实验数据的真伪

掷硬币实验

掷100次硬币，将实验结果记录下来：1表示头像，0表示数字。

- 甲的记录：
  10011101001001101101100011010010111011011110001100
  10110110001011000101100100011011101000101110101010
- 乙的记录：
  10010101001010011010011001010101100100100111001010
  01010101100011000100100100101010101101011001110110
- 丙的记录:
  11001000101001000101000001101001111110001111010000
  00100110000000111110110100010101010001100101110101 1

问题：三人是真正做过实验，还是伪造数据？

游程Excursion

连续的0称作0游程，连续的1称作1游程。
每个0或1的个数称为游程长度。

简单统计：

- 甲的记录：52个1，48个0，连续1最多4个，连续0至多3个，0-游程30个，
- 乙的记录：47个1，53个0，连续1最多4个，连续0至多3个，0-游程36个，
- 丙的记录：46个1，54个0，连续1最多6个，连续0至多6个，0-游程25个。

例1.

将m个不可区分的球放入r个可区分的盒子，可以得到C_{r+m-1}^{m}个不同结果，每个盒子里都有球的不同结果数为$C_{r+m-r-1}^{m-r}=C_{m-1}^{r-1}$.

例2.

对于$m\geq r$，满足方程$x_1+x_2+\cdots+x_r=m$的正整数解$(x_i\geq 1)$共有C_{m-1}^{r-1}个.

例3.

独立重复同一枚硬币n次得到m个反面，分别用1和0表示正面和反面，R表示0游程的个数，则有

$$P(R=r)=\frac{C_{m-1}^{r-1}C_{n-m+1}^{r}}{C_n^m},1\leq r\leq m.$$

由公式计算出

n=100, m=47

R	22	23	24	25	26	27	28	29
P_R	0.064	0.102	0.137	0.157	0.154	0.129	0.092	0.056

以及

R	$R \leq 16$	$17 \leq R \leq 21$	$23 \leq R \leq 27$	$22 \leq R \leq 29$
P_R	1.57×10^{-4}	0.059	0.679	0.891
R	$30 \leq R \leq 34$	$R \geq 35$		
P_R	0.048	1.026×10^{-4}		

结论

甲和乙的游程数目可归为(极)小概率事件，于是，甲和乙大概率是编造数据。丙的很可能是真正实验数据。

概率理论与数理统计

条件概率

条件概率

条件概率

事件A发生的条件下，事件B发生的概率. 若$P(A) > 0$,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

即为事件B在事件A发生的条件下的条件概率.

- 在相同条件下，进行n次实验, 事件A的频数为n_A, 频率为$f_n(A)$, 事件B的频数为n_B, 频率为$f_n(B)$, A与B同时发生的频数为$n_{A\cap B}$, 频率为$f_n(A \cap B)$.

$$\frac{\text{在}A\text{发生条件下}}{B\text{发生的频率}} = \frac{n_{A\cap B}}{n_A} = \frac{n_{A\cap B}/n}{n_A/n} = \frac{f_n(A \cap B)}{f_n(A)}.$$

定理：条件概率也是一个概率

设$P(A)>0$, $P(\cdot|A)$满足概率定义的三个条件:

1. 非负性: $P(B|A)\geq 0$, $B\in\mathcal{F}$,
2. 规范性: $P(\Omega|A)=1$,
3. 可列可加性：若(B_i)为一列可数个两两互斥集合，则

$$P(\cup_{i=1}^{+\infty}B_i|A)=P(\sum_{i=1}^{+\infty}B_i|A)=\sum_{i=1}^{+\infty}P(B_i|A).$$

加法公式

$\mathcal{F}$中任意B_1, B_2,

$$\mathbf{P}(B_1\cup B_2|A)=\mathbf{P}(B_1|A)+\mathbf{P}(B_2|A)-\mathbf{P}(B_1\cap B_2|A)$$

定理

若A_1, A_2, $\cdots$, $A_n\in\mathcal{F}$, 且$P(A_1\cdots A_n)>0$, 则

$$P(A_1\cdots A_n)=P(A_1)P(A_2|A_1)\cdots P(A_n|A_1\cdots A_{n-1}).$$

全概率公式与Bayes公式

事件列$(B_i)_{1\le i\le n}\in\mathcal{F}$, 称$B_1,\cdots,B_n$为$\Omega$的一个划分, 若它满足

1. $B_i\cap B_j=\emptyset$, $i\neq j$,
2. $B_1\cup B_2\cup\cdots\cup B_n=\Omega$.

全概率公式

设事件列$(B_i)_{1\le i\le n}\in\mathcal{F}$为$\Omega$的一个划分, 且$P(B_i)>0$, 那么, 对于事件$A\in\mathcal{F}$

$$P(A)=\sum_{i=1}^{n}P(A|B_i)P(B_i).$$

Bayes公式

设事件列$(B_i)_{1\le i\le n}$为Ω的一个划分, 且$P(B_i)>0$, 事件A满足$P(A)>0$, 那么

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n}P(A|B_i)P(B_i)},i=1,\cdots,n.$$

独立性

定义

两个事件A与B为相互独立的，如果满足

$$P(A\cap B)=P(A)P(B).$$

定理

设A与B相互独立，

- 若$P(A)>0$, 则$P(B|A)=P(B)$;
- A与B^c, A^c与B, A^c与B^c相互独立.

定义: 设三个事件A, B与C

- 若$P(AB)=P(A)P(B)$, $P(BC)=P(B)P(C)$, $P(AC)=P(A)P(C)$, 则称三个事件两两独立;
- 若还有$P(ABC)=P(A)P(B)P(C)$, 则称三个事件相互独立.

推广：事件类的独立性

- 事件类$(A_i)_{i\in I}$的两两独立：若对于任意$i,j\in I$, $i\neq j$, $P(A_iA_j)=P(A_i)P(A_j)$;
- 事件类$(A_i)_{i\in I}$的相互独立：若对于任意I的有限子集J, $P(\cap_{i\in J}A_i)=\prod_{i\in J}P(A_i)$;
- $\mathcal{F}$的子σ-代数$(\mathcal{A}_i)_{i\in I}$称为两两独立，若对于任意$i,j\in I$, $i\neq j$, 任意$A_i\in\mathcal{A}_i$, $A_j\in\mathcal{A}_j$, 总有

$$P(A_i\cap A_j)=P(A_i)P(A_j).$$

- $\mathcal{F}$的子σ-代数$(\mathcal{A}_i)_{i\in I}$称为相互独立，若对于任意I的有限子集J, 任意$A_i\in\mathcal{A}_i$, $i\in J$, 总有

$$P(\cap_{i\in J}A_i)=\prod_{i\in J}P(A_i).$$

在上面的命题中，将部分集合替换为其补集，结论不变.