概率论月考复习

容斥恒等式

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) =$$
$$\sum_{i=1}^{n} P(E_i) - \sum\sum P(E_iE_j) + \sum\sum\sum P(E_iE_jE_l) - \cdots + (-1)^{n+1}P(E_1\cdots E_n)$$

例1： 10对夫妇坐成一圈，计算所有的妻子都不坐在她丈夫旁边的概率。

$$n\text{个人坐成一圈，全排列有}(n-1)!\text{种情况。}$$
$$\text{记}E_i:\text{第}i\text{对夫妇坐在一起。}$$
$$\text{所求即为}:1-P(\bigcup_{i=1}^{10} E_i)$$
$$P(\bigcup_{i=1}^{10} E_i) = \sum_{i=1}^{10} P(E_i) - \sum\sum P(E_iE_j)\cdots$$

- **总结：**

  - 对于n个事件的并，考虑用容斥恒等式转化。

组合数性质

例1：$(2x+y+3z)^5$的展开项有多少项（terms）？x^2y^2z的系数（coefficient）是多少？

$$\text{对于}(a_1x_1+\cdots a_kx_k)^n$$
$$\text{项数为}:\binom{n+k-1}{k-1}$$
$$\text{其中}n\text{为多项式的指数，}k\text{为变量的个数。}$$
$$\text{每一项前面的系数为}\binom{n}{t_1,\cdots,t_k}a_1^{t_1}\cdots a_k^{t_k} = \frac{n!}{t_1!t_2!\cdots t_k!}a_1^{t_1}\cdots a_k^{t_k}$$

连续集函数

定义（递增递减序列）： 事件序列$\{E_n, n \geq 1\}$如果满足:

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

则称为递增序列。定义一个新的事件，记为$\lim\limits_{n\to\infty} E_n$，如下:

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

反之，如果满足:

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

则称为递减序列。定义一个新的事件，记为$\lim\limits_{n\to\infty} E_n$，如下:

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

命题： 事件序列$\{E_n, n \geq 1\}$如果是递增或者递减序列，则:

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

例 6a 概率与悖论 假设有个无限大的坛子以及无限个编了号码 1，2，3，…的球，考虑以下的试验：

在差 1 分到 12 P. M. 的时候，将 1 到 10 号球放进坛子，并把 10 号球拿出来(假设放球和拿球的时间忽略不计)；

在差 1/2 分到 12 P. M. 的时候，将 11 到 20 号球放进坛子，并把 20 号球拿出来；

在差 1/4 分到 12 P. M. 的时候，将 21 号到 30 号球放进坛子，并把 30 号球拿出来；

在差 1/8 分到 12 P. M. 的时候，……

等等.

问题：在 12 P. M. 的时候，坛子里有多少球？

假如现在随机取走一个球，求证：坛子里没有球的概率为1。

$$\text{考虑1号球，在差一分没被抽走的概率为}: \frac{9}{10}$$

$$\text{在差}\frac{1}{2}\text{分没被抽走的概率为}: \frac{9}{10} \times \frac{18}{19}$$

$$\text{在差}\frac{1}{4}\text{分没被抽走的概率为}: \frac{9}{10} \times \frac{18}{19} \times \frac{27}{28}$$

$$\text{在差}\frac{1}{2^n}\text{分没被抽走的概率为}: \prod_{i=1}^{n} \frac{9i}{9i+1}$$

$$\text{显然是一个递减序列，} \lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty}(E_n)) = \prod_{n=1}^{\infty} \frac{9n}{9n+1}$$

$$\text{下证} \prod_{n=1}^{\infty} \frac{9n+1}{9n} = +\infty$$

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{9n}\right) = \sum_{n=1}^{\infty} \ln\left(1 + \frac{1}{9n}\right) \sim \sum_{n=1}^{\infty} \frac{1}{9n} \to \infty$$

同理，每个球的概率都是0。

- **总结:**

  - **概率的连续性质:交换极限顺序。**

乘法规则（Multiplication rule）

公式:

$$P\left(\bigcap_{i=1}^{n} E_i\right) = P(E_1)P(E_2|E_1)\cdots P(E_{n-1}|E_n)$$

这给我们求几个事件的交集提供了方法。

例1: 坛子内有8个红球，4个白球，假定球的质量不同,每个红球的质量为r,每个白球的质量为w,假设每次抽到一给定球的概率是其质量除以当时坛子中球的总质量,则两次取出的都为红球的概率是多少?

$$\text{记}A\text{：第一次抽到红球。}B\text{：第二次抽到红球。考虑乘法公式：}$$
$$P(AB) = P(A)P(B|A) = \frac{8r}{8r+4w} \times \frac{7r}{7r+4w}$$

例2: 一副52张牌随机地分成4堆,每堆13张. 计算每一堆正好有一张A的概率。

$$\text{记}E_i(i=1,2,3,4)\text{为四个事件，其中：}$$
$$E_1\text{：黑桃}A\text{在任意一堆里。}E_2\text{：红桃}A\text{和黑桃}A\text{在不同堆里。}$$
$$E_3\text{：方片}A\text{和红桃，黑桃不在一堆里。}E_4\text{：四个}A\text{在不同堆里。}$$
$$P(E_1) = 1,\ P(E_2|E_1) = 1 - \frac{12}{51} = \frac{39}{51},\ P(E_3|E_1E_2) = 1 - \frac{24}{50} = \frac{26}{50}$$
$$P(E_1E_2E_3E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3) = \frac{39\times 26\times 13}{51\times 50\times 49} \approx 0.105$$

贝叶斯公式（Bayes's formula）

全概率公式:

$$P(E)=\sum_{i=1}^{n}P(E|F_i)P(F_i)$$

综合全概率公式，可以得到贝叶斯公式为：

$$P(E_j|F)=\frac{P(E_jF)}{P(F)}=\frac{P(E_j|F)P(F)}{\sum\limits_{i=1}^{n}P(F|E_i)P(E_i)}$$

其中E_i表示一组互不相容且穷尽的事件，即$\bigcap\limits_{i=1}^{n}E_i=\varnothing$，$P\left(\bigcup\limits_{i=1}^{n}E_i\right)=1$

全概率公式在条件概率下的写法：

$$P(E_1|F)=P(E_1|E_2F)P(E_2|F)+P(E_1|E_2^cF)P(E_2^c|F)$$

推广到若干事件，即：

$$P(E_1|F)=\sum_{i=1}^{n}P(E_1|F_iF)P(F_i|F)$$

定义：优势比（odds）：

$$\frac{P(F)}{P(F^c)}=\frac{P(F)}{1-P(F)}$$

在给定条件E下，优势比变为：

$$\frac{P(F|E)}{P(F^c|E)}=\frac{P(F)}{P(F^c)}\frac{P(E|F)}{P(E|F^c)}$$

这就是原来的优势比乘以新的证据在F，F^c下的概率值。

$$\text{记}A:\text{此人确实患该疾病。}B:\text{某人化验结果为阳性。}$$
$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}=\frac{0.95\times0.5\%}{0.95\times0.5\%+0.01\times99.5\%}$$

例1： 假设某药剂师考虑如下诊断方案:如果我至少有80%的可能确定病人确实有此病,那么我会建议手术;而如果我并不确定,那么我会推荐做进一步的检查,该检查是昂贵的,有时也是痛苦的. 现在,开始我仅仅有60%的把握认为琼斯患有此病,因此我推荐做了检查项目A,该检查对于确有此病的患者给出阳性结果,而对健康人却不会给出阳性结果. 经检查琼斯的结果是阳性后,正当我建议手术时,琼斯给了我另一个信息,他患有糖尿病. 这个信息使问题复杂化,尽管它并不影响我一开始认为他患有此病的60%的把握,但是却影响了检查项目A的效果. 因为虽然该检查项目对健康人不给出阳性,但是对于患有糖尿病却不患有这种疾病的人来说,有30%的可能给出阳性结果. 那么我现在该如何做?是做进一步检查,还是立即手术?

$$\text{记}A:\text{该病人确实患有此病。}B:\text{该人检测结果为阳性。}$$
$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}=\frac{1\times0.6}{1\times0.6+0.3\times0.4}=0.833$$
$$\text{立即手术。}$$

例 3k 一架飞机失踪了，推测它等可能地坠落在 3 个区域. 令 $1-\beta_i(i=1,2,3)$表示飞机事实上坠落在第 i 个区域，且被发现的概率(β_i 称为忽略概率，因为它表示忽略飞机的概率，通常由该区域的地理和环境条件决定). 已知对区域 1 的搜索没有发现飞机，求在此条件下，飞机坠落在第 $i(i=1,2,3)$个区域的条件概率.

记$E_i(i=1,2,3)$表示飞机坠落在第i个区域。F : 对第1个区域的搜索没有发现飞机。

$$P(E_1|F)=\frac{P(F|E_1)P(E_1)}{P(F|E_1)P(E_1)+P(F|E_2)P(E_2)+P(F|E_3)P(E_3)}=\frac{\frac{\beta_1}{3}}{\frac{\beta_1}{3}+\frac{1}{3}+\frac{1}{3}}=\frac{\beta_1}{\beta_1+2}$$

$$P(E_2|F)=P(E_3|F)=\frac{1}{\beta_1+2}$$

例 3l 假设有 3 张形状完全相同但颜色不同的卡片，第一张两面全是红色，第二张两面全是黑色，而第三张是一面红一面黑．将这 3 张卡片放在帽子里混合后，随机地取出 1 张放在地上．如果取出的卡片朝上的一面是红色的，那么另一面为黑色的概率是多少？

记RB,RR,BB分别为两面分别为红黑，红红，黑黑。F : 取出的卡片朝上的一面是红色

$$P(RB|F)=\frac{P(F|RB)P(RB)}{P(F|RB)P(RB)+P(F|RR)P(RR)+P(F|BB)P(BB)}=\frac{\frac{1}{2}\times\frac{1}{3}}{\frac{1}{2}\times\frac{1}{3}+\frac{1}{3}\times 1}=\frac{1}{3}$$

记$A_1: AJ$作案。$A_2: B$: 关系是准确的。

$$P(A_1|B)=\frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)}$$

例 3o 某罪犯在犯罪现场留下了一些 DNA，法医研究后注意到能够辨认的只有 5 对，而且每个无罪的人，与这 5 对相匹配的概率为 10^{-5}，律师认为罪犯就是该城镇 1 000 000 个居民之一．在过去 10 年内，该城镇有 10 000 人刑满释放．他们的 DNA 资料都记录在案，在检查这些 DNA 文档之前，律师认为这 10 000 个有犯罪前科的人犯此罪的概率为 α，而其余 990 000 个居民中的每个人犯此罪的概率为 β，其中 $\alpha=c\beta$.（即他认为最近 10 年内释放的有犯罪前科的人作案的可能性是其他人的 c 倍.）将 DNA 分析结果同这 10 000 个有犯罪前科的人的数据文档对比后，发现只有 AJ 琼斯的 DNA 符合．假设律师关于 α 和 β 之间的关系是准确的，AJ 作案的可能性有多大？

分析 : 首先，犯罪率总和为1，即$10000\alpha+990000\beta=1$

在这10000个人中只有AJ的DNA符合的情况下，AJ作案。

记E : 这10000个人中只有AJ的DNA符合。$F: AJ$作案。

$$P(F|E)=\frac{P(E|F)P(F)}{P(E|F)P(F)+P(EF^c)}$$

一方面，$P(F)=\alpha$，$P(E|F)=(1-10^{-5})^{9999}$，$P(F^c)=1-\alpha$

另一方面，$P(EF^c)=10^{-5}\times(1-10^{-5})^{9999}\times\boxed{(1-10000\alpha)}$(这10000个人里不能有人作案)

$$\therefore P(F|E)=\frac{1}{0.9+\frac{10^{-5}}{\alpha}}$$

记A : 前n次都是正面朝上。B : 第$n+1$次正面朝上。E_i : 是第i个硬币

$$P(B|A)=\sum_{i=0}^{k}P(B|E_iA)P(E_i|A)$$

$$P(B|E_iA)=\frac{i}{k},P(E_i|A)=\frac{P(E_iA)}{\sum_{i=0}^{k}P(AE_i)P(E_i)}=\frac{\left(\frac{i}{k}\right)^n}{\sum_{i=0}^{k}\left(\frac{i}{k}\right)^n\frac{1}{k+1}}$$

$$\therefore P(B|A)=\sum_{i=0}^{k}\frac{i}{k}\frac{\left(\frac{i}{k}\right)^n\frac{1}{k+1}}{\sum_{j=0}^{k}\left(\frac{j}{k}\right)^n\frac{1}{k+1}}=\frac{\sum_{i=0}^{k}\left(\frac{i}{k}\right)^{n+1}}{\sum_{j=0}^{k}\left(\frac{i}{k}\right)^n}$$

1. (15 points) Suppose there are two coins. Coin 1 lands on heads with probability $\frac{1}{2}$, coin 2 lands on heads with probability $\frac{1}{3}$. We randomly select one coin and repeatedly flip it. If the first flip results in heads, what is the conditional probability that there are exactly 3 heads in the first 5 flips? (Express your answer in the fraction form)

$$\text{记}A:\text{第一次是头向上。}B:5\text{次有}3\text{次头向上。}E_i(i=1,2)\text{为第}i\text{枚硬币}$$
$$P(B|A)=P(B|E_1A)P(E_1|A)+P(B|E_2A)P(E_2|A)$$
$$\text{首先，}P(B|E_1A)=\binom{4}{2}(\frac{1}{2})^4(\text{是第一枚硬币且第一次头向上})$$
$$P(B|E_2A)=\binom{4}{2}(\frac{1}{3})^2(\frac{2}{3})^2$$
$$P(E_1|A)=\frac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1)+P(A|E_2)P(E_2)}=\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{6}}=\frac{3}{5},P(E_2|A)=\frac{2}{5}$$
$$P(B|A)=\frac{3}{5}\times\binom{4}{2}(\frac{1}{2})^4+\frac{2}{5}\times\binom{4}{2}(\frac{1}{3})^2(\frac{2}{3})^2$$

- **总结:**

  - 可能出现的英语：**优势比(odds)**
  - 贝叶斯公式
  - 判断在什么事件发生的情况下，另一个事件发生的概率。

独立事件(independent)

定义： 如果满足:

$$P(EF)=P(E)P(F)$$
$$P(E|F)=P(E)$$
$$P(F|E)=P(F)$$

则称E,F互相独立。

此时,$E,F^c \quad E^c,F \quad E^c,F^c$也是互相独立的。

拓展： 对于多个事件独立：事件$E_1,E_2,\cdots,E_n$是独立的，如果对这些事件的任意子集$E_{1'},\cdots,E_{n'}$，都有：

$$P(E_{1'})P(E_{2'})\cdots P(E_{n'})=P(E_{1'}E_{2'}\cdots E_{n'})$$

定义（条件独立）： 如果

$$P(E_1|E_2F)=P(E_1|F)$$
$$\text{或者}P(E_1E_2|F)=P(E_1|F)P(E_2|F)$$

则称E_1和E_2关于F是条件独立的。

随机变量X的分布函数（cumulative distribution function， cdf)

定义：分布函数$F(b)$表示$P(X \le b)$。

<u>概率的连续性质：</u> 概率测度对事件的极限具有连续性。假设考虑一组事件:$b_1 \subseteq b_2 \cdots \subseteq b_n \to b$，则：

$$\lim_{n \to \infty} P\{X \le b_n\} = P\{X \le b\}$$

策略：

- $P(X \le b)$，$F(b)$。
- $P(X > a)$，$1 - F(a)$。
- $P(X < c)$，利用概率的连续性质$P(X < c) = F\left(\lim_{n \to +\infty}(c - \frac{1}{n})\right) = \lim_{n \to +\infty} F\left(c - \frac{1}{n}\right)$。
- $P(X = d)$，$F(d) - \lim_{n \to +\infty} F\left(d - \frac{1}{n}\right)$。
- $P(f < X \le g)$，$F(g) - F(f)$。

对于离散型的随机变量，在分界点的跳跃值分别等于X在该点的取值。

这也提供了一个解题思路：**画出$X \sim f(X)$图像**，间断点的概率值等于分界点的跳跃值，连续则该点概率为0.

总结：

- 可能出现的英文：
  - distribution function of X （随机变量X的分布函数）
- 定义法：$F(A) = P(X \le A)$
- 画图

均值和方差（expectations and variance)

均值定义：$E(X) = \sum_{x \in X(s)} xp(x)$

性质：

- $$E(f(x)) = \sum f(x)p(x)$$
-
- $$E(aX+b) = aE(X)+b$$
-

方差计算方式： $\sigma^2(X) = E(X^2) - (E(X))^2$

性质：

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$

常见的离散分布

离散均匀分布(discrete uniform distribution):

满足：$P(X_i) = \dfrac{1}{n}$，$i = 1, 2, \cdots$

- 步骤：X have a discrete uniform distribution U_n with parameter n
- $$E(X) = \frac{n+1}{2}，\mathrm{Var}(X) = \frac{n^2-1}{12}$$

伯努利分布：

$$E(X^k) = npE[(Y+1)^{k-1}]$$

其中X为参数为(n,p)的二项随机分布(X is a **Bernoulli random variable** with parameter n, p)，Y为参数为$(n-1,p)$的二项随机分布。

由此可以推出：

$$E(X) = np$$
$$\sigma^2(X) = np(1-p)$$

且当$k = (n+1)p$时，$P(X=k)$取最大值。

- **几何分布：**

$$P(X=n)=(1-p)^{n-1}p$$
$$E(X)=\frac{1}{p}$$
$$\operatorname{Var}(X)=\frac{1-p}{p^2}$$

- **负二项分布：**

设 $(S,\mathcal{A},\mathbb{P})$ 是一个概率空间。设 $E\in\mathcal{A}$ 是一个标记为成功的事件，其发生的概率为 p。进行独立的试验，直到累积获得 r 次成功。设 X 为所需的试验次数，则称 X 为参数为 (r,p) 的**负二项分布随机变量。**

特别地，几何随机变量实际上是参数为 $(1,p)$ 的负二项分布随机变量。

负二项分布随机变量的概率质量函数（pmf）为：

$$P(X=n)=\binom{n-1}{r-1}p^r(1-p)^{n-r},\quad n=r,r+1,\cdots$$

- $$E(X)=\frac{r}{p},\operatorname{Var}(X)=\frac{r(1-p)}{p^2}$$