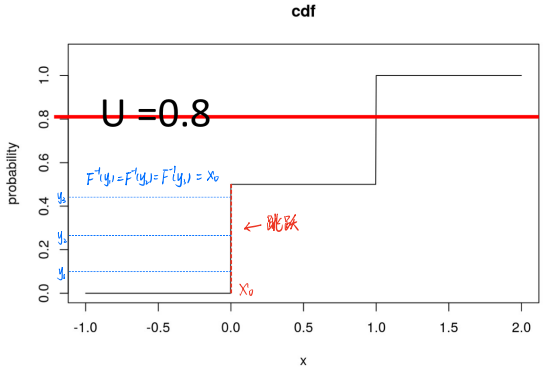
（直观体现：取 $F(x)$ 图象中平台最左端一点，对应的 x_0 的值）



2° 问题二：出现一个 y_0 没有对应的 x 时（$F(x)$ 图象出现竖直方向的跳跃）

若 $F(x_0^-)<y_0<F(x_0^+)$，则 $F^{-1}(y_0)=x_0$

（直观体现：将跳跃处用线连接起来，变成多个 y 对应一个 x_0 的情况）



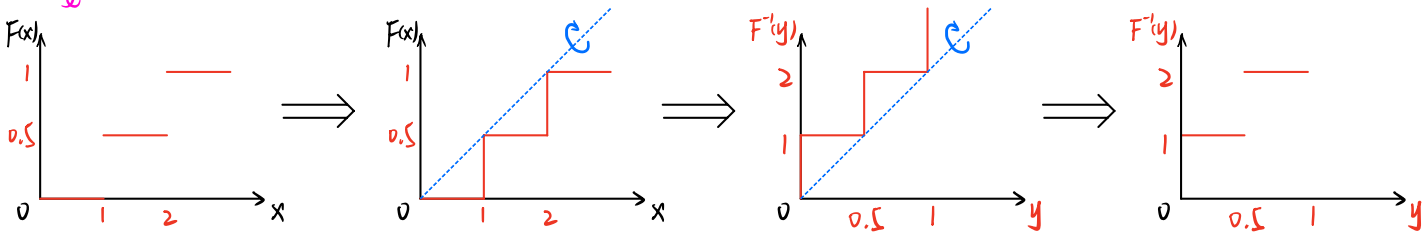
3° 总结

令 $F^{-1}(y)=\inf\{x: F(x)\geq y\}$（满足 $F(x)\geq y$ 的最小的 x）

4° 个人理解

若已知 CDF，仅需将图像跳跃处用实线连接，随后将图像关于 $y=x$ 翻折，最后去除图像上竖直的线，即得到 $F^{-1}(y)$ 关于 y 的图像。（y 为 $F(x)$，$F^{-1}(y)$ 为 x）

如

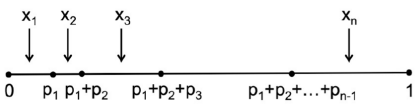


$$F(x)=\begin{cases}0, & x<1\\ 0.5, & 1\leq x<2\\ 1, & 2\leq x\end{cases}\longrightarrow F^{-1}(y)=\begin{cases}1 & 0\leq y\leq 0.5\\ 2 & 0.5<y\leq 1\end{cases}$$

例：

ITM: example 1

- R.v. X takes values x_i with probability $p_i, i=1,2,\ldots,n$ $(\sum_{i=1}^{n} p_i=1)$
- Generate $U\sim Unif(0,1)$
- If $U\in[0,p_1]$, output x_1; if $U\in(p_1,p_1+p_2]$, output x_2; etc
- A picture view



例: # ITM: example 2

- $F(x) = 1\text{-}\, e^{-x/\lambda}$
- $F^{-1}(u) = -\lambda \ln(1\text{-}\, u)$

Calculate E[g(X)]?

- First, generate a sample U from uniform[0,1]
- Then, let X = -λ ln(1-U)
- Calculate E[g(X)]

§5 Simple g, complicated X : Case 2

only know a complicated PDF : The acceptance/rejection method (ARM) (接受拒绝法)

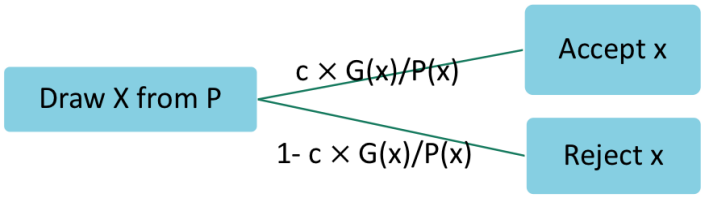
1. Principle & Basic idea

已知随机变量 X 的 PDF : G(x) (无法直接利用 G(x) 对 x 进行 sampling)

选取一个新的，容易 sampling 的 PDF P(x) (两个 PDF 有相同的 sample space)

取一个合适的常数 C，使 $C \le \frac{P(x)}{G(x)}$ 对任意 X 均成立 (C·G(x) 的图像恒在 P(x) 的下方)

- Step 1:
  按照 P(x) 的分布对 X 进行取样
- Step 2:
  对每个取出的 x_0，又有 $\frac{C \cdot G(x)}{P(x)}$ 的概率接受该 x_0.
  否则返回 Step 1 继续按 P(x) 取样，直到此 x_0 有至少一个被接受的样本



- Step 3:
  对接受的样本计算均值，即为 E[X]

ARM: basic idea

- PDF X: f(x) (**easy to sample**, e.g, uniform, exponential, normal).
- Target Y with PDF g(y)

- ARM sampling Procedure:
  - Choose constant c such that c <= f(x)/g(x) for all x
  - **Generate X from f(X)**
  - **If X=x, accept with probability c g(x)/f(x)**

Notice
1. X and Y should have the same sample space.
2. If both X and Y are discrete, simply replace the PDF by PMF

- 原理分析
  - 若总共取了 X 的 N 个样本，$N \to \infty$

- 在 $X=x$ 处，被接受的样本个数为

  $N \times P(x) \times \frac{C \times G(x)}{P(x)} = N \times C \times G(x)$

- 所有被接受的样本个数为

  $\Sigma_x \, N \times C \times G(x) = N \times C$

- 在所有被接受的样本中，$X=x$ 的占比为

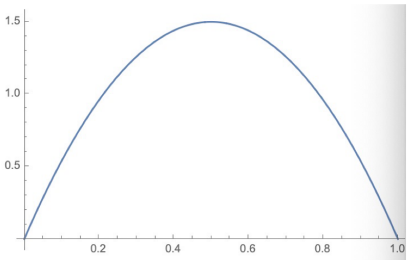
  $\frac{N \times C \times G(x)}{N \times C} = G(x)$

例：

ARM: example

- Generate a sample with g(x) =$6x(1-x)$, 0<x<1
- Use uniform to generate: i.e., f(x) = 1, 0<x<1

Choose c such that c <= f(x)/g(x) for all x



- c = ½, then c <= f(x)/g(x) for all 0<x<1
  - Generate X from a uniform distribution on [0,1]
  - If X=x, accept with probability $3\,x(1-x)$

Summary

Complicated g function, e.g.,

$\int_0^2 e^{x+\cos(x)}dx = \mathrm{E}[2e^{X+\cos(X)}]$

- **Sampling can be used to calculate E[g(X)]**
  - Sample data approximates true model

$\mathrm{E}[g(X)] = \Sigma_{\omega \in \Omega}\, g(\omega)\, P(\omega)$ → True probability

$N \to \infty$

$\Sigma_{\omega}\, g(\omega)\, Q(\omega)$ → Sample Proportions

For common R.V., we can directly use built-in functions in python

Summary

- Suppose we only knowing a nontrival CDF, how to sample?
  - **The Inverse Transform Method (ITM)**

- Suppose we only knowing a nontrival PDF, how to sample?
  - **The Acceptance/Rejection Method (ARM)**