用直接抽样和舍选法抽样

[算法及公式]：

1： 直接抽样

先将所给的数据归一化为几率,即$P_i = \frac{\sigma_i}{\sum_i \sigma_i}$.

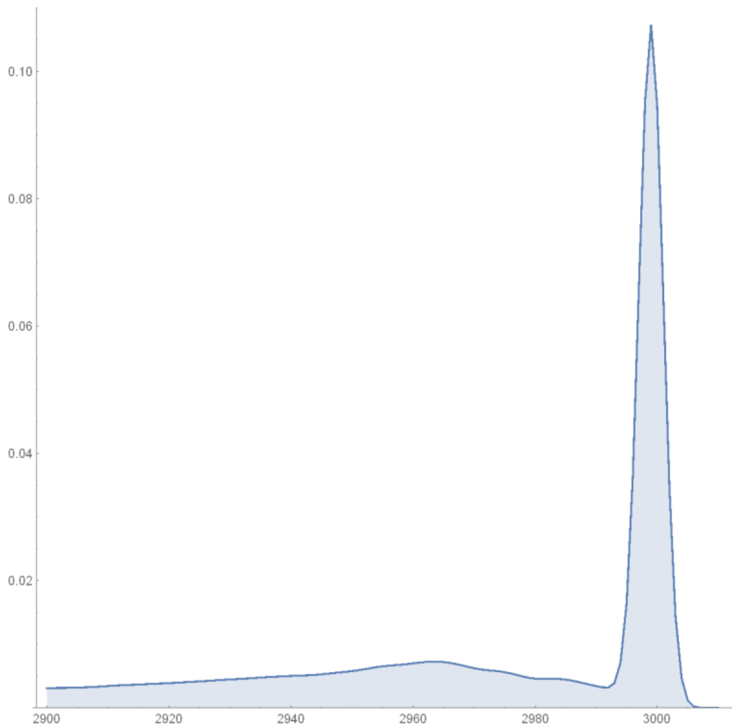
产生[0,1]之间的随机数ξ，当随机数ξ满足下式

$\sum_{i=1}^{n-1} P_i < \xi < \sum_{i=1}^{n} P_i$,物理量$x$取值为$x_n$.

本题中，x可取 111 个值。故可设一个计数数组 cy[111]，产生 N 个随机数，每个随机数产生一个对应的x_n，便将 cy[n]加一。由此便可得到一个直接抽样法的数组，然后再将数组归一化，即 cy[n]= $\frac{cy[n]}{N}$.与原分布比较。

2： 舍选法抽样

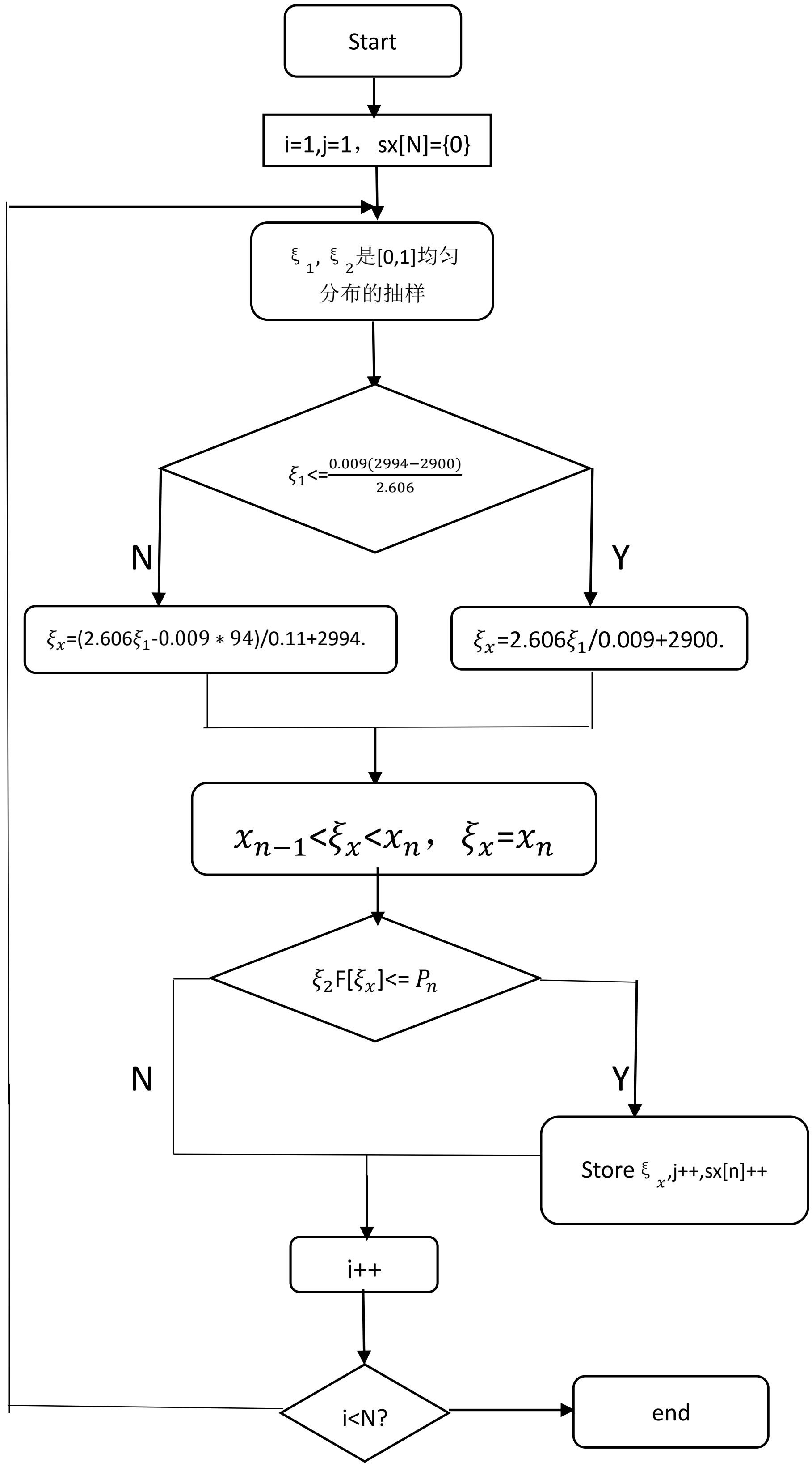
先自设 F[x]



根据分布图形，可用分段阶梯函数，将函数分成两个部分

$$F[x]=\begin{cases}0.009, x \in [2900,2994] \\ 0.11, x \in [2994,3010]\end{cases},$$

根据舍选抽样，ξ_1，ξ_2是[0,1]间的随机数组，



$\xi_1=\frac{\int_{2900}^{\xi_x}F[x]dx}{\int_{2900}^{3010}F[x]dx}$,

$\int_{2900}^{3010}F[x]dx=\int_{2900}^{2994}0.009dx$+$\int_{2994}^{3010}0.11dx$=2.606,

当ξ_x<=2994 时，

$\xi_1=\frac{\int_{2900}^{\xi_x}F[x]dx}{\int_{2900}^{3010}F[x]dx}=\frac{0.009(\xi_x-2900)}{2.606}$<=$\frac{0.009(2994-2900)}{2.606}$

ξ_x=2.606ξ_1/0.009+2900.

当ξ_x>2994 时,

$$\xi_1=\frac{\int_{2900}^{\xi_x}F[x]dx}{\int_{2900}^{3010}F[x]dx}=\frac{0.009*94+0.11(\xi_x-2994)}{2.606}$$

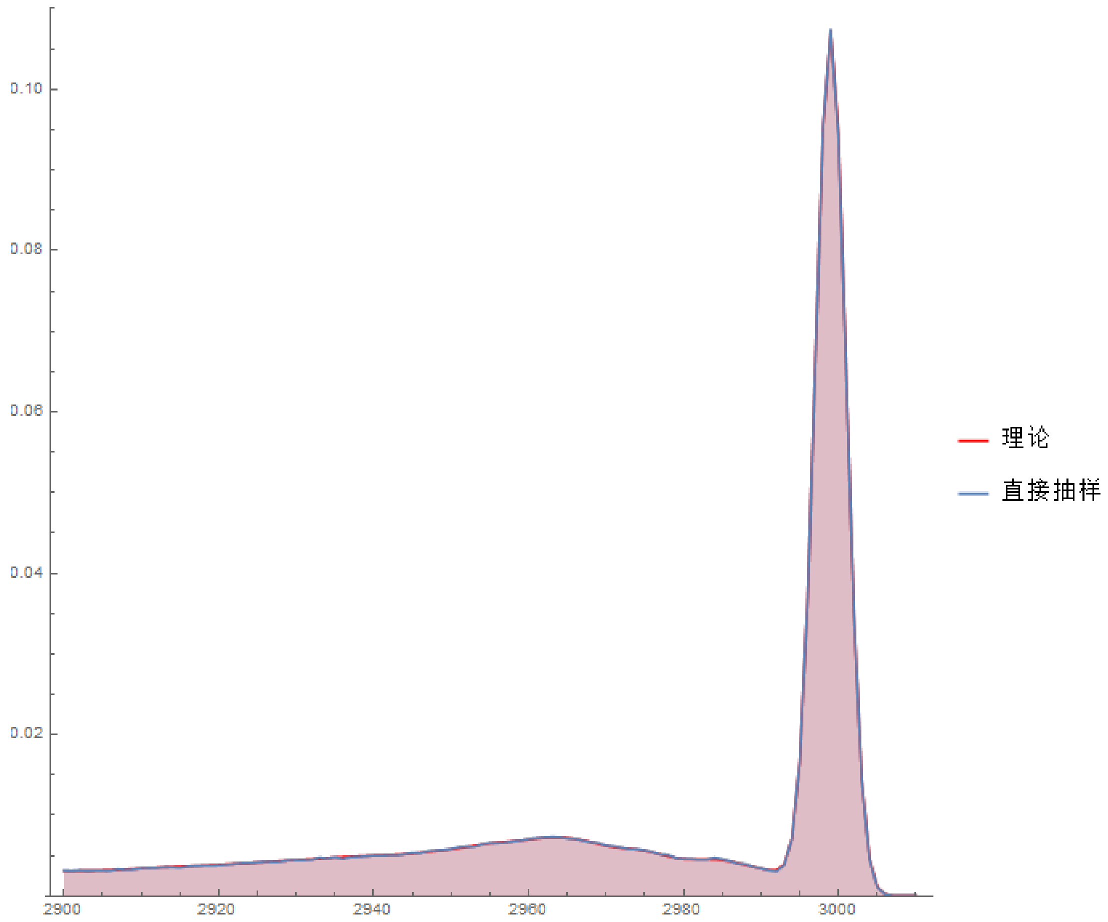
ξ_x=(2.606ξ_1-0.009 $*$ 94)/0.11+2994.

当x_{n-1}<ξ_x<x_n，ξ_x=x_n，如果ξ_2F[ξ_x]<= P_n，则取ξ_x.

同时将统计计数数组 sx[n]加一。由此便可得到一个舍选抽样法的数组，然后再将数组归一化，与 P[x]比较。

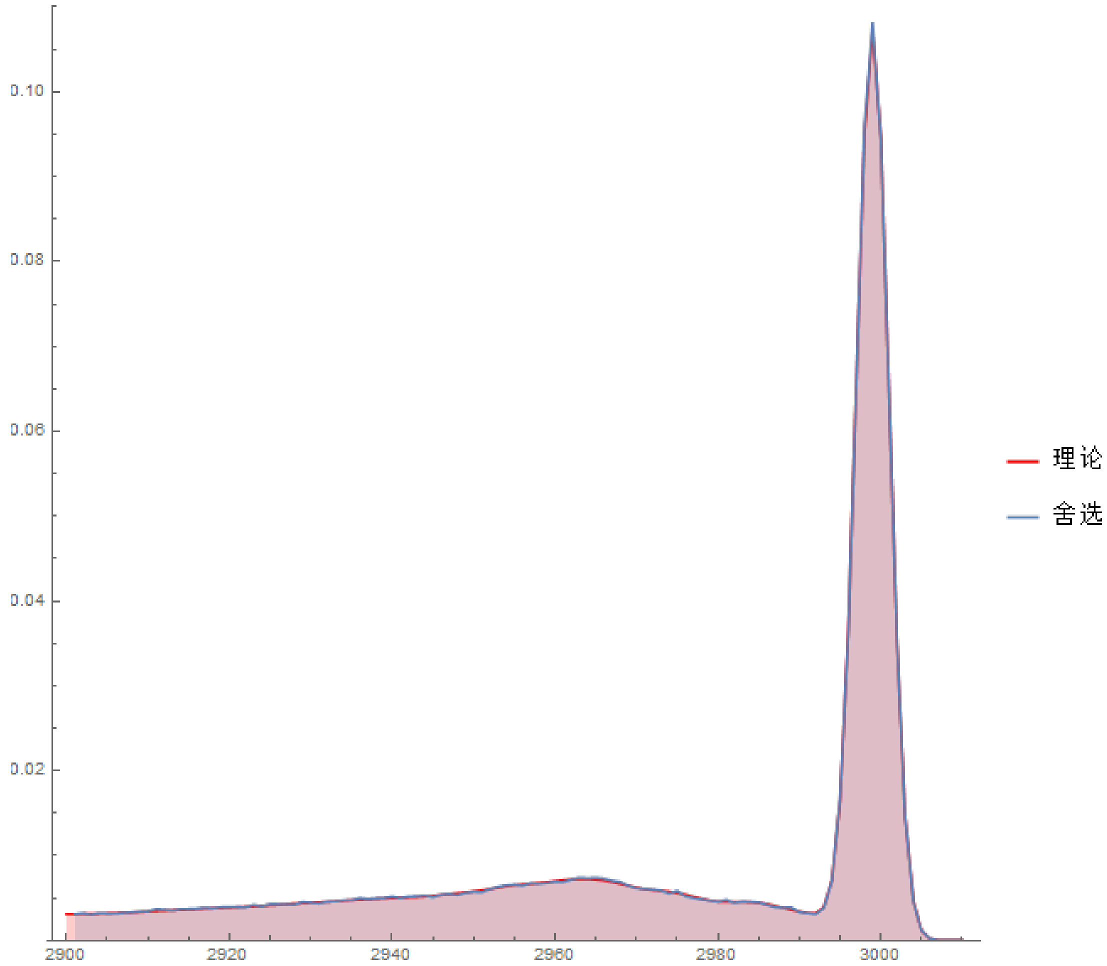
[结果与讨论]:

1：直接抽样



上图 N 取 100W，共 110 个区间，每个区间平均落入接近 10000 的随机数，所以得到的曲线与原曲线基本一致。

2：舍选抽样



N 同样取 100W，但由于舍选法抽样效率较低，实际取的点数比 100W 少，得到的曲线不如直接抽样法光滑。

[讨论]：

1：两种抽样方法的比较。

在分布已知的情况下，用直接抽样法更为简单且效率更高，而使用舍选法，由于曲线不规则，用阶梯函数法舍选时，若区间分得过多则过于麻烦，区间分的较少则抽样效率低，本题中，舍选法抽样效率为 38.3%，较为低下。故在分布已知的离散型变量分布，应采用直接抽样法。

2：舍选法抽样效率讨论。

本题分为两个区间，整体抽样效率为38.3%，较为低下，如果划分更多的区间，必然能提高抽样效率，但是更为麻烦