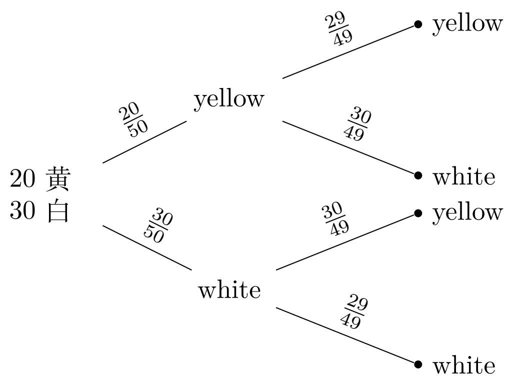
1 概率统计

1.1 tips

Use tree diagram or venn diagram，it will be very easy.

But venn diagram is usually meant for sets, when used in Probability, it must be reconsidered in your mind.
公式类的题多用 venn，叙述性的题多用 tree

But how to describe the tree branch is the most important



1.2 排列、组合

$P_n^m = n(n-1)(n-2)\cdots(n-m+1) = \frac{n!}{(n-m)!}$

$C_n^m = \frac{P_n^m}{P_m^m} = \frac{P_n^m}{m!} = \frac{n!}{(n-m)!m!}$

0,1,2,3,4,5,6,7 排成三位数：1. loci: 7x7x6; 2.$P_8^3 - P_7^2$; 3. loci $P_7^1 P_7^2$
1,1,1,2,2,3(可用来理解 C) 排六位数：$P_6^6/3!2!$

捆绑法：5 男 3 女，3 女必须在一起：$=P_6^6 P_3^3$
插空法：8 学生 4 老师，老师不能相邻，且学生在中间 -0-0-0-0-0-0-0-0- $=P_8^8 P_7^4$
隔板法: n 元素, m 堆, 每堆至少一个, 相当于 m-1 个木板插入 n-1 个空 $=C_{n-1}^{m-1}$

1.3 Probability

全概率公式：$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$, $A_1 \cdots A_n$ 为一个完备事件组

So we have: $P(B) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$

贝叶斯定理 (Bayes' theorem): $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

So we have: $P(AB) = P(A|B)P(B) = P(B|A)P(A)$

Visualizing bayes theorem[1]
$P(A|B) = \frac{|AB|}{|B|} = \frac{\frac{|AB|}{|U|}}{\frac{|B|}{|U|}} = \frac{P(AB)}{P(B)}$, that means take B as Universe(U)

$P(A-B) = P(A\bar{B}) = P(A) - P(AB)$
$P(A \cup B) = P(A) + P(B) - P(AB)$

1.4 独立和互斥

独立和互斥是两个概念

A 和 B 独立 (independent) $\Leftrightarrow P(AB) = P(A)P(B)$

SO, when independent $P(A|B) = P(A)$ and $P(B|A) = P(B)$

独立情况下 venn 图大多是 AB 交叉的
互斥情况下 venn 图是不交叉的

1.5 分布函数、期望、方差

分布函数：$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)dt, -\infty < x < \infty$
$F'(x) = f(x)$

期望 (大数定律) 与平均值：期望是样本趋于无穷的极限
例：掷色子 (可以用来理解 E, 离散型和连续型), 2,2,2,6,4 average=(2+2+2+6+4)/5=3
而 E 是固定的 $=1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$, 但是 3.5 数值是扔不出的

方差 (Variance): 描述它的离散程度，变量离期望值的距离，每个样本值与 E 之差的平方值 d 的平均数
例：色子 2,3,3,4,4, $D = \frac{(2-3.5)^2+(3-3.5)^2+(3-3.5)^2+(4-3.5)^2+(4-3.5)^2}{5}$

$$EX = \begin{cases} \sum p_i x_i & \text{离散函数} \\ \int_{-\infty}^{\infty} x f(x) d(x) & \text{连续函数} \end{cases}$$

$$E(X-EX)^2 = DX = Var(x) = E(X^2) - (EX)^2 = \begin{cases} \sum (x_i - EX)^2 P_i & \text{离散函数} \\ \int_{-\infty}^{\infty} (x_i - EX)^2 f(x) d(x) & \text{连续函数} \end{cases}$$

期望值 E 是线性函数
$E(aX + bY + C) = aE(X) + bE(Y) + C$
$E(XY) = EX \cdot EY$ (XY 的协方差为 0(此时称不相关，两个随机变量独立是一种情况))

[1] https://oscarbonilla.com/2009/05/visualizing-bayes-theorem/

$E(E(X)) = E(X)$
$E(X - E(X)) = 0$

$D(C) = 0$
$D(aX) = a^2DX$
$D(aX + C) = a^2DX$
$D(X \pm Y) = DX + DY \pm 2E\{(X - EX)(Y - EY)\}$

1.6 常见分布

1.6.1 0-1、Bernulli Trial

同条件下重复、相互独立的试验，只有两种结果，发生或者不发生，如掷硬币
$E = \mu = 1 \cdot P + 0 \cdot (1 - P) = P$
$D = \sigma^2 = (1 - p)^2p + (0 - p)^2(1 - p) = p(1 - p)$

1.6.2 二项分布

N 次独立的 Bernulli Trial

$P\{X = k\} = C_n^k p^k(1 - p)^{n-k}$
$E = \sum_{i=1}^{n} \mu = np$
$D = \sum_{i=1}^{n} \sigma^2 = np(1 - p)$

1.6.3 泊松分布

$P_k = \frac{\lambda^k}{k!}e^{-\lambda}$
$E = D = \lambda$

1.6.4 指数分布

$P(x) = \lambda e^{-\lambda x}, x \geq 0$
$E = 1/\lambda$
$D = 1/\lambda^2$

1.6.5 均匀分布

$f(x) = \frac{1}{b-a}, F(x) = \frac{x}{b-a}$
$E = \int_a^b xf(x)dx = \frac{a+b}{2}$

$D = E(X^2) - (EX)^2 = \frac{(a+b)^2}{2}$, if you use $D = \int_a^b (x-EX)^2 f(x)dx$ 很繁琐

1.6.6 正态分布、高斯分布

众数 = 中位数 =E=μ

标准正态分布 $\mu = 0$ $\sigma^2 = 1$

1.7 Examples

A 和 B 独立,$P(A \cup B) = 0.8, P(A) = 0.4, P(\bar{B}|A) = ?, hint$[2]

共 50 球，20 黄球，30 白球，两人依次取出不放回，求第二人取黄球的 P，求如第二人黄球第一人黄球的 P, 思考第一人黄球第二人也是黄球的 P, hint[3]

参加考试考生中，本专业学生占 6 成，本专业通过率是 85%，非本专业通过率是 50%。某位考生通过了考试，他是本专业的 P 是多少, hint[4]

10 个产品, 4 次品, 任取两个 (第一次第二次), 至少一个正品的 P 是多少, hint[5]

[2] $P(A \cup B) = P(A) + P(B) - P(A)P(B)$ answer is 1/3

[3] Tree diagram, final answer is 2/5, 19/49, 19/49

[4] 重点是怎么画 tree, tree 的各节点是同性质的吗??，考生 {考生-本专业 {考生-通过、考生-未通过}，考生-非本专业 {考生-通过、考生-未通过}}, answer is 51/71

[5] 1. 1-$\frac{C_4^2}{C_{10}^2}$ 2. Tree diagram