数学实验大作业四

塞傲霖 电 82 2018010919

以下通过四个角度对于长度为 200 的二进制序列进行随机性验证，其中角度 1~3 可以进行较为完备的 p 值计算，得到定量结果；而角度 4 理论分析有一定难度，有待完善，目前给出了半定量半定性的随机性验证结果。

初步结论：序列 A 的随机性强于序列 B 的随机性。

角度 1：平均值

经过统计：

meanA=0.465，meanB=0.545

由中心极限定理，正面次数二项分布在试验次数足够大时，逼近正态分布。用正面次数除以总试验数可以得到每次试验的正面频率。

$$\Sigma X \sim B(200,0.5) \xrightarrow{L} N(100,50)$$

$$\bar{X} = \frac{\Sigma X}{200} \sim N\left(0.5, \frac{50}{200^2}\right)$$

因此可以计算出

$$P(\bar{X} \leq 0.465) = 0.1611$$
$$P(\bar{X} \geq 0.545) = 0.1015$$
$$\therefore P(\bar{X} \leq 0.465) > P(\bar{X} \geq 0.545)$$

角度 1 的结论：序列 A 的平均值出现的概率更大。

角度 2：二值跳跃次数

统计出 A 的二值跳跃次数为 104 次，B 的二值跳跃次数为 90.

由于 200 次试验中间有 199 个间隔，每个间隔都有相等的概率跳跃或不跳跃（相当于前一次试验已知，后一次试验取相同/不同值的条件概率）。因而总二值跳跃次数 Y 服从：

$$Y \sim B(199,0.5) \xrightarrow{L} N(99.5,49.75)$$

因此可以计算出

$$P(Y \geq 104) = 0.4640$$
$$P(Y \leq 90) = 0.4243$$

$$\therefore P(Y \geq 104) > P(Y \leq 90)$$

角度 2 的结论：序列 A 的二值跳跃次数出现的概率更大。

角度 3：连续 1/0 串的长度

经过统计：

A 最大 1 串长=6，B 最大 1 串长=10.

A 最大 0 串长=6，B 最大 0 串长=10.

如果在总长度为 n 的二进制序列中要有 k 个连续的 1（0 同理）：

$$count = 2^{n-k} + 2^{n-k-1} + \ldots + 2^0 = 2^{n-k+1} - 1$$

$$P = \frac{count}{2^n} = \frac{2^{n-k+1} - 1}{2^n}$$

n=200，k=6：

$$P(k = 6) = 0.0312$$

$$P(k = 10) = 0.0020$$

$$\therefore P(k = 6) > P(k = 10)$$

角度 3 的结论：序列 A 出现的连续 0/1 串长度更短，出现概率更大。

角度 4：傅里叶变换

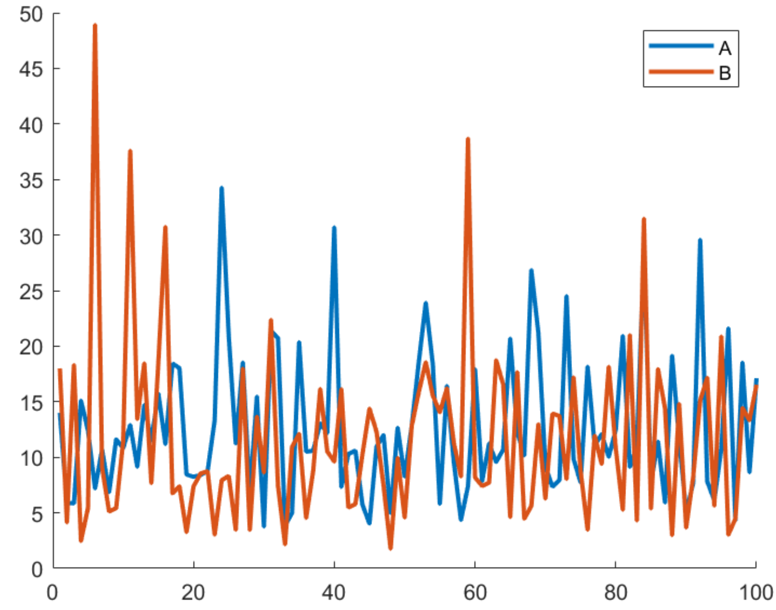
这种方法借鉴了参考文献[1]，方法的详细推导可以参照参考文献[1]。

离散傅立叶变换测试关注的是序列的离散傅立叶变换的峰值高度。该测试的目的是检测序列中周期性的模式，如果检测到这种模式，则说明序列不随机。

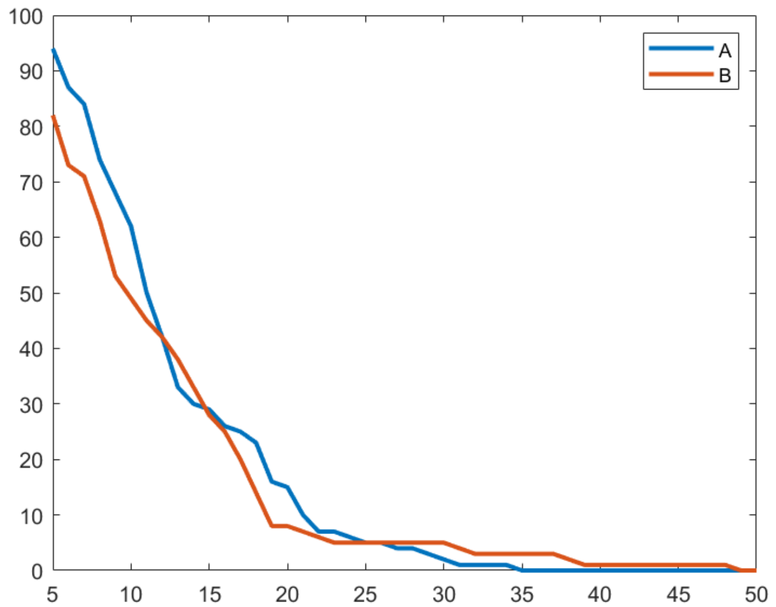
测试步骤如下：

1. 首先将待测二进制序列 ε 变换成序列 $X=x_1,x_2,...,x_n$，其中 $x_i=2\varepsilon_i-1=\pm1$。
2. 对序列 X 应用离散傅立叶变化（DFT），得到 S=DFT(X)。S 为复数序列，表示了原二进制序列中周期性的成分，并按不同的频率将其分离开来。
3. 记 S′ 为序列 S 的前 n/2 个元素组成的序列。对 S′ 取模得 M=modulus(S′)。M 表示一系列的峰值高度。
4. 选取一系列阈值 T，记录各种阈值下信号的峰值个数。
5. 如果信号随机，则峰值个数-阈值曲线稳定在一定范围内，如果信号不随机，则该曲线有

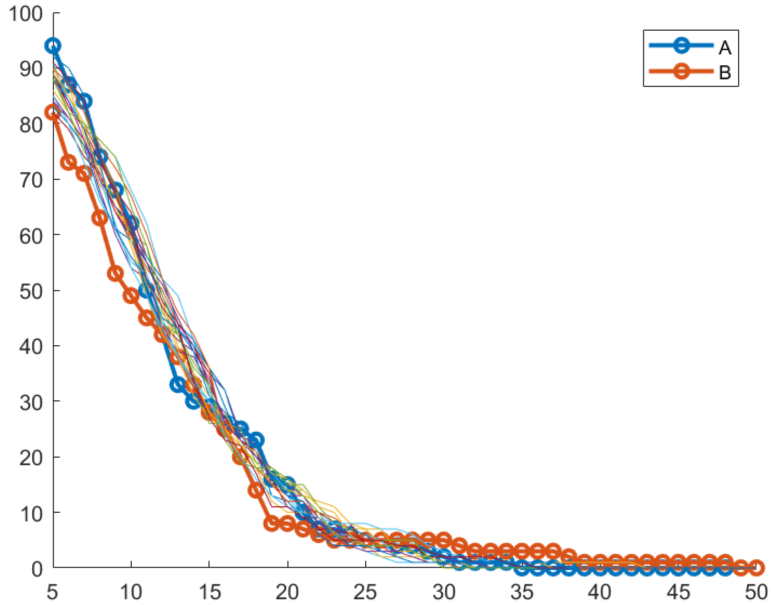
较大偏差。可以用待测序列与随机序列的均方根误差 RMSE 来定量衡量误差大小。



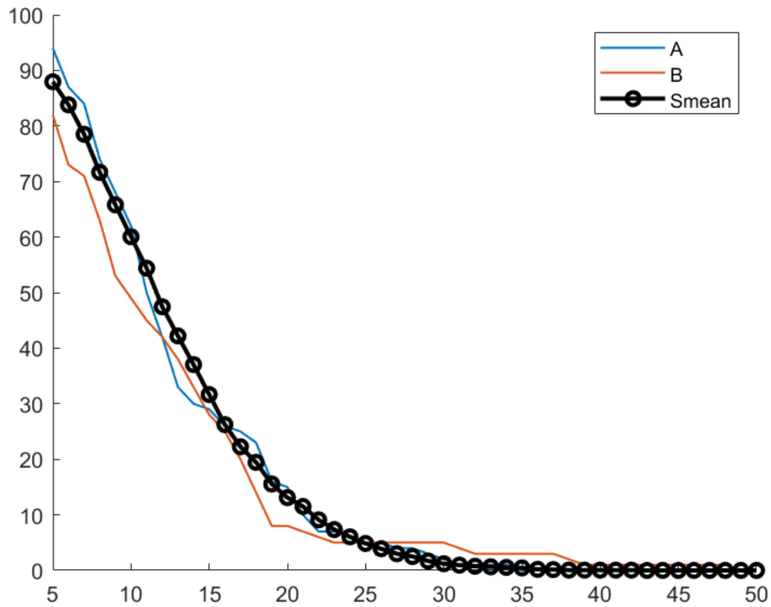
A&B 序列的傅里叶变换结果



A&B 序列的峰值个数-阈值曲线



A&B&随机模拟峰值个数-阈值曲线



A&B&平均随机模拟峰值个数-阈值曲线

```
拟合
f(x)=99.88.*exp(-0.005079.*x.^2)
Coefficients (with 95% confidence bounds):
       a =        99.88  (99.3, 100.5)
       b =     0.005079  (0.00503, 0.005129)
Goodness of fit:
  SSE: 7.353
  R-square: 0.9998
  Adjusted R-square: 0.9998
  RMSE: 0.4088
```

可以以 f(x)作为基准，计算待测序列与随机序列的均方根误差 RMSE 来定量衡量误差大小。计算出

A 序列与基准的 RMSE=2.6375

B 序列与基准的 RMSE=4.6366

模拟随机序列与基准的 RMSE 平均值（20 次试验）=1.7829

目前可以根据 RMSE 的大小进行判断：A 序列 RMSE 显著小于 B 序列 RMSE，A 序列 RMSE 更接近模拟随机序列的 RMSE，因此认为 A 序列出现的概率更大。后续可以进行更加详细的理论分析，得到 RMSE 的概率分布，可以进行拟合优度分析，利用 p 值进行判断。

参考文献：

[1]刘骋昺. 随机性统计测试的评估[D].上海交通大学,2014.