

(2023-2024-1)

100 , 120 .

: 本次测试为限时测试, 请尝试每一个题目, 可以参看任何相关资料以及我在课堂上介绍过的任何结论 (若使用某些在课堂上我没有讲过的结论 (其实不需要), 请详细标注出来), 但是不可以相互讨论、抄袭. 若发现抄袭现象, 则平时成绩记为 0 分. 每个大题请在新的一页开始, 切勿紧接着前面的题目书写. 以下三题合计 90 分, 书写整洁规范、符合要求 10 分, 总计 100 分.

1. (30) DNA , (A, G, C, T). 0 (), (: (iii) (iv) , (iii)):

(i) 10 , DNA AG ; 10 分

记事件 $B_i, i = 1, 2, \dots, 9$ 分别表示第 $i, i+1$ 时刻复制的碱基为 A, G 这一事件, 我们先考虑所求事件的对立事件, 即到时刻 10 为止, DNA 复制期间曾出现 AG 这一片段的概率 $\mathbb{P}(B_1 \cup B_2 \cup \dots \cup B_9)$. 利用 Jordan 公式, 我们得到此概率等于

$$\sum_i \mathbb{P}$$

```
# A=1, T=2, C=3, G=4
x<-1:4
y=vector(length=10^6)
for(i in 1:10^6)
{z=sample(x,10,replace=T)
 z=paste(z[1],z[2],z[3],z[4],z[5],z[6],z[7],z[8],z[9],z[10],sep="")
 y[i]=grepl("14",z)}
1-sum(y)/10^6
# [1] 0.538296
```

$$(2) \quad \sum_{k=0}^n \binom{n}{k} \left(\frac{1}{4} \right)^k \left(\frac{3}{4} \right)^{n-k} = 1$$

(ii) 10, DNA ; .. 10 分

记 $C_i, i = 1, 2, 3, 4$ 分别表示 A, G, C, T 在到时刻 10 为止没有出现这一事件, 那么所求概率为 $1 - \mathbb{P}(C_1 \cup C_2 \cup C_3 \cup C_4)$. 利用 Jordan 公式可知

$$\begin{aligned} \mathbb{P}(C_1 \cup C_2 \cup C_3 \cup C_4) &= \sum_i \mathbb{P}(C_i) - \sum_{i < j} \mathbb{P}(C_i C_j) + \sum_{i < j < k} \mathbb{P}(C_i C_j C_k) \\ &= 4 * (3/4)^{10} - 6 * (2/4)^{10} + 4 * (1/4)^{10} \approx 0.22 \end{aligned}$$

故所求概率到时刻 10 为止, DNA 复制期间四种类型的碱基曾全部出现的概率 ≈ 0.78 .

(iii) DNA AG I; DNA
AT II.
AG AT I ;
10 分

记事件 D 表示最后编译出来的蛋白质类型为 I , 关于前两个时刻的编译结果使用全概率公式可得到 ($*A$ 表示第二个时刻编译出来的是碱基 A 这一事件)

$$\mathbb{P}(D) = \frac{1}{16} + \frac{4}{16} \mathbb{P}(D|*A) + \frac{10}{16} \mathbb{P}(D)$$

对 $\mathbb{P}(D|*A)$ 关于第三个时刻的编译结果继续使用全概率公式得到

$$\mathbb{P}(D|*A) = \frac{1}{4} + \frac{1}{4} \mathbb{P}(D|*A) + \frac{1}{4} \mathbb{P}(D)$$

由上面两个式子可以解出 $\mathbb{P}(D) = 1/2$.

: , T, G ,
; (?),
1/2.

(iv) $\xi \sim \text{Pois}(4)$, A
 G .

: Poisson 10 分

利用 Poisson 随机变量的可分性, 我们知道最终 A 的个数和 G 的个数相互独立且均为 $\text{Pois}(1)$ 分布, 我们分别用 X, Y 表示, 那么所求概率即为 $\mathbb{P}(X > Y)$. 利用对称性, 我们仅需要计算 $\mathbb{P}(X = Y)$ 即可, 为此注意到

$$\mathbb{P}(X = Y) = \sum_{k=0}^{\infty} \mathbb{P}(X = Y = k) = e^{-2} \sum_{k=0}^{\infty} \frac{1}{(k!)^2} := c,$$

故最终所求概率为 $(1 - c)/2 \approx 0.41$, 其中 c 由上式定义给出.

: (i) (ii) (iv), , .

2. (30) $0, 1$.
 $p(\quad 1 \quad 1 \quad 0 \quad 0)$
 $q(\quad 1 \quad 0 \quad 0 \quad 1)$.
 A $1(\quad , \quad 0, \quad 1), A_n$
 n 1. $\alpha = \mathbb{P}(A), \beta = \mathbb{P}(\bar{A})$.

(i)

$$\mathbb{P}(A|A_1) = \frac{\alpha + \alpha(p - q)}{1 + (\alpha - \beta)(p - q)};$$

..... 10 分

利用 Bayes 公式以及 $p + q = 1, \alpha + \beta = 1$ 的事实, 可知

$$\begin{aligned} \mathbb{P}(A|A_1) &= \frac{\mathbb{P}(A_1|A)\mathbb{P}(A)}{\mathbb{P}(A_1|A)\mathbb{P}(A) + \mathbb{P}(A_1|A^c)\mathbb{P}(A^c)} \\ &= \frac{\alpha p}{\alpha p + \beta q} \\ &= \frac{2\alpha p}{2\alpha p + 2\beta q} \\ &= \frac{\alpha p + \alpha(1 - q)}{\alpha p + \alpha(1 - q) + \beta q + \beta(1 - p)} \\ &= \frac{\alpha + \alpha(p - q)}{1 + (\alpha - \beta)(p - q)} \end{aligned}$$

(ii) $n \geq 2$, $\mathbb{P}(A|A_n)$; 15 分

我们最后的结论是

$$\mathbb{P}(A|A_n) = \frac{\alpha + \alpha(p - q)^n}{1 + (\alpha - \beta)(p - q)^n}, \quad n \geq 1.$$

$$\mathbb{P}(A_n|A) = \frac{1}{2} + \frac{1}{2}(p-q)^n, \quad \mathbb{P}(A_n|A^c) = \frac{1}{2} - \frac{1}{2}(p-q)^n \quad ? \quad 9$$
$$\begin{aligned}\mathbb{P}(A_n|A) &= \mathbb{P}(A_n|A, A_{n-1})\mathbb{P}(A_{n-1}|A) + \mathbb{P}(A_n|A, A_{n-1}^c)\mathbb{P}(A_{n-1}^c|A) \\ &= p\left[\frac{1}{2} + \frac{1}{2}(p-q)^{n-1}\right] + q\left[\frac{1}{2} - \frac{1}{2}(p-q)^{n-1}\right] \\ &= \frac{1}{2} + \frac{1}{2}(p-q)^n.\end{aligned}$$

$\xi \sim B(n, q)$, 这是因为每个中继站是相互独立的传递错误且概率一致, 故符合二项分布的背景.

$$(i) \quad F_X, F_X^{-1}, \Phi, \Phi^{-1}, \quad g, \quad Y := g(X) \quad N(0, 1),$$

$$\mathbb{P}(Y \leq y) = \mathbb{P}(\Phi^{-1} \circ F_X(X) \leq y) = \mathbb{P}(X \leq F_X^{-1}(\Phi(y))) = F_X(F_X^{-1}(\Phi(y))) = \Phi(y),$$
$$(ii) \quad X_1, X_2, \dots \quad F_X(\cdot),$$

$$N := \min \left\{ n \mid n \geq 2, X_n > X_{n-1}^2 \right\}$$

由 (i), 不妨假定这里的 $X_1, X_2, \dots$ 为独立同分布的 $U(0, 1)$, 因为 $X_1, X_2, \dots$ 之间的相对大小在单调递增函数的变换下保持序不变. 此外, 注意到对任意的 $n \geq 2$, $\{N > 2\} = \{X_1 \geq X_2 \geq X_3 \geq \dots \geq X_n\}$, 再由提示可知

$$\begin{aligned}\mathbb{P}(N > n) &= \mathbb{P}(X_1 \geq X_2 \geq X_3 \geq \dots \geq X_n) \\ &= \mathbb{P}(X_1 > X_2 > X_3 > \dots > X_n) \\ &= 1/n!,\end{aligned}$$

注意上式子对 $n = 1$ 时也成立, 故最终得到 N 的分布列

$$\mathbb{P}(N = n) = \mathbb{P}(N > n-1) - \mathbb{P}(N > n) = \frac{1}{(n-1)!} - \frac{1}{n!}, \quad n = 2, 3, \dots$$

(iii) $X \sim U(0, 1)$, ,

$$Z = \begin{cases} 1/X, & ; \\ 1-X, & . \end{cases}$$

Z (; ,); 10 分

我们先求分布函数, 注意到 $Z > 0$, 故对任意的 $z > 0$, 利用全概率公式

$$\begin{aligned}\mathbb{P}(Z \leq z) &= \mathbb{P}(Z \leq z|H) * (1/2) + \mathbb{P}(Z \leq z|T) * (1/2) \\ &= \mathbb{P}(1/X \leq z|H) * (1/2) + \mathbb{P}(1-X \leq z|T) * (1/2) \\ &= \mathbb{P}(1/X \leq z) * (1/2) + \mathbb{P}(1-X \leq z) * (1/2)\end{aligned}$$

为计算上述概率, 我们需要对 z 进行讨论

若 $z \geq 1$, 那么 $\mathbb{P}(Z \leq z) = 1/2 + (1/2) * \mathbb{P}(X \geq 1/z) = 1 - 1/(2z)$;

若 $0 < z < 1$, 那么 $\mathbb{P}(Z \leq z) = z/2$.

可见, Z 的分布函数在 $(0, +\infty)$ 上连续且处处可导, 故 Z 是连续型随机变量且其密度函数为

$$f_Z(z) = \frac{1}{2z^2} \mathbb{I}_{(1, +\infty)}(z) + \frac{1}{2} \mathbb{I}_{(0, 1]}(z) = \frac{1}{2} \min\{1, z^{-2}\} \mathbb{I}_{(0, +\infty)}(z).$$

(iv) (iii) , $W = XZ$, W ? ,
; ; 5 分

不是! 注意到 $\mathbb{P}(W = 1) \geq \mathbb{P}(\{H\}) = 1/2$ 即可.

(v) (iii) Z ?
 $x \in \mathbb{R}$, $\{Z \leq x\}$ 5 分

按照定义, 我们知道对任意的 $x < 0, \{Z \leq x\} = \emptyset$, 对任意的 $x \geq 0$,

$$\begin{aligned}\{Z \leq x\} &= \{Z \leq x, H\} + \{Z \leq x, T\} \\ &= \{1/X \leq x, H\} + \{1 - X \leq x, T\} \\ &= \{X \geq x^{-1}, H\} + \{X \leq 1 - x, T\} \\ &= \{X \geq x^{-1}\} \cap \{H\} + \{X \leq 1 - x\} \cap \{T\},\end{aligned}$$

上式中的四个集合均为事件, 故经过上述有限交、并运算后 (利用事件域的封闭性) 依然是事件.

： “+”， .