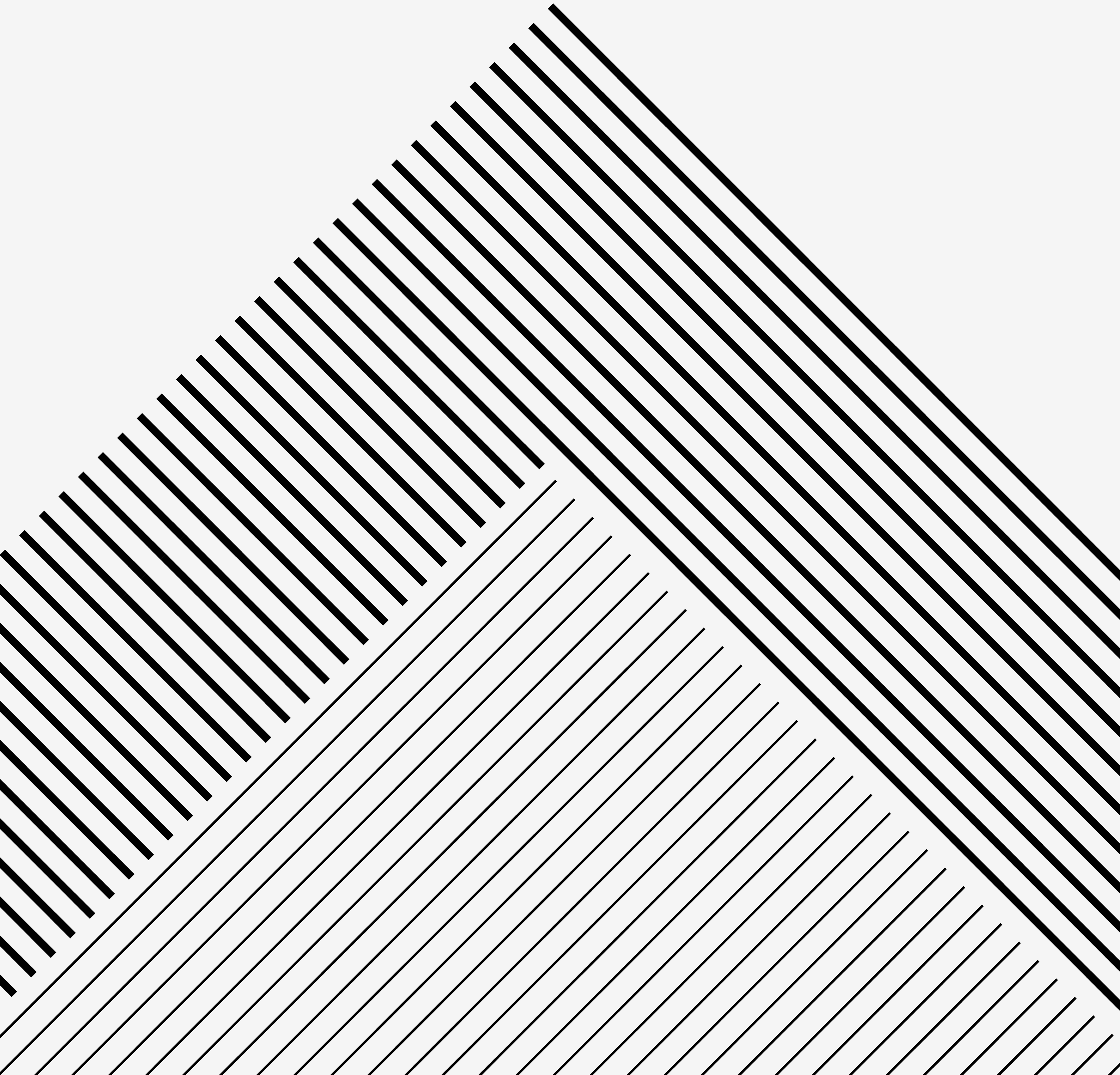


Part 一、随机事件与概率

§1.1 随机事件及其运算

一、随机现象

- 随机现象
- 确定性现象

二、样本空间

样本空间 $\Omega=\{\omega\}$. 随机现象的一切可能基本结果组成的集合. (至少有2个样本点)

样本点：随机试验基本结果.

Ω 包含所有特征，ω 为一个特征.

分类：
- 离散样本空间：样本点个数有限或可列个 → $\{1,2,3,4,\cdots\}$ 可列
- 连续样本空间：样本点个数无限或不可列个

三、随机事件

1. 随机现象的某些样本点组成的集合. 简称为事件.

   基本事件：事件只含一个试验结果.

   Ω 本身为必然事件. ϕ 为不可能事件.

2. 事件间关系及运算

① 包含：$\phi\subset A\subset\Omega$ $A=B \iff A\subset B$ 且 $B\subset A$

② 并：$A\cup B$ 交：$A\cap B\ (AB)$

差：$A-B=A-AB$

$\{X=a\}=\{X\le a\}-\{X<a\}$.

对立：$\overline{A}=\Omega-A$ $\overline{\overline{A}}=A$

$\overline{\Omega}=\phi$ $\overline{\phi}=\Omega$

A与B互相对立 $\iff AB=\phi$ 且 $A\cup B=\Omega$

互斥：$AB=\phi$

$$\boxed{\text{对立事件} \underset{\times}{\overset{\checkmark}{\rightleftarrows}} \text{互不相容事件}}$$

$A-B=A\cap\overline{B}$

⇓

A发生且B不发生

$AB=A-A\overline{B}$

AB都发生 = A发生 − A发生B不发生

③运算性质：交换律、结合律、分配律

对偶律： $\overline{A\cup B}=\bar{A}\cap\bar{B}$ $\overline{A\cap B}=\bar{A}\cup\bar{B}$

A或B对立 = A对立且B对立 A且B对立 = A对立或B对立

例 A、B、C都不发生： $\bar{A}\bar{B}\bar{C}=\overline{A\cup B\cup C}$

A、B、C不都发生： $\overline{ABC}=\bar{A}\cup\bar{B}\cup\bar{C}$

A、B、C至少有2个发生： $AB\bar{C}\cup A\bar{B}C\cup\bar{A}BC\cup ABC=AB\cup AC\cup BC$

ABC中仅A发生： $A\bar{B}\bar{C}=A\bar{B}-C=A-B-C=A(\overline{B\cup C})=A-(B\cup C)$

四、随机变量

用来表示随机现象结果的变量.

五、事件域

$\underset{\emptyset}{\underset{\downarrow}{C_n^0}}+C_n^1+\cdots+\underset{\Omega}{\underset{\downarrow}{C_n^n}}=2^n$ 集合的子集个数.

§1.2 概率的定义及其确定方法

一、排列组合

1. 排列：$A_n^r(P_n^r) = \underbrace{n\times(n-1)\times\cdots\times(n-r+1)}_{r个} = \frac{n!}{(n-r)!}$　　$r\le n$

$A_n^n = n!$

2. 重复排列：n^r　　r 可以大于 n

3. 组合：$C_n^r = \binom{n}{r} = \frac{P_n^r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$　　$r\le n$

$\binom{n}{r} = \binom{n}{n-r}$　　$P_n^r = r!\,C_n^r$

4. 重复组合：$\binom{n+r-1}{r}$　　r 可以大于 n

n个元素为n个盒子（n+1 根木棒）用 0表示取过几次.

|00|　|0|……|000|

不可以动　　不可以动

有r个0 (n+1)个木棒 ⇒ 在 n+r-1 个位置上放r个"0"或(n-1)个"|" ⇒ $\binom{n+r-1}{n-1} = \binom{n+r-1}{r}$

二、几个模型　古典概型

抽样模型

例1(抽样模型) 设 N 件产品中有 M 件是次品, $N-M$ 件是正品。现从 N 件中随机地不放回地抽取 n 件产品。求: 事件 A_m={所取的 n 件产品中恰有 m 件次品}的概率. ($m = 0,1,2,\ldots,n$)

$$P(A_m) = \frac{C_{N-M}^{n-m}\cdot C_M^m}{C_N^n}$$

有放回抽样

例2 (有放回抽样) 设 N 件产品中有 M 件是次品, $N-M$ 件是正品。现从 N 件中随机地有放回地抽取 n 件产品。求: 事件 B_m ={所取的 n 件产品中恰有 m 件次品}的概率.

$$P(B_m) = \frac{C_n^m M^m\cdot(N-M)^{n-m}}{N^n} = C_n^m\left(\frac{M}{N}\right)^m\left(1-\frac{M}{N}\right)^{n-m}$$

或者：$\frac{M}{N}$ 为p(次品)　$1-\frac{M}{N}$ 为p(正品)　要有 m次品 ⇒ $C_n^m\cdot\left(\frac{M}{N}\right)^m$

n-m 正品 ⇒ $\left(1-\frac{M}{N}\right)^{n-m}$

盒子模型

例3 （盒子模型）设有 n 个(不同)球, 每个球等可能地落入 N 个不同的盒子中（$n \le N$），
设每个盒子容球数不限, 求下列事件的概率:

（1）A =“指定的 n 个盒子中各有一球”；

（2）B =“恰有 n 个盒子中各有一球”.

(1) $P(A) = \dfrac{n!}{N^n}$

(2) $P(B) = \dfrac{C_N^n\, n!}{N^n} = \dfrac{N!\, n!}{(N-n)!\, n!\, N^n} = \dfrac{N!}{(N-n)!\, N^n}$

抽签问题

袋中有 a 只白球，b 只红球，它们除颜色不同外，其他方面没有差别，现在把球随机地一只只摸出来，作（1）放回取样，（2）不放回取样，求第 k 次摸出的球是白球的概率($k \le a+b$)。

(1) $P_k = \dfrac{(a+b)^{k-1}\cdot a}{(a+b)^k} = \dfrac{a}{a+b}$

(2) 样本空间包含样本点的总数为: P_{a+b}^k (排列)

第k个为白球的事件为: $C_a^1 \times P_{a+b-1}^{k-1}$

选1个白球　　剩下的排列

$$\Rightarrow P_k = \frac{a \times P_{a+b-1}^{k-1}}{P_{a+b}^k} = \frac{a\times(a+b-1)\cdots(a+b-k+1)}{(a+b)\cdots(a+b-k+1)} = \frac{a}{a+b}$$

三. 概率的几何意义

几何概型

$$P(A) = \frac{G\text{的测度}}{\Omega\text{的测度}}$$

会面问题　比丰投针问题

必然事件的概率为1, 但概率为1的事件不一定是必然事件.

比如向[0,1]投点, A:点在(0,1)上. $P(A)=1$, 但A未必发生.

四. 概率的公理化定义

公理1 (非负性)　$\forall A \in F$, $P(A) \ge 0$;

公理2 (规范性)　$P(\Omega) = 1$;

公理3 (可列可加性)　$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, 其中 $A_1, A_2, \cdots$ 为两两互不相容事件.

§1.3 概率的性质

一、概率的其他性质

1、$P(\phi)=0$

2、有限可加性：若有限个事件 $A_1, A_2, \cdots, A_n$ 互不相容，则有：

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

推论：$P(\bar{A}) = 1 - P(A)$

3. $\boxed{\text{可减性}}$：对任意事件 A, B 有：$P(A-B) = P(A) - P(AB)$ ☆

二、概率加法公式

对任意事件有：

2个：$P(A\cup B) = P(A) + P(B) - P(AB)$

3个：$P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

n个：$P(A_1\cup A_2\cup\cdots\cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i\le j\le n} P(A_iA_j) + \sum_{1\le i\le j\le k\le n} P(A_iA_jA_k) + \cdots + (-1)^{n} P(A_1A_2\cdots A_n)$

匹配问题：**例（匹配问题）某人写好 n 封信，又写好 n 只信封，然后在黑暗中把每封信放入一只信封中，试求至少有一封信放对的概率。**

记 A_k 表示第 k 封信放对了.

$P = P(A_1\cup A_2\cup\cdots\cup A_n) = \sum P(A_i) - \sum P(A_iA_j) + \sum(A_iA_jA_k) + \cdots + (-1)^{n-1}\sum(A_1A_2\cdots A_n)$

$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$

$P(A_iA_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

$P = \binom{n}{1}\frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \binom{n}{3}\frac{1}{n(n-1)(n-2)} + \cdots + (-1)^{n-1}\frac{1}{n!}$

$= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}$

三、概率的连续性

§1.4 条件概率

一、条件概率

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(B|A)P(A)}{P(B)} \qquad P(B)>0$$

given

性质：$P(\bar{A}|B)=1-P(A|B)$

$P((A_1\cup A_2)|B)=P(A_1|B)+P(A_2|B)-P(A_1A_2|B)$

注意：$P(A|B)+P(A|\bar{B})\neq 1$.

二、乘法公式

n个事件发生有明显的先后顺序

$$\begin{cases}若P(B)>0，则：P(AB)=P(B)P(A|B)\\ 若P(A)>0，则：P(AB)=P(A)P(A|B)\end{cases}$$

若$P(A_1A_2\cdots A_{n-1})>0$，则 $P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$

波利亚罐模型

例(波利亚罐模型). 设罐中装有b只黑球，r只红球，每次自袋中任取一只球，观察其颜色然后放回，并再放入c只同色球和d只异色球。若在袋中连续取球三次，试求第一、二次取到红球且第三次取到黑球的概率。

记B_i为第i次取出黑球，R_i为第i次取出红球.

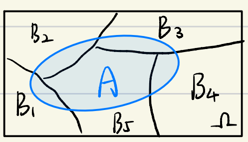
$$P(R_1R_2B_3)=P(R_1)P(R_2|R_1)P(B_3|R_1R_2)$$
$$=\frac{r}{b+r}\cdot\frac{r+c}{b+r+c+d}\cdot\frac{b+2d}{b+r+2c+2d}.$$

$$\begin{cases}c=-1,\ d=0：不返回抽样\\ c=0,\ d=0：返回抽样\\ c>0,\ d=0：传染病模型\\ c=0,\ d>0：安全模型\end{cases}$$

三、全概率公式

1.完备事件组：若$B_1B_2\cdots B_n$互不相容，且$W=\bigcup_{i=1}^{n}B_i$，则称$B_1B_2\cdots B_n$为一个完备事件组，或称$B_1B_2\cdots B_n$为Ω的一个分割.

n可以为∞.



2. 设 $B_1 B_2 \cdots B_n$ 为样本空间 Ω 的一个分割，如果 $P(B_i)>0$，$i=1,2,\cdots,n$，则对任一事件 A 有：

$$P(A)=\sum_{i=1}^{n} P(B_i)P(A|B_i)$$

最简形式：若 $0<P(B)<1$，则 $P(A)=P(B)P(A|B)+P(\bar{B})P(A|\bar{B})$.

摸彩模型　设在 n 张彩票中有一张可中奖，求第二人抽到中奖彩票的概率.

设 A_i 表示第 i 个人抽到中奖彩票.

$P(A_1)=\frac{1}{n}$　　$P(\bar{A_1})=\frac{n-1}{n}$

$P(A_2)=P(A_1)P(A_2|A_1)+P(\bar{A_1})P(A_2|\bar{A_1})=\frac{1}{n}\times 0+\frac{n-1}{n}\times\frac{1}{n-1}=\frac{1}{n}$

类推：$P(A_3)=P(A_4)=\cdots P(n)=\frac{1}{n}$

若有 k 张中奖，则 $P(A_1)=P(A_2)=\cdots=P(n)=\frac{k}{n}$.

机会均等.

敏感性问题调查

例（敏感问题调查）在调查家庭暴力所占家庭的比例 p 时，被调查者往往不愿回答真相，这使得调查数据失真。

为得到实际的 p 同时又不侵犯个人隐私，调查人员将袋中放入比例是 p_0 的红球和比例是 $q_0=1-p_0$ 的白球。被调查者在袋中任取一球窥视后放回，并承诺取得红球就讲真话，取到白球就讲假话。

被调查者只需在匿名调查表中选“是”或“否”，然后将表放入投票箱。没人知道被调查者是否讲真话和回答的是什么。

如果声称有家庭暴力的家庭比例是 p_1，求 p。

设 B 表示声称为是，A 为存在家庭暴力.

则：$P(B)=p_1$，　$P(B|A)=p_0$，　$P(B|\bar{A})=q_0=1-p_0$

$P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})$

$p_1=P(A)p_0+(1-P(A))q_0$

$\Rightarrow p=P(A)=\frac{p_1-q_0}{p_0-q_0}$　$(p_0\neq q_0)$

传球问题

例　m 个人相互传球，球从甲手中传出，每次传球时，传球者等可能地把球传给其余 m-1 个人中的任何一个，求第 n 次传球时仍由甲传出的概率。

第 n 次是否由甲传出 和第“n-1”次谁传有关系.

设 A_i 为第 i 次由甲传出 $\Rightarrow P(A_n)=P(A_{n-1})\underbrace{P(A_n|A_{n-1})}_{0}+P(\overline{A_{n-1}})\underbrace{P(A_n|\overline{A_{n-1}})}_{\frac{1}{m-1}}$

$\Rightarrow P(A_n) = \frac{1}{m-1} P(\overline{A_{n-1}})$

$\Rightarrow P(A_n) = \frac{1}{m-1}(1-P(A_{n-1}))$ 一阶线性递推数列

$P(A_n)-M = N[P(A_{n-1})-M]$

$P(A_n) = NP(A_{n-1}) + M - MN$

对比系数: $N = -\frac{1}{m-1}$, $M = \frac{1}{m}$

$P(A_n)-\frac{1}{m} = \frac{1}{1-m}(P(A_{n-1})-\frac{1}{m})$ $\quad P(A_1)=1$.

令 $P(A_n)-\frac{1}{m} = b_n$.

$b_n = \frac{1}{1-m}\cdot b_{n-1} = \left(\frac{1}{1-m}\right)^2 \cdot b_{n-2} = \cdots = \left(\frac{1}{1-m}\right)^{n-1}\cdot b_1$

$= \left(\frac{1}{1-m}\right)^{n-1}\cdot\left(1-\frac{1}{m}\right) = \left(\frac{-1}{m-1}\right)^{n-2} \times \frac{-1}{m}$

$\Rightarrow P(A_n) = b_n + \frac{1}{m} = \frac{1}{m}\left[1-\left(\frac{-1}{m-1}\right)^{n-2}\right]$, $n=2,3,4,\cdots$

赌徒破产模型

例 (赌徒破产模型) 甲有本金 a 元，决心再赢 b 元停止赌博。设甲每局赢的概率是1/2，每局输赢都是一元钱，甲输光后停止赌博，求甲输光的概率 $q(a)$。

A表示第一局赢. B_k表示本金为 k元时最后输光.

有: $q(0)=1$, $q(a+b)=1$.

以第一局输赢情况分割.

$q(k) = P(B_k) = P(A)P(B_k|A) + P(\bar{A})P(B_k|\bar{A})$

$\downarrow \frac{1}{2} \quad \downarrow P(B_{k+1}) \quad \downarrow \frac{1}{2} \quad \downarrow P(B_{k-1})$

$P(B_k|A) = P(B_{k+1})$ "第一局赢了,本金就多了一块,输光就要成了 $k+1$"

$\Downarrow$

$P(B_k) = \frac{1}{2}P(B_{k+1}) + \frac{1}{2}P(B_{k-1})$ 二阶线性递推数列.

$x^2-2x+1=0 \quad (x-1)^2=0 \quad x_1=x_2=1$.

$P(B_k) = (C+Dk)$

而 $P(0)=1$, $P(a+b)=0$ $\Rightarrow C=1$, $D=-\frac{1}{a+b}$

$\Rightarrow P(B_k) = 1-\frac{k}{a+b}$

$\Rightarrow q(a) = 1-\frac{a}{a+b} = \frac{b}{a+b}$

(当 $b\to\infty$, $q(a)\to 1$)

四. 贝叶斯(Bayes)公式

设 $B_1, B_2, \cdots, B_n$ 是样本空间 Ω 的一个分割. 如果 $P(A)>0$, $P(B_i)>0$, $i=1,2,\cdots,n$, 则:

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)} \quad i=1,2,\cdots,n$$

例

例. 盒中装有 8 个乒乓球，其中有 6 个新的。第一次练习时，从中任取 2 个来用，用完后放回盒中。第二次练习时，再从盒中任取 2 个。

(1) 求第二次取出的球都是新球的概率；

(2) 求在第二次取出的球都是新球条件下，第一次取到的球都是新球的概率。

"新球用后成旧球"

(1) 利用古典概型:

记 B 为第二次取出的球是 2 个新球.

$$P(B)=\frac{C_6^2C_4^2}{C_8^2C_8^2}+\frac{C_6^1C_2^1C_5^2}{C_8^2C_8^2}+\frac{C_2^2C_6^2}{C_8^2C_8^2}=\frac{225}{784}$$

(2) 记 A 为第一次取出的球是 2 个新球.

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{C_6^2C_4^2}{C_8^2C_8^2}\times\frac{784}{225}=\frac{2}{5}.$$

例

例. 在肝癌普查中发现，某地区的自然人群中，每十万人内平均有40人患原发性肝癌，有34人甲胎球蛋白高含量，有32人既患原发性肝癌又出现甲胎球蛋白高含量。从这个地区的居民中任抽一人，若他患有原发性肝癌则记为*C*，甲胎球蛋白高含量记为*D*，这时

$P(C)=0.0004$，$P(D)=0.00034$，$P(CD)=0.00032$ $\Rightarrow P(C|D)=\frac{P(CD)}{P(D)}=0.9412$

例.(续) 假定用血清甲胎球蛋白法诊断肝癌，

$$P(A\mid C)=0.95,\ P(\overline{A}\mid\overline{C})=0.90,$$

若以 *A* 表示事件"诊断结果被检验者患有肝癌"，以 *C* 表示事件"被诊断者患有肝癌"，现在对自然人群进行普查，试求 $P(C|A)$.

$$P(C|A)=\frac{P(AC)}{P(A)}=\frac{P(A|C)P(C)}{P(C)P(A|C)+P(\overline{C})P(A|\overline{C})}=\frac{0.95\times0.0004}{0.95\times0.0004+0.1\times0.9996}=0.0038$$

若先检验了甲胎蛋白后再求 $P(C|A)$ 呢? 将 $P(C)$ 修正成 $P(C|D)$

§1.5 独立性

一、两个事件的独立性

1. 若两事件 A、B满足：$P(AB)=P(A)P(B)$，则称A与B相互独立(A、B独立)。

2. 若 $P(A)>0$，A、B独立 $\Longleftrightarrow$ $P(B)=P(B|A)$

   若 $P(B)>0$，A、B独立 $\Longleftrightarrow$ $P(A)=P(A|B)$

3. AB独立不能由Venn表示.

   若A、B互不相容，且 $P(A)>0$，$P(B)>0$，则A与B不独立；

   若A、B独立，且 $P(A)>0$ $P(B)>0$，则A与B并不互不相容.

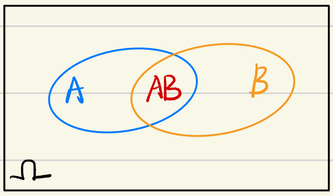
   对于正概率事件，互斥 ⇒ 不独立.

4. 四对事件：A,B；$\bar{A},B$；$A,\bar{B}$；$\bar{A},\bar{B}$ 任何一对独立，其它三对也独立.

   若 $A,\bar{B}$ 独立，证 A,B 独立.

   → $A\cap\overline{AB}=A\overline{B}$　　$P(A-B)=P(A)-P(AB)$

$$
\begin{aligned}
P(AB)&=P(A-A\bar{B})=P(A)-P(A\bar{B})\\
&=P(A)-P(A)P(\bar{B})\\
&=P(A)-P(A)[1-P(B)]\\
&=P(A)P(B)
\end{aligned}
$$



二、多个事件的相互独立性

1. 设A,B,C是三个事件，如果有：

$$
\begin{cases}
P(AB)=P(A)P(B)\\
P(AC)=P(A)P(C) \qquad (1)\\
P(BC)=P(B)P(C)
\end{cases}
$$

则A、B、C 两两独立.

若还有：$P(ABC)=P(A)P(B)P(C)$　　(2)

则A、B、C相互独立.

相互独立 $\xrightarrow{\checkmark}$ 两两独立；两两独立 $\xrightarrow{\times}$ 相互独立

n个事件相互独立：

$$
\begin{cases}
P(A_iA_j)=P(A_i)P(A_j),\\
P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k),\\
\cdots\cdots\\
P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)
\end{cases}
$$

$1\le i<j<k<\cdots\le n.$

2. 几个结论:

(1) n个事件独立的充要条件是任意取出其中的k个事件都是相互独立的.

(2) 若$A_1 A_2 \cdots A_n$独立,则将他们任意分成l组,每一组中的事件经过任何运算产生一个新的事件,这样的l个新的事件相互独立.

(3) 利用独立事件的性质计算其和事件的概率:

若$A_1, A_2, \cdots, A_n$相互独立,则 $P(\bigcup_{i=1}^{n} A_i) = 1-\prod_{i=1}^{n}(1-P(A_i))$

特别地,若$P(A_i)=p$,则 $P(\bigcup_{i=1}^{n} A_i) = 1-(1-p)^n$.

证:

$$
\begin{aligned}
P(\bigcup_{i=1}^{n} A_i) &= P(A_1 \cup A_2 \cup \cdots \cup A_n) \\
&= 1-P(\bar{A_1}\bar{A_2}\cdots\bar{A_n}) \\
&= 1-P(\bar{A_1})P(\bar{A_2})\cdots P(\bar{A_n}) \\
&= 1-\prod_{i=1}^{n}(1-P(A_i))
\end{aligned}
$$

例

例 若每个人的呼吸道中有感冒病毒的概率为0.002,求在有1500人看电影的剧场中有感冒病毒的概率。

设A_i表示第i个人有病毒. A_i间相互独立.

$$
\begin{aligned}
p &= P(A_1 \cup A_2 \cup \cdots \cup A_n) \\
&= 1-P(\bar{A_1}\bar{A_2}\cdots\bar{A_n}) \\
&= 1-P(\bar{A_1})P(\bar{A_2})\cdots P(\bar{A_n}) \\
&= 1-(1-0.002)^{1500} \approx 0.95
\end{aligned}
$$

"小概率事件效应"

例

例 甲、乙两人进行乒乓球比赛,每局甲胜的概率为p,$p>1/2$。问对甲而言,采用三局二胜制有利,还是采用五局三胜制有利。设各局胜负相互独立。

三局: 甲甲 p^2 甲乙甲 $p^2(1-p)$ 乙甲甲 $p^2(1-p)$

$\Rightarrow p_1 = p^2 + 2p^2(1-p)$

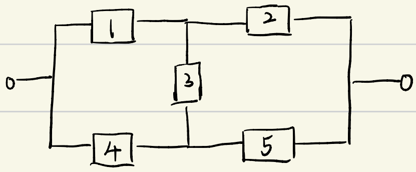
五局: 甲甲甲 (比3局) p^3

□□□甲 (比4局) □中填2个甲 $\binom{3}{2}p^3(1-p)$ $\Rightarrow p_2 = p^3 + 3p^3(1-p) + 6p^3(1-p)^2$

□□□□甲(比5局) 4个(2个)乙即可. $\binom{4}{2}p^3(1-p)^2$

$$\Rightarrow P_2-P_1=p^3+6p^3(1-p)^2+3p^3(1-p)-p^2-2p^2(1-p)$$
$$=3p^2(p-1)^2(2p-1)\geq 0.$$

例



每个元件正常工作概率均为p，且各元件工作独立。
求电路S正常工作概率。

设A_i为第i个元件正常工作。

$$P(S)=P(A_3)P(S|A_3)+P(\bar{A_3})P(S|\bar{A_3})$$

串： 记$S_1=A_1A_2$. $P(S_1)=P(A_1A_2)=p^2$

$$P(S|\bar{A_3})=P(A_1A_2\cup A_4A_5)=1-P(\overline{A_1A_2}\ \overline{A_3A_4})$$ $\overline{A_1A_2}$与$\overline{A_3A_4}$独立

$$=1-P(\overline{A_1A_2})P(\overline{A_3A_4})=1-(1-p^2)^2=1-(p^4-2p^2+1)$$
$$=2p^2-p^4.$$

并： 记$S_2=A_1\cup A_4$. $P(S_2)=P(A_1\cup A_4)=1-P(\overline{A_1\cup A_4})=1-P(\bar{A_1}\bar{A_4})=1-P(\bar{A_1})P(\bar{A_4})$
$$=1-(1-p)^2=1-(p^2-2p+1)$$
$$=2p-p^2.$$

$$P(S|A_3)=P((A_1\cup A_4)\cap(A_2\cup A_5))$$
$$=P(A_1\cup A_4)P(A_2\cup A_5)=(2p-p^2)^2$$
$$=p^4-4p^3+4p^2$$

$$\Rightarrow P(S)=p(p^4-4p^3+4p^2)+(1-p)(2p^2-p^4)$$
$$=p^5-4p^4+4p^3+2p^2-p^4-2p^3+p^5$$
$$=2p^5-5p^4+2p^3+2p^2$$

三、试验的独立性

n重Bernoulli试验：

$\begin{cases}n\text{重独立重复试验（结果独立）}\\ \text{每次试验只有2个可能的结果 } A,\bar{A} \text{ 且 } 0<P(A)<1.\end{cases}$

Part 二、随机变量及其分布

§2.1 随机变量及其分布

一、随机变量的分布

1、定义：设 $X(w)$ 是定义于可测空间 (Ω, F)（F 记作 Ω 的事件域）上的单值实函数，如果对于任意实数 x 有：$\{w: X(w) \le x\} \in F$，则称 $X(w)$ 为可测空间 (Ω, F) 上的随机变量。

简记为：r.v. X.

比如：在 $[0,1]$ 上随机取一个数，X 表示取到数的坐标，有：

$$\{X > 0.8\} = \{\overline{X \le 0.8}\}$$

$$\{0.1 < X \le 0.99\} = \{X \le 0.99\} - \{X \le 0.1\}$$

△ $$\{X < 1\} = \bigcup_{n=1}^{\infty} \{X \le 1 - \frac{1}{n}\}$$

二、随机变量的分布函数

1、定义：设 X 为 r.v.，对任意实数 x，称函数 $F(x) = P(X \le x)$ $\quad -\infty < x < +\infty$ 为 X 的分布函数。

记为 $X \sim F(x)$.

任意一个 r.v. X（离散 or 连续）都有一个分布函数。

2. 性质：**单调性** $F(x)$ 是定义在整个实数轴 $(-\infty, +\infty)$ 上的单调非减函数，

即对 $\forall x_1 < x_2$，有 $F(x_1) \le F(x_2)$

有界性 对任意的 x，有 $0 \le F(x) \le 1$，且 $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$. $F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$

右连续性 $F(x)$ 是 x 的右连续函数，即对任意的 x_0，有：

$$\lim_{x \to x_0+0} F(x) = F(x_0) \quad 或 \quad F(x_0+0) = F(x_0)$$

有界：$$1 = P(-\infty < X < +\infty) = P\left(\bigcup_{i=-\infty}^{\infty} \{i-1 \le X \le i\}\right) = \sum_{i=-\infty}^{\infty} P(i-1 \le X \le i)$$

$$= \lim_{\substack{n \to \infty \\ m \to -\infty}} \sum_{i=m}^{n} P(i-1 \le X \le i) = \lim_{n \to \infty} F(n) - \lim_{m \to -\infty} F(m)$$

$$\xRightarrow{\text{单调性}} \lim_{n \to \infty} F(n) = 1 \qquad \lim_{m \to -\infty} F(m) = 0.$$

注意：① 分布函数的不连续点（跳跃间断点）最多可列个。

② 以上三个性质是 $F(x)$ 是否为某一 r.v. X 的分布函数的充要条件。

利用分布函数表示概率:

$$P(X>a)=1-F(a)$$
$$P(X<a)=\lim_{n\to\infty}P\{X\leq a-\frac{1}{n}\}=F(a-0)$$
$$P(X\geq a)=1-F(a-0)$$
$$P(X=a)=F(a)-F(a-0)$$
$$P(a\leq X<b)=F(b-0)-F(a-0)$$

特别地，$F(x)$ 在 $x=a$ 处连续时，有 $F(a-0)=F(a)$ 或者说 $F(x)$ 在 $x=a$ 处连续当且仅当 $P(X=a)=0$

若 $F(x)$ 在 $x=a$ 处发生跳跃，是因为 $P(X=a)>0$.

三、离散随机变量及其分布列

1、定义：设 X 是一个离散型 r.v. 若 X 所有可能取值是 $x_1,x_2,\cdots,x_n,\cdots$ 则 X 取 x_i 的概率

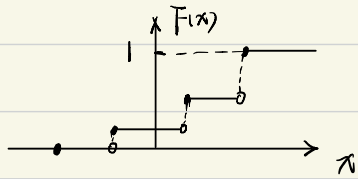
$$P(X=x_i)=p_i=p(x_i),\quad i=1,2,\cdots,n,\cdots$$

为 X 的概率分布列（分布列），记：$X\sim\{p_i\}$.

2.性质：非负性 $p(x_i)\geq 0,\ i=1,2,\cdots$

正则性 $\sum_{i=1}^{\infty}p(x_i)=1$.

是否为分布列的充要条件.



3、由离散 r.v. X 的分布列 $\Rightarrow F(x)$ $\quad F(x)=\sum_{x_i\leq x}p(x_i)$

$F(x)$ 是分段阶梯函数. 在 X 可能取值 x_i 处发生间断（跳跃间断点），间断处跃度为 $p(x_i)$.

4.单点分布（退化分布）

r.v. X 只取常值 c：$P(X=c)=1$

则：$F(x)=\begin{cases}0, & x<c\\ 1, & x\geq c\end{cases}$

四、连续型随机变量及其概率密度函数

1.定义：设 X 是 r.v. 若存在一个非负可积函数 $p(x)$，使得对任意实数 x 有：

$$F(x)=\int_{-\infty}^{x}p(t)dt\quad -\infty<x<+\infty \longrightarrow F'(x)=p(x)$$

则 $p(x)$ 是 X 的概率密度函数（概率密度 or 密度）. 记 $X\sim p(x)$

$p(x)$ 的值不是概率. 而 $p(x)dx$ 才表示概率: $p(x)dx\approx P(x<X<x+dx)$

2. 性质 非负性 $p(x)\geq 0$ } 是否为概率密度的充要条件.

正则性 $\int_{-\infty}^{+\infty} p(x)=1$

3. 连续型 r.v. 的两个重要性质:

① 对于连续型 r.v. X, $P\{X=a\}=0$.

$$0\leq P(X=a)\leq P(a-\Delta x<X\leq a)=\int_{a-\Delta x}^{a} p(t)dt \xrightarrow{\Delta x\to 0} 0$$

这表明 不可能事件概率为0, 概率为0 事件未必 不可能发生.

② 连续型 r.v. 的分布函数 F(x) 是一个连续函数.

i. 反之不一定为真. 若分布函数连续, 则对应的 r.v. 未必连续. 比如分布函数 F(x) 为康托尔函数.

ii. 一个连续分布的密度函数不唯一 (可以修改个别点而不影响积分值) 而分布函数唯一.

iii. 对一个连续型 r.v. 而言. 其概率分布就是指密度.

4. 用密度计算概率公式:

连续型 r.v. 密度为 $p(x)$, 则对任意 $a<b$, 有:

$$P(a<X\leq b)=P(a\leq X<b)=P(a<X<b)=P(a\leq X\leq b)=\int_a^b p(x)dx.$$

例

例 向区间 $[a,b]$ 内任意投点, 用 X 表示落点的坐标. 设这个点落在任一子区间的概率与这个小区间的长度成正比, 而与小区间的位置无关, 求 X 的分布函数与密度函数.

X 服从 $[a,b]$ 上的 均匀分布 $X\sim U(a,b)$

$x<a$ 时: $F(x)=P(X\leq x)=0$

$a\leq x<b$ 时: $F(x)=P(X\leq x)=P(a\leq X\leq x)=k(x-a)$ k 为比例系数

$x\geq b$ 时: $F(x)=P(X\leq x)=1$

求 k: $F(b)=F(x\leq b)=P(a\leq x\leq b)=k(b-a)=1 \Rightarrow k=\frac{1}{b-a}$

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\leq x<b\\ 1, & x\geq b\end{cases}$$

$$p(x)=F'(x)=\begin{cases}\frac{1}{b-a}, & a\leq x<b\\ 0, & 其他\end{cases}$$

例

例 已知某型号电子管的使用寿命 X 为连续型随机变量, 其密度函数为

$$p(x)=\begin{cases}\dfrac{c}{x^2}, & x>1000\\ 0, & \text{其它}\end{cases}$$

(1) 求常数 c；

(2) 计算 $P\{X\le 1700 \mid 1500<X<2000\}$;

(3) 已知一设备装有 3 个这样的电子管, 每个电子管能否正常工作相互独立, 求在使用的最初 1500 小时只有一个损坏的概率.

(1) $\int_{-\infty}^{+\infty} p(x)\,dx = \int_{1000}^{+\infty} \frac{c}{x^2}\,dx = -\frac{c}{x}\Big|_{1000}^{+\infty} = \frac{c}{1000} = 1 \Rightarrow c=1000$

(2) $P(X\le 1700 \mid 1500<X<2000) = \dfrac{P(X\le 1700)\,P(1500<X<2000)}{P(1500<X<2000)} = \dfrac{P(1500<X\le 1700)}{P(1500<X<2000)}$

$P(1500<X\le 1700) = \int_{1500}^{1700} \frac{1000}{x^2}\,dx = 1000\left(\frac{1}{1500}-\frac{1}{1700}\right) = \frac{4}{51}$

$P(1500<X<2000) = \int_{1500}^{2000} \frac{1000}{x^2}\,dx = 1000\left(\frac{1}{1500}-\frac{1}{2000}\right) = \frac{1}{6}$

$P_2 = \frac{24}{51}$

(3) $P(A) = P(0\le X<1500) = \int_{1000}^{1500} \frac{1000}{x^2}\,dx = \frac{1}{3}$

$P_3 = C_3^1\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^2 = \frac{4}{9}$

§2.2 随机变量的数学期望

一、离散型随机变量的数学期望

1.定义：设X为离散型r.v.，分布列为：$p(x_i)=P(X=x_i)$　$i=1,2,\cdots,n,\cdots$

如果：$\sum_{i=1}^{\infty}|x_i|p(x_i)<\infty$　绝对收敛.

则称：$E(X)=\sum_{i=1}^{\infty}x_i p(x_i)$ 为r.v. X的数学期望（期望 or 均值）

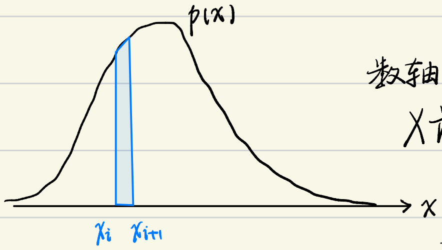
r.v. X分布列为：$P\left(X=(-1)^n\frac{2^n}{n}\right)=\frac{1}{2^n}$，$n=1,2,\cdots$　则无法定义数学期望.

二、连续型随机变量的数学期望

1.定义：设X为连续型r.v.，其密度函数为p(x)，

如果：$\int_{-\infty}^{+\infty}|x|p(x)dx<\infty$

则称：$E(X)=\int_{-\infty}^{+\infty}|x|p(x)dx$ 为X的数学期望（期望 or 均值）.



数轴密分成：$x_0<x_1<\cdots<x_i<\cdots$

近似为蓝条面积.

X落在$[x_i,x_{i+1})$上概率为：$\int_{x_i}^{x_{i+1}}p(x)dx\approx p(x_i)(x_{i+1}-x_i)\approx p(x_i)\Delta x_i$

所以用 $p(x)dx$ 代表 x 处概率.

$$\sum_{i=1}^{\infty}x_i p(x_i)=\int_{-\infty}^{+\infty}x p(x)dx$$

例　若$X\sim U[a,b]$，求$E(X)$.

X的概率密度函数 $p(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & \text{其他}\end{cases}$

$\left(\begin{aligned}&F(x)=F(X\le x)=P(a<x\le x)=k(x-a)\quad F(b)=1\Rightarrow k=\frac{1}{b-a}\\ &p(x)=F'(x)=\frac{1}{b-a}\quad a\le x\le b.\end{aligned}\right)$

$$\int_{-\infty}^{+\infty}|x|p(x)dx=\int_a^b\frac{|x|}{b-a}dx=\frac{1}{b-a}\int_a^b|x|dx<\infty.$$

$$\int_{-\infty}^{+\infty}xp(x)dx=\frac{1}{b-a}\int_a^b x\,dx=\frac{\frac{1}{2}(b^2-a^2)}{b-a}=\frac{a+b}{2}.$$

例　设r.v. X的期望存在，概率密度p(x)关于$x=\mu$对称：$p(\mu+x)=p(\mu-x)$. 证：$E(X)=\mu$.

$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_{-\infty}^{+\infty}\mu p(x)dx+\int_{-\infty}^{+\infty}(x-\mu)p(x)dx$　（第一项积分值为μ，是凑出来μ）

$=\mu\int_{-\infty}^{+\infty}p(x)dx+\int_{-\infty}^{+\infty}(x-\mu)p(x)dx=\mu+\int_{-\infty}^{+\infty}(x-\mu)p(x)dx$　（第二项考虑对称消去）

$\underline{令t=x-\mu}$　$\mu+\int_{-\infty}^{+\infty}t\,p(t+\mu)dt$ ← 中心对称，积分为0　$=\mu$　QED.

记$f(t)=tp(t+\mu)$　$f(-t)=-tp(\mu-t)=-tp(t+\mu)\Rightarrow f(t)+f(-t)=0$　f(t)为奇函数

例 为普查某种疾病, n 个人需验血. 验血方案有如下两种:
(1) 分别化验每个人的血, 共需化验 n 次;
(2) 分组化验, k 个人的血混在一起化验, 若结果为阴性, 则只需化验一次; 若为阳性, 则对 k 个人的血逐个化验, 找出有病者, 此时 k 个人的血需化验 $k+1$ 次.

设每人血液化验呈阳性的概率为 p, 且每人化验结果是相互独立的.试说明选择哪一方案较经济.

对(2): 设 n 是 k 倍数, 分成 $\frac{n}{k}$ 组.

第 i 组要化验次数 X_i 为 1 或 $(k+1)$ 次.

$p(X_i=1)=(1-p)^k \quad p(X_i=k+1)=1-(1-p)^k$

X_i	1	$k+1$
P	$(1-p)^k$	$1-(1-p)^k$

$$E(X_i)=(1-p)^k+(k+1)[1-(1-p)^k]$$
$$=(k+1)-k(1-p)^k$$
$$E(X)=\frac{n}{k}E(X_i)=\frac{n}{k}[(k+1)-k(1-p)^k]$$
$$=n\left[1-\left((1-p)^k-\frac{1}{k}\right)\right]$$

若 $(1-p)^k-\frac{1}{n}>0$, 则 $E(X)<n$ 此时(2)更经济.

所以数学期望就是一种"加权平均"

从随机变量 → 数.

三、随机变量函数的数学期望 从 X 到 $g(X)$

1、定理：设 X 为一个 r.v.，$Y=g(X)$，则

$$E(Y)=E(g(X))=\begin{cases}\sum_{i=1}^{\infty} g(x_i)p_i, & X\text{离散型}(P(X=x_i)=p_i)\\ \int_{-\infty}^{+\infty} g(x)p(x)dx, & X\text{连续型}(p(x)\text{为}X\text{密度函数})\end{cases}$$

求 $E[g(X)]$，不用知 $g(X)$ 分布，用 X 分布直接求即可。

例 **例2** 国际市场上对我国某种出口商品的每年需求量是个随机变量 X(吨)．X 服从区间[300, 500]上的均匀分布．每销售出一吨商品，可为国家赚取外汇1.5千元；若销售不出，则每吨商品需贮存费0.5千元．求：应组织多少货源，才能使平均收益最大？

设组织 t 吨货源．$300\le t\le 500$．收益 $Y=g(X)$

$$Y=g(X)=\begin{cases}1.5t, & t\le X\\ 1.5X-0.5(t-X)=2X-0.5t, & t>X\end{cases}$$

X 的概率密度为 $p(x)=\begin{cases}\frac{1}{200}, & 300\le t\le 500\\ 0, & \text{其他}\end{cases}$

$$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty} g(x)p(x)dx=\frac{1}{200}\int_{300}^{500} g(x)dx$$

$$=\frac{1}{200}\left[\int_{300}^{t}(2x-0.5t)dx+\int_{t}^{500}1.5t\,dx\right]$$

$$=\frac{1}{200}\left[-t^2+900t-300^2\right]$$

t 取 450，$E(Y)_{max}$.

四、数学期望的性质

1. 设 C 是常数，$E(C)=C$．　$(P(X=c)=1,\ E(X)=E(c)=c\times 1=c)$
2. 对任意常数 a，有 $E(aX)=aE(X)$
3. $E[g_1(X)\pm g_2(X)]=E[g_1(X)]\pm E_2[g_2(X)]$

   特别地，$E(aX+b)=aE(X)+b$

   推广：$E\left[\sum_{i=1}^{n} g_i(X)\right]=\sum_{i=1}^{n} E[g_i(X)]$
4. 若 $a\le X\le b$，则 $a\le E(X)\le b$.

§2.3 随机变量的方差与标准差

一、方差与标准差的定义

1.定义：若r.v. X^2的数学期望$E(X^2)$存在，则称偏差平方$(X-E(X))^2$的数学期望$E(X-E(X))^2$为r.v. X的方差.

记为：$Var(X) = E(X-E(X))^2$　(不是E的平方，是平方的E)

$$= \begin{cases} \sum_i (x_i - E(X))^2 p(x_i) & \text{离散r.v.} \\ \int_{-\infty}^{+\infty} (x-E(X))^2 p(x)dx & \text{连续r.v.} \end{cases}$$

从离散r.v.看：$\sum_i (x_i - E(X))^2 p(x_i)$，$(x_i - E(X))^2$刻画$x_i$的波动程度，$p(x_i)$表示$x_i$出现概率.

标准差：$\sqrt{Var(X)}$，记为$\sigma(X)$.

$\sigma(X)$与$E(X)$，X量纲相同，所以$\sigma(X)$常用.

r.v. X数学期望存在 $\underset{一定}{\overset{不一定}{\rightleftarrows}}$ 方差存在

二、方差的性质

假设r.v.方差存在.

1、$Var(X) = E(X^2) - [E(X)]^2$

$$Var(X) = E(X-E(X))^2 = E(X^2 - 2XE(X) + (E(X))^2)$$
$$= E(X^2) - E(2XE(X)) + E((E(X))^2)$$
$$\text{而}E(c) = c\ (c\text{为常数}) \Rightarrow = E(X^2) - 2E(X)E(X) + (E(X))^2$$
$$= E(X^2) - (E(X))^2.$$

2、常数的方差为0. 即$Var(c)=0$. c为常数.

3、若a,b是常数，则$Var(aX+b) = a^2 Var(X)$

$$Var(aX+b) = E(aX+b-E(aX+b))^2 = E(aX+b-aE(X)-b)^2 = E(aX-aE(X))^2$$
$$= E(a(X-E(X)))^2 = E(a^2(X-E(X))^2) = a^2 E(X-E(X))^2 = a^2 Var(X)$$

另外，$Var(X) = E(X^2) - (E(X))^2 \geq 0$ 有：$\boxed{\text{若}E(X^2)=0 \Rightarrow E(X)=0，\text{且} Var(X)=0}$

仅知r.v.的期望和方差不可以确定其分布.

例 设r.v. X的期望和方差分别为$E(X)$和$Var(X)$，且$Var(X)>0$. 求：

$Y = \dfrac{X-EX}{\sqrt{Var(X)}}$的期望和方差　(X的标准化)

注：$EX = E(X)$，$EX^2 = E(X^2)$

$\to EX=C \quad E(c)=C$

$$E(Y)=E\left(\frac{X-EX}{\sqrt{Var(X)}}\right)=E\left(\frac{X}{\sqrt{Var(X)}}\right)-E\left(\frac{EX}{\sqrt{Var(X)}}\right)=0$$

$$Var(Y)=Var\left(\frac{X-EX}{\sqrt{Var(X)}}\right)=Var\left(\frac{X}{\sqrt{Var(X)}}\right)=\left(\frac{1}{\sqrt{Var(X)}}\right)^2\cdot Var(X)=1$$

$\frac{EX}{\sqrt{Var(X)}}=C_1,\ Var(X+C_1)=Var(X)$ $\qquad$ $\frac{1}{\sqrt{Var(X)}}=C_2,\ Var(aX)=a^2Var(X)$

三、Chebyshev 不等式

设r.v. X 的期望和方差都存在，则对任意常数 $\varepsilon>0$，有：

$$P(|X-E(X)|\geq\varepsilon)\leq\frac{Var(X)}{\varepsilon^2}$$

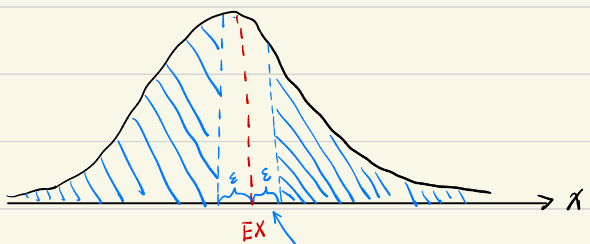
“记忆：$\geq\varepsilon\leq$ 像颜文字.”

或 $$P(|X-E(X)|<\varepsilon)\geq 1-\frac{Var(X)}{\varepsilon^2}$$

设X是一个连续型r.v.，其密度函数为 $p(x)$，记 $E(X)=a$.

$$P(|X-a|\geq\varepsilon)=\int_{\{x:|x-a|\geq\varepsilon\}}p(x)dx\leq\int_{\{x:|x-a|\geq\varepsilon\}}\frac{(x-a)^2}{\varepsilon^2}p(x)dx$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-a)^2p(x)dx=\frac{Var(X)}{\varepsilon^2}$$



X不落在这一块区域的概率不超过 $\frac{Var(X)}{\varepsilon^2}$

事件 $\{|X-E(X)|\geq\varepsilon\}$ 称为大偏差（不可控），给出不可控的概率上限.

若r.v. X的方差存在，则 $Var(X)=0$ 的充要条件是X几乎处处为某个常数a，即 $P(X=a)=1$.

就是E(X)

例

投资组合问题：现有1万元，用于投资两只股票及固定收益，已知投资信息如下：

	股票1	股票2	固定收益
年收益率	R_1	R_2	0.018
$E(R_i)$ 收益率的均值	0.1	0.06	0.018
$D(R_i)$ 收益率的方差	0.065	0.028	0

求：平均收益不低于600元时的最优投资组合方案.

稳定.

设股票1，2，固定收益分别投入 S_1, S_2, S_3 元.

$$S_1+S_2+S_3=10000$$

总收益：$T=S_1R_1+S_2R_2+S_3\cdot 0.018$

总收益均值：$E(T)=E(S_1R_1+S_2R_2+0.018S_3)$

$$=0.1S_1+0.06S_2+0.018S_3$$

总收益方差：$Var(T)=Var(S_1R_1+S_2R_2+0.018S_3)$

$$=S_1^2\,Var(R_1)+S_2^2\,Var(R_2)+0$$

$$=0.065S_1^2+0.028S_2^2$$

模型：$\min\ Var(T)=0.065S_1^2+0.028S_2^2$

$$s.t.\begin{cases}S_1+S_2+S_3=10000\\ S_1,S_2,S_3\geqslant 0\\ E(T)=0.1S_1+0.06S_2+0.018S_3\geqslant 600.\end{cases}$$

非线性规划 → 最优化.

§2.4 常用离散分布

一、二项分布 (binomial distribution)

1. 定义：n重 Bernoulli 试验中，X是在n次试验中事件A发生的次数，$P(A)=p$，则：

$$P(X=k)=C_n^k p^k (1-p)^{n-k}\quad ,k=0,1,2,\cdots,n$$

称X服从参数为n,p的二项分布，记作 $X\sim b(n,p)$.

$$\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=[p+(1-p)]^n=1$$

二项概率 $\binom{n}{k}p^k(1-p)^{n-k}$ 恰好是n次二项式 $(p+(1-p))^n$ 展开式中的第 k+1 项.

当n=1时 $b(1,p)$ 称为二点分布(0-1分布). 分布列为：

X	0	1
P	1-p	p

二项分布 $b(n,p)$ 和两点分布 $b(1,p)$ 重要关系：

设X为n次独立试验中结果A出现次数，则 $X\sim b(n,p)$

把第i次试验中A出现次数记作 X_i，$X_i\sim b(1,p)$. 且 $X_1,X_2,\cdots,X_n$ 也是相互独立的，有：

$$X=X_1+X_2+\cdots+X_n$$

2. 二项分布的期望与方差

设r.v. $X\sim b(n,p)$，则： $E(X)=np,\ Var(X)=np(1-p)$

证：

$$E(X)=\sum_{k=0}^{n}kC_n^kp^k(1-p)^{n-k}=\sum_{k=0}^{n}k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$=np\boxed{\sum_{k=0}^{n}k\frac{(n-1)!}{k!(n-k)!}p^{k-1}(1-p)^{n-k}}$$

由结果凑　　　　目标为1，利用二项式展开

$$=np\cdot\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k}$$

$$=np\cdot\boxed{\sum_{k=1}^{n}C_{n-1}^{k-1}p^{k-1}(1-p)^{(n-1)-(k-1)}}=[p+(1-p)]^{n-1}=1$$

$$=np.$$

类似地：$Var(X)=E(X^2)-(E(X))^2=E(X^2)-n^2p^2$ 　$(Var(X)=np(1-p)\Rightarrow \underbrace{E(X^2)=n(n-1)p^2+np}_{\text{目标}})$

$$E(X^2)=\sum_{k=0}^{n}k^2\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$=\underbrace{\boxed{\sum_{k=0}^{n}k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}}}_{np}+\sum_{k=0}^{n}k(k-1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}.$$

$$=np+n(n-1)p^2\cdot\sum_{k=2}^{n}\frac{(n-2)!}{(k-2)!(n-k)!}p^{k-2}(1-p)^{n-k}$$

$$=np+n(n-1)p^2\cdot\boxed{\sum_{k=2}^{n}C_{n-2}^{k-2}p^{k-2}(1-p)^{(n-2)-(k-2)}}=[p+(1-p)]^{n-2}=1$$

$$=np+n(n-1)p^2$$

若 $X \sim b(n,p)$，则 $X=k$ 时，$P(X=k)$ 最大？

$$\frac{b(k+1;n,p)}{b(k;n,p)} = \frac{C_n^{k+1}p^{k+1}(1-p)^{n-k-1}}{C_n^k p^k (1-p)^{n-k}} = \frac{(n-k)p}{(k+1)(1-p)} = 1 - \frac{k-[(n+1)p-1]}{(k+1)(1-p)}$$

当 $(n+1)p$ = 整数时，在 $k=(n+1)p$ 与 $(n+1)p-1$ 处概率取最大值

当 $(n+1)p \neq$ 整数时，在 $k=[(n+1)p]$ 处概率取最大值

取整

k 叫最可能成功次数 $b(k;n,p)$

例

例. 设每颗子弹打中飞机的概率为 0.01，问在 500 发子弹中打中飞机的最可能次数是多少？并求其相应的概率。

$X \sim b(0.01, 500)$

$\Rightarrow$ 最可能次数 $m = [(n+1)p] = [501 \times 0.01] = 5$

$b(5;500,0.01) = \binom{500}{5}(0.01)^5(0.99)^{495} = 0.176$

二、超几何分布 (Hypergeometric distribution)

1. 定义：对 N 件产品进行不放回抽样检查，若这批产品中有 M 件次品，现从整批产品中随机抽出 $n\ (\le N)$ 件产品，则在这 n 件产品中出现的次品数 X 是随机变量，其概率分布为超几何分布。

记作：$X \sim h(n,N,M)$

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \qquad 0 \le k \le n \le N,\ k \le M$$

2. 期望与方差：

$X \sim h(n,N,M) \Rightarrow E(X) = n\frac{M}{N}$ $\qquad Var(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$ (不用背)

3. 近似计算：

当 $n \ll N$ 时，近似为二项分布进行计算.

三、几何分布（Geometric distribution）

1、定义：在事件A发生的概率为p的bernoulli试验中，以X记作A首次出现时的试验次数，则X为r.v.

其概率分布为几何分布，记作 $X\sim Ge(p)$.

$$P(X=k)=(1-p)^{k-1}p,\quad k=1,2,\cdots$$

2、期望与方差：

$$X\sim Ge(p)\Rightarrow E(X)=\frac{1}{p}、Var(X)=\frac{1-p}{p^2}$$

证：令 $q=1-p$

$$E(X)=\sum_{k=1}^{\infty}kq^{k-1}p=p\sum_{k=1}^{\infty}kq^{k-1}$$ （利用无穷级数中的逐项积分）

$$=p\sum_{k=1}^{\infty}\frac{dq^k}{dq}=p\frac{d}{dq}\left(\sum_{k=0}^{\infty}q^k\right)=p\cdot\frac{d}{dq}\left(\frac{1}{1-q}\right)=\frac{p}{(1-q)^2}$$

$$=\frac{1}{p}$$

类似：$$E(X^2)=\sum_{k=1}^{\infty}k^2q^{k-1}p=p\Big[\underbrace{\sum_{k=1}^{\infty}k(k-1)q^{k-1}}_{\text{变形后逐项积分}}+\underbrace{\sum_{k=1}^{\infty}kq^{k-1}}_{\text{已知的，凑出来}}\Big]$$

$$=pq\sum_{k=1}^{\infty}k(k-1)q^{k-2}+\frac{1}{p}$$

$$=pq\sum_{k=1}^{\infty}\frac{d^2q^k}{dq^2}+\frac{1}{p}=pq\frac{d^2}{dq^2}\left(\sum_{k=0}^{\infty}q^k\right)+\frac{1}{p}$$

$$=pq\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right)+\frac{1}{p}=pq\frac{2}{(1-q)^3}+\frac{1}{p}$$

$$=\frac{2q}{p^2}+\frac{1}{p}$$

$$\Rightarrow Var(X)=E(X^2)-[E(X)]^2=\frac{2q}{p^2}+\frac{1}{p}-\frac{1}{p^2}=\frac{2(1-p)}{p^2}+\frac{1}{p}-\frac{1}{p^2}$$

$$=\frac{1-p}{p^2}$$

3. 几何分布的无记忆性

设 $X\sim Ge(p)\Leftrightarrow$ 对 $\forall$ 正整数 m,n 有：$P(X>m+n\mid X>m)=P(X>n)$

在离散型分布中，只有几何分布才有这样特殊性质。

特殊地：$P(X=k+1\mid X>k)=P(X=1)$

证：$$P(X>n)=\sum_{k=n+1}^{\infty}(1-p)^{k-1}p=p\cdot\frac{(1-p)^n}{1-(1-p)}=(1-p)^n$$

$$P(X>m+n\mid X>m)=\frac{P(X>m+n)}{P(X>m)}=\frac{(1-p)^{m+n}}{(1-p)^m}=(1-p)^n$$

解释：

在前m次试验中A没发生，那么在接下来n次试验A仍不发生的概率只与n有关，而与m无关，像是忘记了前m次经历。$\searrow$ 独立.

四、负二项分布（帕斯卡分布）

1、定义：在bernoulli试验中，记每次试验中事件A发生概率为p，记X为事件A第r次成功出现时的实验次数，则X为r.v.

X服从负二项分布（帕斯卡分布），记 $X\sim Nb(r,p)$.

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

k次中最后一次一定是A，从剩下(k-1)次中选出(r-1)次.　　最后一次出现A的概率也要乘上，所以是r次.

2、期望与方差

$$X\sim Nb(r,p)\Rightarrow E(X)=\frac{r}{p}\qquad Var(X)=\frac{r(1-p)}{p^2}$$

对比 $X\sim Ge(p)$，就是多乘上r.

可以理解成：由几何分布的无记忆性，每个出现的E都是 $\frac{1}{p}$，则r次就是 $\frac{r}{p}$.

$Nb(r,p)$ 与 $Ge(p)$ 之间的关系：

记 X_i 为试验A第r次出现时的次数（每一次出现后再从0开始计数），则 $X_1,X_2,\cdots,X_r$ 独立同分布，$X_i\sim Ge(p)$，有：$X=X_1+X_2+\cdots+X_r\sim Nb(r,p)$

$$\underbrace{\bar{A}\bar{A}\cdots\bar{A}A}_{X_1}\ \underbrace{\bar{A}\bar{A}\bar{A}\cdots\bar{A}A}_{X_2}\cdots\underbrace{\bar{A}\cdots\bar{A}A}_{X_r}$$

例

分赌注问题 甲、乙两个赌徒按某种方式下注赌博，说定先胜 t 局者将赢得全部赌注，但进行到甲胜 r 局，乙胜 s 局（$r<t$，$s<t$）时，因故不得不中止，试问如何分配这些赌注才公平合理？

设 $n=t-r$，$m=t-s$ 记为甲，乙获取胜利要再赢的场次．设每局甲赢概率为p．　　$q=1-p$.

甲赢为事件A．赌注为a.

$$P(\text{甲胜})=p^n+C_n^1p^nq+\cdots+C_{n+m-1}^{m-1}p^nq^{m-1}$$
$$=\sum_{k=0}^{m-1}\binom{n+k-1}{k}p^nq^k$$

甲：$aP(\text{甲胜})$

五. 泊松分布 (Poisson distribution)

1、引入：设X表示放射性物质t秒内发射出k个粒子，设X为离散型r.v.

将t秒等分成n份，$\Delta_n=\frac{t}{n}$，对于充分大的n，设：

(1) 在Δ_n内最多有一个粒子放出，放出一个粒子概率为$p_n=\mu\cdot\frac{t}{n}$　μ为正常数

(2) 各个小时间段Δ_n内是否放出粒子独立。

则$X\sim b(n,p_n)$.

$$P(X=k)=\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}$$

$$=\lim_{n\to\infty}\frac{n!}{k!(n-k)!}\left(\frac{\mu t}{n}\right)^k\left(1-\frac{\mu t}{n}\right)^{n-k}$$

$$=\frac{(\mu t)^k}{k!}\cdot\lim_{n\to\infty}\frac{n!}{(n-k)!\,n^k}\left(1-\frac{\mu t}{n}\right)^{n-k}$$

$$=\frac{(\mu t)^k}{k!}\cdot\underbrace{\lim_{n\to\infty}\left(\frac{n-1}{n}\times\frac{n-2}{n}\times\cdots\times\frac{n-k+1}{n}\right)}_{1}\cdot\underbrace{\lim_{n\to\infty}\left(1-\frac{\mu t}{n}\right)^{n-k}}_{1^\infty}$$

$$=\frac{(\mu t)^k}{k!}\lim_{n\to\infty}\left[\left(1-\frac{\mu t}{n}\right)^{-\frac{n}{\mu t}}\right]^{-\frac{\mu t}{n}\cdot(n-k)}=\frac{(\mu t)^k}{k!}\cdot\lim_{n\to\infty}\exp\left(-\frac{\mu t}{n}(n-k)\right)$$

$$=\frac{(\mu t)^k}{k!}\cdot e^{-\mu t}$$

令$\lambda=\mu t$，则$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$.

2、定义：若$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$，$k=0,1,2,\cdots$，$\lambda>0$是常数

则称X服从参数为λ的泊松分布，记作$X\sim P(\lambda)$.（$X\sim\pi(\lambda)$）

3. 期望与方差

$X\sim P(\lambda)\Rightarrow E(X)=Var(X)=\lambda$

证：

$$E(X)=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda\cdot e^{-\lambda}\cdot e^{\lambda}=\lambda.$$

$$E(X^2)=\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}k\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$=\sum_{k=1}^{\infty}[(k-1)+1]\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$=\sum_{k=1}^{\infty}(k-1)\frac{\lambda^k}{(k-1)!}e^{-\lambda}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$=\lambda^2e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda^2e^{-\lambda}\cdot e^{\lambda}+\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda^2+\lambda.$$

$$Var(X)=E(X^2)-[E(X)]^2=\lambda^2+\lambda-\lambda^2=\lambda.$$

4. 二项分布的泊松近似

泊松定理 在n重bernoulli试验中，记事件A在一次试验中发生概率为p_n（与试验次数n有关），若当$n\to\infty$时，有$np_n\to\lambda$，则：

$$\lim_{n\to\infty} P(X=k) = \lim_{n\to\infty}\binom{n}{k}(p_n)^k(1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$

当n很大，p很小，而$\lambda=np$大小适中时可以用Poisson分布近似二项分布，即：

$$b(k;n,p) \approx \frac{(np)^k}{k!}e^{-np}$$

§2.5 常用连续分布

一、均匀分布 (Uniform distribution)

1、定义：若 r.v. X 的密度函数为

$$p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & 其他\end{cases}$$

则称 X 服从区间 (a,b) 上的均匀分布，记作 $X\sim U(a,b)$.

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$$

若对 $\forall (c,d)\subset(a,b)$, $P(c<X<d)=\int_c^d \frac{1}{b-a}dx=\frac{d-c}{b-a}$

例. **例1 设X~U(0, 10), 现对 X 作4 次独立观测, 试求至少有 3 次观测值大于5 的概率.**

$$p(x)=\begin{cases}\frac{1}{10}, & 0<x<10\\ 0, & 其他\end{cases}$$

$$P(X>5)=\int_5^{\infty}p(x)dx=\int_5^{10}\frac{1}{10}dx=\frac{1}{2}.$$

$$P=C_4^3\left(\tfrac{1}{2}\right)^4+C_4^4\left(\tfrac{1}{2}\right)^4=\frac{5}{16}.$$

2、期望与方差

$X\sim U(a,b)\Rightarrow E(X)=\frac{a+b}{2}$ $\quad Var(X)=\frac{(b-a)^2}{12}$

二、指数分布 (Exponential distribution)

1、定义：若 r.v. X 的密度函数为

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}\qquad \lambda>0$$

则称 X 服从指数分布，记作 $X\sim Exp(\lambda)$.

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$$

$$\int_0^x \lambda e^{-\lambda x}dx=-e^{-\lambda x}\Big|_0^x=1-e^{-\lambda x}$$

2、期望与方差