练习二

1. 名词解释: 互斥事件; 独立事件; 随机变量; 离散型随机变量; 连续型随机变量; 两点分布; 两项分布; 正态分布; 似然函数; 极大似然估计; 似然比检验; 最小二乘估计; 孟德尔分离定律; 孟德尔自由结合定律; 世代矩阵

2. 在孟德尔的豌豆杂交试验中, 控制黄色籽粒的基因用 Y 表示, 其等位基因用 y 表示, Y 对 y 为显性, 纯合基因型 yy 的籽粒为绿色. 现有两个纯系, 分别用 P_1 和 P_2 表示, P_1 的基因型为 YY, 籽粒颜色为黄色; P_2 的基因型为 yy, 籽粒颜色为绿色.

(1) 给出 P_1 和 P_2 杂交产生 F_1 杂种的基因型和籽粒颜色;

(2) 给出杂种 F_1 与 P_1 的回交群体中各种基因型和表型的频率; 给出杂种 F_1 与 P_2 的回交群体中各种基因型和表型的频率; 由此说明, 为什么遗传学研究中测交试验一般选择与隐性亲本做回交;

(3) 给出杂种 F_1 的自交群体 (即 F_2) 中, 各种基因型和表型的频率;

(4) 利用条件概率公式, 计算 F_2 群体黄色籽粒的基因型为 YY 的频率.

3. 在孟德尔的豌豆杂交试验中, 控制黄色籽粒的基因用 Y 表示, 其等位基因用 y 表示, Y 对 y 为显性, 纯合基因型 yy 的籽粒为绿色; 控制圆型籽粒的基因用 R 表示, 其等位基因用 r 表示, R 对 r 为显性, 纯合基因型 rr 的籽粒为皱缩型; 已知这两对基因位于不同染色体上. 现有两个纯系, 分别用 P_1 和 P_2 表示, P_1 的基因型为 YYRR, 为黄色和圆型籽粒; P_2 的基因型为 yyrr, 为绿色和皱缩型籽粒.

(1) P_1 和 P_2 杂交产生的 F_2 群体中, 列出各种基因型的频率, 以及各种基因型对应的表型;

(2) F_2 群体中的表型有哪些? 利用概率加法定律, 计算各种表型的频率.

4. 一个离散随机变量 X 四种可能取值的概率为:

X	0	1	2	3
P	0.729	0.243	0.027	0.001

(1) 给出 X 的分布函数, 绘制分布函数曲线;

(2) 计算 $P(X\leqslant 0.5)$ 和 $P(1\leqslant X\leqslant 1.5)$;

(3) 计算随机变量 X 的均值和方差.

5. 随机变量 X 服从均值为 5, 方差为 4 的正态分布, 即 $X\sim N(5, 4)$. 利用 Excel 中的正态分布函数,

(1) 计算满足条件 $P(X<a)=0.9$ 的 a 的取值;

(2) 计算满足条件 $P(|X\text{-}5|>a)=0.01$ 的 a 的取值;

(3) 在 Excel 中绘制正态分布 $N(5, 4)$的概率密度函数曲线和概率分布函数曲线.

6. 在孟德尔的豌豆杂交试验中，控制黄色籽粒的基因用 Y 表示，其等位基因用 y 表示, Y 对 y 为显性，纯合基因型 yy 的籽粒为绿色. 现有两个纯系，分别用 P_1 和 P_2 表示, P_1 的基因型为 YY，籽粒颜色为黄色; P_2 的基因型为 yy，籽粒颜色为绿色. P_1 和 P_2 杂交产生的 F_2 群体中，黄色和绿色籽粒的频率分别为 0.75 和 0.25. 今随机调查 8 颗 F_2 籽粒，用变量 X 表示黄色籽粒的数目.

(1) 变量 X 有哪些可能的取值? 服从什么样的分布?

(2) 利用 Excel 中的两项分布函数，计算 X 不同取值的概率;

(3) 在 Excel 中绘制概率分布的柱形图;

(4) 一次调查中，正好有 6 个黄色籽粒和 2 个绿色籽粒的可能性是多大? 10 次重复调查中，每次都是有 6 个黄色籽粒和 2 个绿色籽粒的可能性有多大?

7. 利用 Excel 中的随机数发生器，即 RAND()函数，产生一个 7×5 矩阵，用 **A** 表示. 利用 Excel 中的与矩阵运算相关的函数,

(1) 计算 **A** 的转置矩阵，用 $\mathbf{A}^{\mathrm{T}}$ 表示;

(2) 计算乘积矩阵 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 和 $\mathbf{A}\mathbf{A}^{\mathrm{T}}$;

(3) 计算矩阵 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 和 $(\mathbf{A}\mathbf{A}^{\mathrm{T}}+\mathbf{I})$ 的逆矩阵，其中 **I** 表示单位矩阵.

8. 已知两个亲本 P_1 和 P_2 的基因型为 AA 和 aa，杂种 F_1 代基因型的频率为 $f_{AA}^{(0)}=0$ 和 $f_{Aa}^{(0)}=1$，用向量的形式表示为 $f^{(0)}=\begin{bmatrix} f_{AA}^{(0)} \\ f_{Aa}^{(0)} \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. 类似地用 $f^{(t)}=\begin{bmatrix} f_{AA}^{(t)} \\ f_{Aa}^{(t)} \end{bmatrix}$表示 F_1 与亲本 P_1 连续回交 t 代的群体中两种基因型频率构成的向量. 试用世代矩阵（转移矩阵）表示回交 t 代的基因型频率向量和 t-1 代的关系，并由此计算回交 3 代后，群体中 2 种基因型的频率.

9. 设有两个基因座位 A-a 和 B-b 间一次交换的重组率为 r=0.05，两个亲本的基因型假定为 AABB 和 aabb. 从 AABB×aabb 的 F_1 代开始，在以后的自交世代中，有 10 种可能的基因型，这些基因型可分为以下 5 类: (i) AABB 和 aabb; (ii) AAbb 和 aaBB; (iii) AABb, aaBb, AaBB 和 Aabb; (iv) AB/ab; (v) Ab/aB. 各类中，不同基因型的频率相等. 从各类基因型自交产生的后代类型可以得到世代矩阵为 $\mathbf{T}=\begin{bmatrix} 1 & 0 & \frac{1}{4} & \frac{1}{2}(1-r)^2 & \frac{1}{2}r^2 \\ 0 & 1 & \frac{1}{4} & \frac{1}{2}r^2 & \frac{1}{2}(1-r)^2 \\ 0 & 0 & \frac{1}{2} & 2r(1-r) & 2r(1-r) \\ 0 & 0 & 0 & \frac{1}{2}(1-r)^2 & \frac{1}{2}r^2 \\ 0 & 0 & 0 & \frac{1}{2}r^2 & \frac{1}{2}(1-r)^2 \end{bmatrix}$.

(1) 写出重组率 r=0.05 时的世代矩阵 **T**;

(2) F_1 代开始, 计算自交二代 (即 F_3) 群体中, 五类基因型的频率;

(3) F_1 代开始, 计算自交五代 (即 F_6) 群体中, 五类基因型的频率;

(4) F_1 代开始, 计算自交八代 (即 F_9) 群体中, 五类基因型的频率;

(5) 在(2)-(4)中, 计算类型(i)和(ii) (即纯合基因型) 在群体中的比例; 由此说明, 为什么创制植物重组近交家系群体时, 一般需要 5 代或更多代的连续自交过程.

10. 在 r=0.5 的条件下, 重复练习 9, 由此说明纯合基因型频率与有无连锁的关系不大.

11. 利用 Excel 中的随机数发生器, 即 RAND()函数, 产生 100 组随机样本, 每组样本由 5 个服从均匀分布 U(0, 1)的随机数组成.

(1) 计算每组样本的均值和方差;

(2) 绘制 100 个均值和 100 个方差的次数分布图;

(3) 如将样本量由 5 增加到 10, 重复 (1) 和 (2), 并由此说明样本量对估计总体参数 (如总体的均值和方差等) 的重要性.

12. 利用下面的两个结论, 在 Excel 中产生 100 组随机样本, 每组样本由 5 个服从正态分布 N(5, 10)的随机数组成.

结论 I: 如果 X_1 和 X_2 是两个独立的 U(0, 1)随机数, 则变换 $Y_1=\sqrt{-2\ln(X_1)}\sin(2\pi X_2)$ 和 $Y_2=\sqrt{-2\ln(X_1)}\cos(2\pi X_2)$ 得到的 Y_1 和 Y_2 服从标准正态分布 N(0, 1), 且相互独立.

结论 II: 如果 X 是标准正态分布 N(0, 1)随机数, 则变换 $Y=\sigma X+\mu$ 得到的 Y 服从正态分布 $N(\mu,\sigma^2)$.

(1) 计算每组样本的均值和方差;

(2) 绘制 100 个均值和 100 个方差的散点图和次数分布图;

(3) 如将样本量由 5 增加到 10, 重复 (1) 和 (2), 并由此说明样本量对估计总体参数 (如总体的均值和方差等) 的重要性.