概率与期望

By i207M

Graduated from SJZEZ 2020

Studying @THU

Powered by Marp

概率

在一次随机试验 E 中可能发生的不能再细分的结果被称为**单位事件**。在随机试验中可能发生的所有单位事件的集合称为**事件空间**，用 S 来表示。

也就是说，进行一次随机试验 E ，其结果一定符合 S 中的恰好一个元素，不可能是零个或多个。例如在一次掷骰子的随机试验中，如果用获得的点数来表示单位事件，那么一共可能出现 6 个单位事件，则事件空间可以表示为 $S = 1, 2, 3, 4, 5, 6$ 。

一个**随机事件**是事件空间 S 的子集，它由事件空间 S 中的单位元素构成，用大写字母 $A, B, C, \dots$ 表示。例如在掷两个骰子的随机试验中，设随机事件 A 为“获得的点数和大于 10 ”，则 A 可以由下面 3 个单位事件组成： $A = (5, 6), (6, 5), (6, 6)$ 。

某个随机事件的某个结果的可能性。如果让这个随机事件不断重复发生，频率趋近于概率。
可以将一个随机事件的所有结果划分成T个等可能基本事件，如果某个结果包括了其中t个基本事件，那么发生的概率为t/T。

一些公式

P(A并B)=P(A)+P(B)-P(A交B)

可以扩展为容斥原理。

条件概率

如果A发生了，那么B发生的概率，记作P(B|A)

考虑四种情况：
P(AB都发生)=a,P(AB都不发生)=b
P(只有A发生)=c,P(只有B发生)=d

那么P(B|A)=a/(a+c),即P(B|A)=P(AB)/P(A)

全概率公式

用分类讨论来处理概率问题时常用，其实很多人用过但是不知道这是个公式。

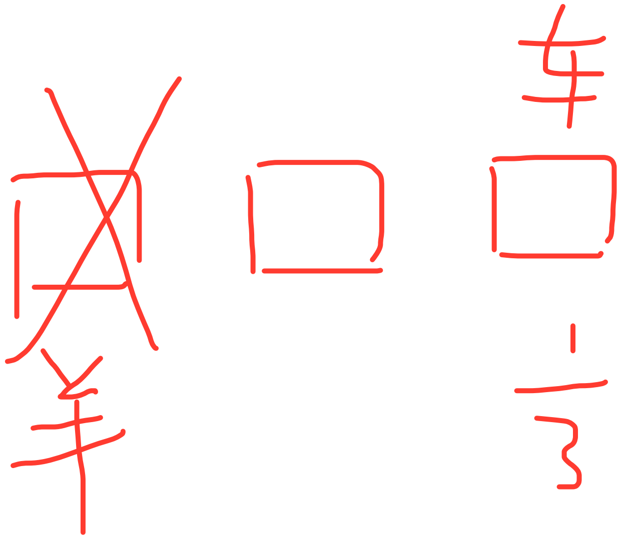
假如事件B1,B2,B3...Bn两两之间交集为空且它们的并集就是样本空间，那么对于另一个事件A:

P(A)=P(B1)*P(A|B1)+P(B2)*P(A|B2)...P(Bn)*P(A|Bn)

例如:已知某场OI比赛的第一题有30%概率出DP题，70%概率出数据结构题，DP题是难题的概率为50%，数据结构题为难题的概率为60%，问第一题是难题的概率。

P(是难题)=P(是DP题)*P(是难题|是DP题)+P(是数据结构题)*P(是难题|是数据结构题)

30 50 70 60



12.6 羊和车

有 3 个门，2 个后面是羊，1 个后面是车。

你可以选一个门，得到门后面的东西。但是你不知道门后面是什么。

这时你随机选了一个，然后主持人打开了一个门，后面是羊，问你改不改变自己的选择。

问题是改变是否可以提高门后面是车的概率？

另外

应该改变，不改变是车的概率是 $\frac{1}{3}$，改变是车的概率 $\frac{2}{3}$。

门后面是车概率提高了，如果不便于理解，可以想象有 $n = 10^9$ 个门，其中一个后面是车，然后主持人打开了 $n-2$ 个。

当然，改变的前提是你想要一个车而不是羊。

生日悖论

不考虑出生年份，问：一个房间中至少多少人，才能使其中两个人生日相同的概率达到 50% ?

解：假设一年有 n 天，屋子中有 k 人，用整数 $1, 2, ..., k$ 对这些人进行编号。假定每个人的生日均匀分布于 n 天之中，且两个人的生日相互独立。

设 k 个人生日互不相同为事件 A , 则事件 A 的概率为

$$P(A) = \frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-k+1}{n}$$

至少有两个人生日相同的概率为 $P(\overline{A}) = 1 - P(A)$ 。根据题意可知 $P(\overline{A}) \geq \frac{1}{2}$, 那么就有 $1 \times \frac{n-1}{n} \times \cdots \times \frac{n-k+1}{n} \leq \frac{1}{2}$

最后答案是23人。这十分有趣。

贝叶斯公式

由P(A|B)算P(B|A)时常用。

$P(B)*P(A|B)=P(AB)=P(A)*P(B|A)$

$P(B|A)=P(B)*P(A|B)/P(A)$ [这个叫贝叶斯公式]

$P(A|B)$

在人工智能决策方面有广泛应用。

例：已知某人咳嗽，则他感冒的概率？

A B

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

$$P(\text{癌} \mid +) = \frac{P(\text{癌}) \, P(+ \mid \text{癌})}{P(+)}$$

深入理解贝叶斯公式

推荐3b1b的视频

想象有1000人，其中1%患有癌症，他们都接受了筛查，筛查的灵敏度（1-假阴性率）为90%，特异度（1-假阳性率）为90%。若一个人的检测结果为阳性，则他患有癌症的概率约为?

$$10 \begin{cases} 9 & + \\ 1 & - \end{cases} \qquad 990 \begin{cases} 99 & + \\ 891 & - \end{cases}$$

$$\frac{9}{99+9} = 8.3\%$$

<1/10

正确的理解：

检测的结果不能说明一个人是否患有疾病；
检测的结果不能说明一个人患有疾病的概率；
检测的结果“更新”了患病的概率。

通过观测我们获得了更多的信息，也就由此改变了我们对概率的看法。

小\大	♀	♂
♀		
♂		

$\frac{2}{3}$

神奇的概率

某人家有两个孩子，已知其中一个是女生，问另一个是男生的概率？

结合概率的定义：

$$\frac{2}{3}$$

$$P(男|女日)=\frac{\frac{1}{2}\cancel{P(男)}\cdot\cancel{P(女日|男)}\quad\frac{\ }{49\times3}}{P(女日)}$$

其中

某人家有两个孩子，已知其中一个是女生，且生日是星期日，问另一个是男生的概率？

小\大	♀ 0 ~ 6	♂ 0 ~ 6
♀ 0 ~ 6		49
♂ 0 ~ 6	49	49

$$\frac{14}{27}\quad 27=14\times2-1$$

$$\frac{14}{27}$$

蛮神奇的

12.17 天选之人

为了简化题目，我们认为每个人的分数是一个 0 到 1 之间随机的实数。

某位同学考试完之后问其他同学的分数。

问了 100 个人，发现自己的分数都更高。

问自己比第 101 个人分数更高的概率？

$$P(101\text{高} \mid 100\text{高}) = \frac{P(101\text{高})}{P(100\text{高})} = \frac{101}{102}$$

答案

$(x^n)' = nx^{n-1}$

$\int x^n = \frac{1}{n+1} x^{n+1}$

$\left(\frac{1}{n+1} x^{n+1}\right)' = x^n$

$$\Pr[\text{比第 101 个人分高}|\text{比前 100 个人分高}] = \Pr[\text{比前 101 个人分高}] / \Pr[\text{比前 100 个人分高}]$$

$$\Pr[\text{比前 101 个人分高}] = \int_0^1 x^{101} \mathrm{d}x = \frac{1}{102}$$

全概率公式

$\sum_{x \in [0,1]} P(x) \cdot x$

$$\Pr[\text{比前 100 个人分高}] = \int_0^1 x^{100} \mathrm{d}x = \frac{1}{101}$$

$= \frac{1}{101} x^{101} \Big|_0^1 = \frac{1}{101}$

所以最终答案是 $\frac{101}{102}$。

这个结论可以推广，如果你问了 $p+q$ 个人，你比其中 p 个人高，比其中 q 个人低。

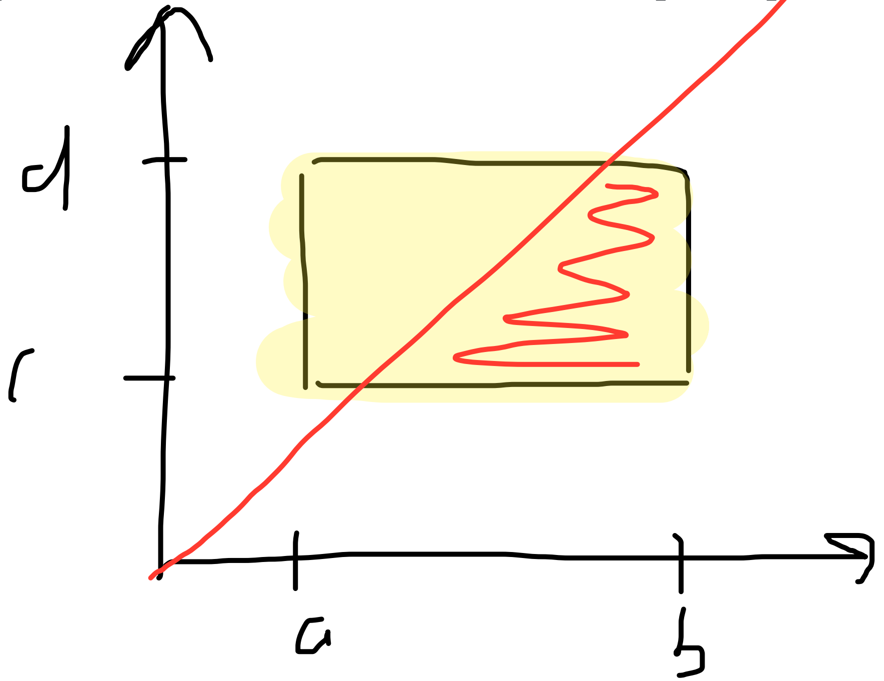
那么你比下一个人高的概率是 $\boxed{\frac{p+1}{p+q+2}}$。

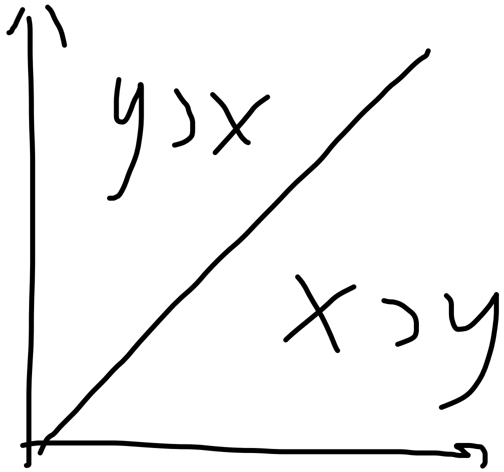
$= \int_0^1 x^{100} \cdot \frac{dx}{1-0}$

貌似有道省选题考的是这个，但是我找不着了。

给定两个变量，一个在$[a, b]$实数范围内随机，一个在$[c, d]$范围内随机，求$x > y$的概率

(x, y)





将两个变量转化为坐标，所有点形成一个矩形，发现题目就是求这个矩形在直线$y = x$上方的面积与矩形总面积的比值；

几何意义！！！

期望

随机变量X的均值E(X):

基本算法: $\sum P(x)x$

线性性(可加性):E(X+Y)=E(X)+E(Y),E(aX)=aE(X)

两个(或者多个)随机变量的和的期望等于期望的和。

只有两个事件独立时，乘积的期望=期望的乘积。

期望的平方不等于平方的期望，期望的max不等于max的期望。

期望的线性性常用来用来把一个大的期望拆成多部分去算。

例：给一张n个点m条边的图，第i个点有p_i的概率染黑，$1-p_i$的概率染白，问两端颜色不同的边的数目的期望?

y 50% 6 50% 12

x 60% 10 40% 5

$E(x)^2 = 64$

$10 \cdot 0.6 + 5 \cdot 0.4$

$E(x) = 8$

$100 \cdot 0.6 + 25 \cdot 0.4$

$E(x^2) = 70$

$E(X \cdot X)$

$E(XY) \neq E(X)E(Y)$

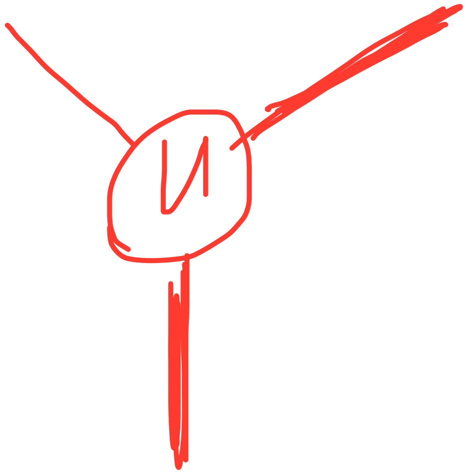
X Y Z

$\max(E(x), E(y), E(z)) \neq E(\max(x, y, z))$

$$E(\sum \text{color 不同}) = \sum E(\text{这条边 color 不同})$$

$$\sum p_u \cdot (1-p_v) + (1-p_u)\, p_v$$

期望的线性性，分别考虑每条边对期望的贡献，即每条边两端颜色不同的概率。



2^n

COGS1065 绿豆蛙的归宿

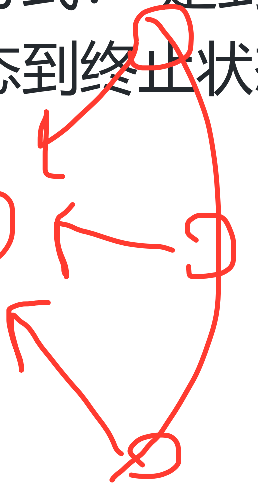
给出一个有向无环图，起点为1终点为N，每条边都有一个长度，并且从起点出发能够到达所有的点，所有的点也都能够到达终点。绿豆蛙从起点出发，走向终点。
到达每一个顶点时，如果有K条离开该点的道路，绿豆蛙可以选择任意一条道路离开该点，并且走向每条路的概率为 1/K 。
现在绿豆蛙想知道，从起点走到终点的所经过的路径总长度(也就是路径上的边数)期望是多少？

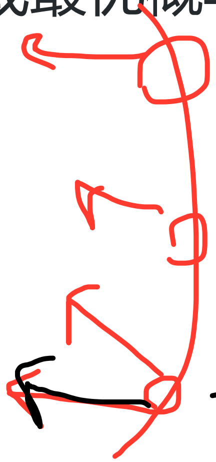
法一

定义f[i]表示当前位于点i，从这里出发走到终点的期望步数。按拓扑序递推。

注意这里的定义方式：“走到终点的期望步数”，其实很多概率题目都是类似的定义方式，定义某个状态到终止状态的期望步数或最优概率，这样往往比正着定义要方便。

$f(n)=0$





$f(i)=\frac{1}{2}(f(v_1)+f(v_2))$

$$u \longrightarrow v$$

$$E(edge) = \frac{f(u)}{deg(u)}$$

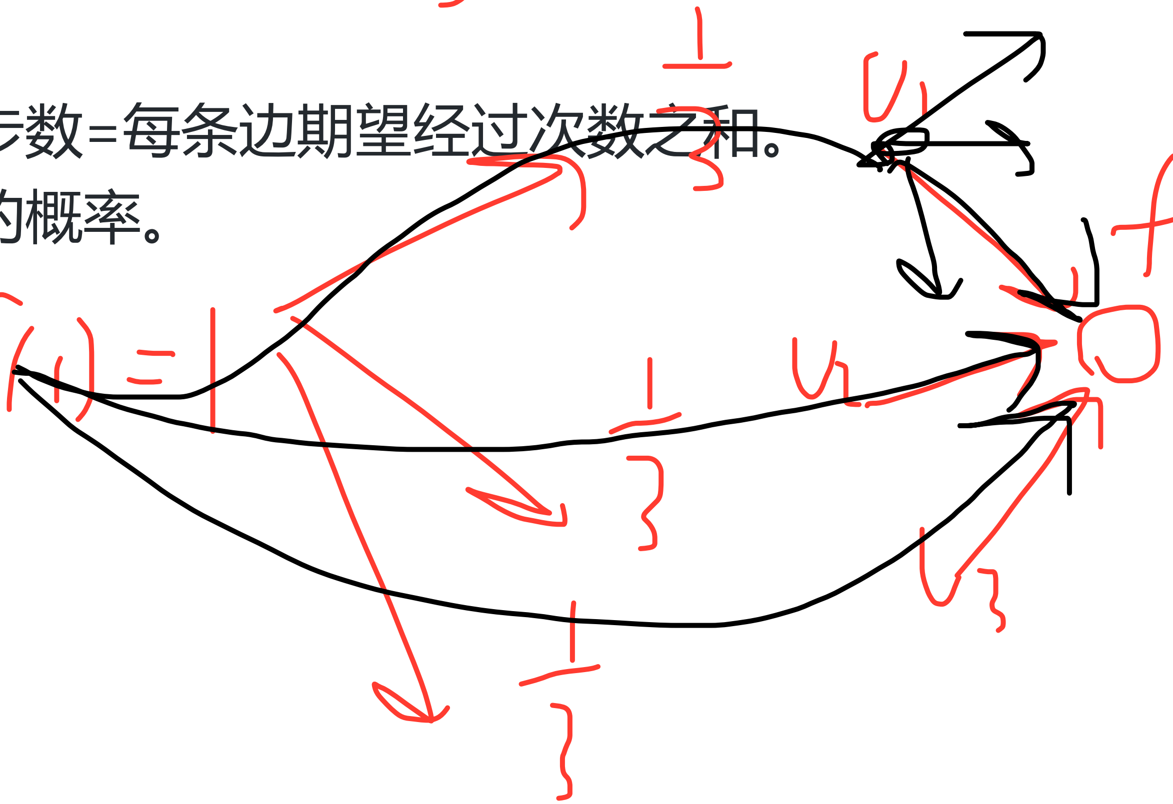
法二

期望的线性性，期望的步数=每条边期望经过次数之和。

也就是求每条边被经过的概率。

点

$f(1)=1$



$$f(u) = \sum_{v} \frac{f(v)}{deg(v)}$$

如果一条边从u指向v，要经过这条边就必须经过点u.如果经过了点u，就有1/(u的出度)的概率经过这条边。

对于点u,我们考虑它的k条入边(from[1]->u),(from[2]->u)...(from[k]->u)，应用全概率公式,经过点u的概率等于P(从(from[1]->u)到达u)+P(从(from[2]->u)到达u)...+P(从(from[k]->u)到达u)，即从每条边到达u的概率之和。按照拓扑序递推即可。

$$① \; E=\sum_i i\cdot P(i) = 1\times p + 2\times(1-p)\cdot p + \cdots$$

$$= p\left(1 + 2\cdot(1-p) + 3\cdot(1-p)^2 + \cdots\right)$$

神奇的期望

如果事件P发生的概率是p，则期望几次后发生?

$var = \sum (x_i-\bar{x})^2 / n$

方差的期望=平方的期望-期望的平方

$E(x)=\mu$

$$\sigma^2 = E[(X-\mu)^2] = E(X^2-2\mu X+\mu^2) = E(X^2)-2\mu E(x)+\mu^2 = E(X^2)-\mu^2$$

$E(X^2)-2\mu^2+\mu^2$

$P(\geq 1) \quad P(\geq 2) \quad P(\geq 3)$

$= 1 + (1-p) + (1-p)^2 \cdots (1-p)^n \cdots$

$= \dfrac{1-(1-p)^n}{1-(1-p)} \quad (1-p)^n \to 0 \quad = \dfrac{1}{p}$

$$② \; \sum i\cdot P(i) = \sum_i P(\geq i)$$

$$= \sum i\cdot\left(P(\geq i) - P(\geq(i+1))\right) = 1\times(P(\geq 1) - P(\geq 2)) + 2\times(P(\geq 2) - P(\geq 3)) + 3\times P(\geq 3)$$

12.10 生日问题三

如果认为每天出生概率相同。（虽然由于政策，季节，并不相同。）

并且认为任意两个人生日相互独立。

并且不考虑闰年。

你是一个良心老板，要雇人打工，你要雇佣 n 个人。

在一年中，对于一天来说，如果他是任意一个人的生日，那么这天所有人都放假。

作为一个良（黑）心老板，你当然想最大化人数乘以工作日，你希望知道期望的最大值是多大。

答案

注意期望是可以相加的，你只需要关注一天（比如第一天）有多少人能上班。
你并不需要以一年为周期考虑这个问题。
每个人会导致上班的概率乘以 $(1-\frac{1}{365})$。
n 个人上班的话，收益是

$$n(1-\frac{1}{365})^n$$

你可以通过打表，二分，求导，或者是解不等式来处理这个问题。
可以发现 $n=364$ 和 $n=365$ 同时是最大值。

12.19　计划生育

如果认为所有孩子性别随机且独立。

如果每个家庭都持续生育，直到出现一个男孩。

那么社会上的男女比例会是多少?

$$1 : 1$$

题

可能会混进去一些奇奇怪怪的题目。

一定要积极思考！

收集邮票

有n种不同的邮票，皮皮想收集所有种类的邮票。唯一的收集方法是到同学凡凡那里购买，每次只能买一张，并且买到的邮票究竟是n种邮票中的哪一种是等概率的，概率均为1/n。但是由于凡凡也很喜欢邮票，所以皮皮购买第k张邮票需要支付k元钱。

现在皮皮没有邮票，皮皮想知道自己得到所有种类的邮票需要花费的钱数目的期望。

$N \le 10^5$

f

$$cost = \frac{f(f+1)}{2} = \frac{1}{2}(f^2 + f)$$

$$\neq (E(f))^2$$

$f[n]=0$

设$f[i]$为当前有i种邮票，到结束的期望次数：

$$\frac{1}{\frac{n-i}{n}}$$

$$f[i]=f[i+1]+\frac{n}{n-i}$$

设$g[i]$ =当前价格为1，有i种邮票，到结束的期望花费：

$$g[i]=\frac{i}{n}\times(g[i]+f[i]+1)+\frac{n-i}{n}\times(g[i+1]+f[i+1]+1)$$

化简：

$g[i][j+1]=g[i][j]+f[i]$

$$g[i]=\frac{i}{n-i}\times f[i]+g[i+1]+f[i+1]+\frac{n}{n-i}$$

$g[i][j]$

价格

$\sum_{i=1}^{n} f(i)$

OSU

P 1 , 1-P 0

给定长度为n的01串。在这个串中连续的X个1可以贡献 X^3 的分数，这X个1不能被其他连续的1所包含（也就是极长的一串1）。

求期望收益? $N \leq 10^5$

直接考虑比较困难，我们一级一级考虑。

设$x1[i]$表示以i结尾的期望连续1的长度：$x1[i] = (x1[i-1]+1) \times p[i]$ +0.

平方的期望不等于期望的平方

$(x+1)^2 = x^2+2x+1$

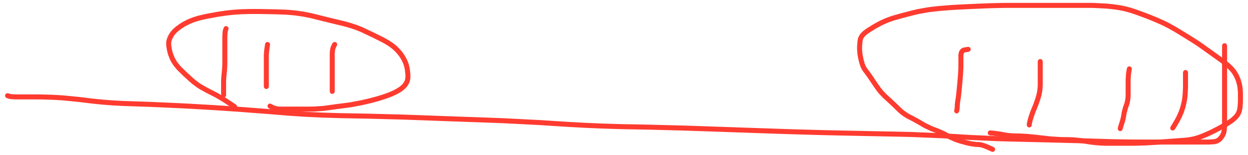
$E(x^2) \to E((x+1)^2)$

平方的期望：$x2[i] = (x2[i-1]+2 \times x1[i-1]+1) \times p[i]$

同理，立方的期望：$x3[i] = (x3[i-1]+3 \times x2[i-1]+3 \times x1[i-1]+1) \times p[i]$

系数就是二项式系数。可以理解为从一段中选择两/三个位置的方案数。

len^2 = 从$[1, len]$中有序选2个数的方案



注意答案不是$x3[n]$。$x3$表示以每个点结尾的长度的立方的期望，但并没有保证这就是这个1串的末端，所以答案会被重复统计，于是我们可以维护另外一个数组，$ans[i] = ans[i-1] + (3 \times x2[i-1] + 3 \times x1[i-1] + 1) \times p[i]$，表示这个点，相对上一个点，对答案的贡献期望增加了多少，最后答案就是ans[n].

why?

$$(x+1)^3 - x^3 = 3x^2 + 3x + 1$$

P4562 [JXOI2018]游戏

https://www.luogu.com.cn/problem/P4562

[l, r]

n个数 只有k个关键

最终对答案有贡献的只有“独立的数”，也就是它不能被$[l, r]$中的任何一个数整除；统计独立的数，可以埃拉托色尼筛法$O(N \log\log N)$；对于一个排列最末尾的关键数字所在的位置即是它的检查时间。

法一：

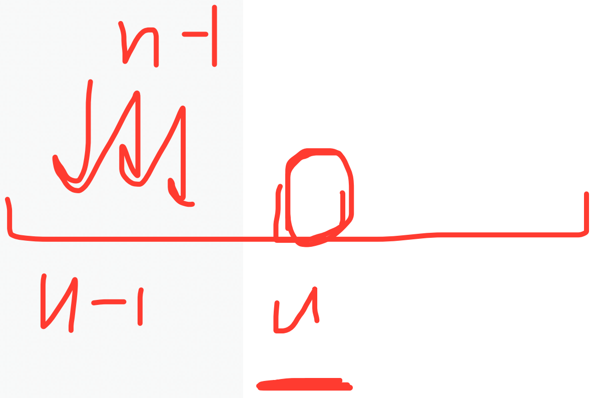
设独立的数的总数有n个，那么最后一个独立的数可以出现在$[n, r-l+1]$的任何一个位置u上，对答案的贡献就是u，方案数的话，首先得保证剩下的$n-1$独立的数要在前$u-1$个位置中，那么就有$n!(r-l-n+1)!\binom{u-1}{n-1}$种，这样一来

$$
\begin{aligned}
\text{ans}/[n!(r-l-n+1)!] &= \sum_{u=n}^{r-l+1} u\binom{u-1}{n-1} \quad ① \\
&= n\sum_{u=n}^{r-l+1}\binom{u}{n} \\
&= n\left(\sum_{u=0}^{r-l+1}\binom{u}{n} - \sum_{u=0}^{n-1}\binom{u}{n}\right) \quad ② \\
&= n\binom{r-l+2}{n+1} - n\binom{n}{n+1} \\
&= n\binom{r-l+2}{n+1} - 0 \\
&= \frac{n\cdot(r-l+2)!}{(n+1)!(r-l-n+1)!}
\end{aligned}
$$

$$\therefore \text{ans} = \frac{n(r-l+2)!}{n+1}$$

$$u\binom{u-1}{n-1} = n\binom{u}{n}$$

$$\frac{(u-1)!\cdot u}{(n-1)!(u-n)!} = \frac{n\cdot u!}{n!(u-n)!}$$

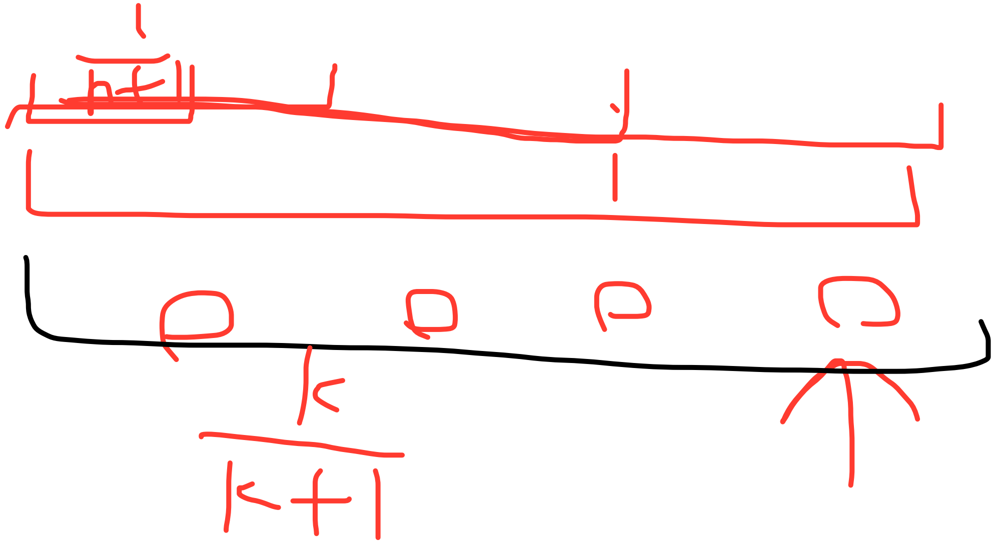


$$\frac{n}{n+1}$$

$$\sum_{i=0}^{m}\binom{i}{n} = \binom{m+1}{n+1}$$

n 球

k 个红球



法二：

问题转换为在n个数中随机取k个数，最末尾的数字的位置的**期望**乘以n的阶乘（即所有排列总数）。如果是期望，那么平均下来是平均分布,可以得出位置的期望为k/(k+1)×(n+1) (因为位置从1开始数起)，则总答案即为 $\frac{k}{k+1} \times (n+1)!$

直觉

再说一些

大家将来会做很多很多很多很多期望题，学习各种高端操作。

- 注意更换角度，巧妙转化，统计贡献
- 认真推式子（组合数学技巧）
- 巧设状态
- 询问期望步数，往往可以设“从i到i+1的期望步数”（即，差分的思想）（即，$P(X=i) \to P(X \le i)$）

Thank you

祝大家取得好成绩！

感谢 `OI-Wiki` 的信息

如果你比较强，可以尝试以下几个题目：

P3750 [六省联考2017]分手是祝愿

如果你很强，可以尝试以下几个题目：

[ZJOI2015]地震后的幻想乡，[SDOI2017]硬币游戏，