
概率论与数理统计课程群
2022
该二维码7天内(10月3日前)有效，重新进入将更新

Chpt.2 Random Variables & Probability Distributions

第二章 随机变量及其分布

上节回顾

- 全概率公式

B_1,B_2, ---, B_n为样本空间S的一个划分，且 $P(B_i)>0$, $i=1,2,\ldots,n$。

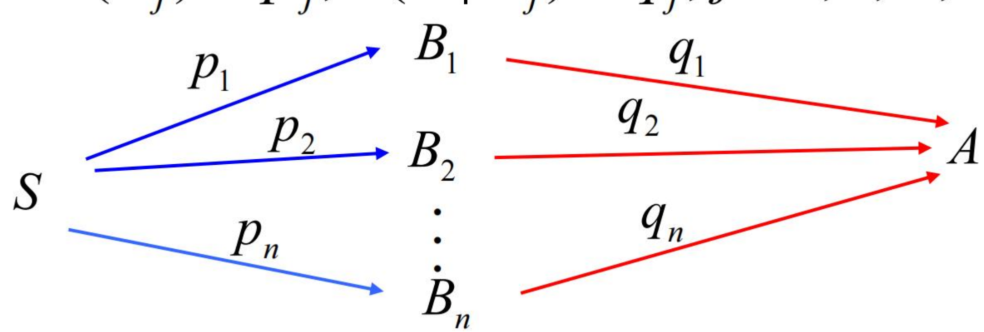
则E的任意事件A的概率为

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\ldots+P(A|B_n)P(B_n)$$

- 贝叶斯公式

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

设$P(B_j)=p_j, P(A|B_j)=q_j, j=1,2,\ldots,n.$



思考题

根据以往的临床记录，某种诊断癌症的试验具有5%的假阳性及3%的假阴性。

若设A={试验反应是阳性}，C={被诊断者患有癌症}，则有

$$P(A|\bar{C}) = 5\%, P(\bar{A}|C) = 3\%,$$

现对自然人群进行普查，设患有癌症的概率为0.005, P(C)=0.005，问这种方法能否用于普查？

提示：计算P(C|A)

解 已知 $P(A|C)=0.95$，$P(A|\bar{C})=1-P(\bar{A}|\bar{C})=0.05$，$P(C)=0.005$，$P(\bar{C})=0.995$，由贝叶斯公式

$$P(C|A)=\frac{P(A|C)P(C)}{P(A|C)P(C)+P(A|\bar{C})P(\bar{C})}=0.087.$$

本题的结果表明，虽然 $P(A|C)=0.95$，$P(\bar{A}|\bar{C})=0.95$，这两个概率都比较高. 但若将此试验用于普查，则有 $P(C|A)=0.087$，亦即其正确性只有 8.7%(平均 1000 个具有阳性反应的人中大约只有 87 人确患有癌症). 如果不注意到这一点，将会得出错误的诊断，这也说明，若将 $P(A|C)$ 和 $P(C|A)$ 混淆了会造成不良的后果. □

回顾随机事件

在古典（几何）概型中，我们定义了随机事件，试图使用随机事件来解决一个复杂的概率问题：

其中我们将分析的复杂的随机事件分解成若干个基本随机事件，计算对应的概率。

这是一个case study，一事一议。一个题就要研究一次。

随机事件的定义依赖于样本空间；而在语言描述的古典概型问题中，出题人往往是没有清晰定义样本空间的。此时需要答题人自行凝练样本空间。此时由于文字语言的不够精确性，所以往往样本空间可能存在问题。

概率研究中的Case Study方法是否能解决所有问题？

前面已介绍一些概率模型、很多例题，都有代表性；

但是可以找出更多的例子进行研究，永无止境！

思考：不同的case之间有无本质、统一的东西？

进一步，不同的概率现象之间有什么样的联系？

2.1 Random Variables

2.1.1 概率空间*

概率空间包含三个元素，$\{\Omega, \mathcal{F}, P\}$。

- Ω是样本空间，
- $\mathcal{F}$是随机事件的集合（即Ω子集的集合），必须是一个σ代数

  1) $\Omega \in \mathcal{F}$；

  2) $A \in \mathcal{F} \Rightarrow A^c = \Omega \backslash A \in \mathcal{F}$；

  3) $A_n \in \mathcal{F} \Rightarrow \bigcup_n A_n \in \mathcal{F}$， $n \in \mathbb{N}$
- $P: \mathcal{F} \to [0,1]$是满足非负性、规范性、可列可加性的一个函数，即给每个事件赋予一个0到1之间的数。

2.1 Random Variables

2.1.1 概率空间*

例子：连续抛两次硬币，正面朝上H，反面朝上T

$$\Omega = \{HH, HT, TH, TT\}.$$

$$\begin{aligned}\mathcal{F}_1 = \{&\{\varnothing\}, \{HH\}, \{HT\}, \{TH\}, \{TT\},\\ &\{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\},\\ &\{HH, HT, TH, \}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\},\\ &\{HH, HT, TH, TT\}\}\end{aligned}$$

$$\mathcal{F}_2 = \{\{\varnothing\}, \{HH, HT, TH, TT\}\}$$

P(HH)=P(HT)=P(TH)=P(TT)=1/4

$\{\Omega, \mathcal{F}_1, P\}$、$\{\Omega, \mathcal{F}_2, P\}$都是概率空间

注意：$\mathcal{F}$和P都不是唯一的

2.1 Random Variables

2.1.1 Introduction

对于数值函数我们已经有足够多的研究成果

自然希望把一般的概率问题 $\Rightarrow$ 数值函数问题。

事件空间 $\Rightarrow$ 实数集合

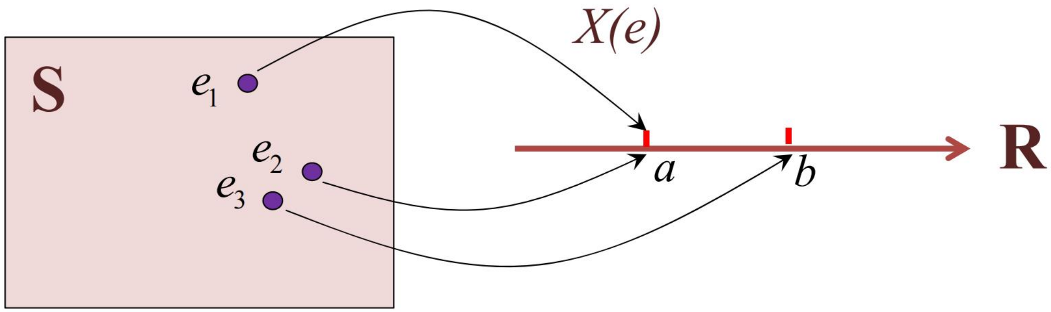
研究事件 $\Rightarrow$ 研究数值

那么，概率函数是否可用？

概率函数定义在$\mathcal{F}$上， $\forall A \in \mathcal{F}$有$P(A)$与之对应，但由于事件的多样性，导致使抽取其共性变得困难；且对P的要求较多，不太便于研究。

2.1.2 随机变量的定义

随机试验E，其样本空间为Ω，如果对每个e∈ Ω，有唯一一个实数X(e)与之对应，就得到一个定义在Ω上的单值实函数X＝X(e)，称为**随机变量**。



例子：两个孩子的问题

X=女孩的个数

$\Omega = \{BB, BG, GB, GG\}$，X={0，1，2}

事件{至少有一个女孩} = {e: X(e)≥1} = {X≥1}

概率P({至少有一个女孩})=P({X≥1})=P(X≥1)

Ω => X

A =>$\{X \in L\}$

P(A) =>P($X \in L$)

2.1 Random Variables

随机变量定义方法的优势：

1. 成功去除了随机事件必须使用语言描述的弊端，研究中讨论的是一个实数域上的数值函数的概率
2. 更为抽象化，避免了“一事一议”，对于不同的随机事件，只要能够合理的映射到同样一个随机变量的数值空间，都可以做一样的研究。

2.1 Random Variables

- 概率 vs 随机变量

  $P: \mathcal{F} \rightarrow [0,1]$，满足非负，规范，可列可加性

  $X: \Omega \rightarrow \mathbb{R}$

- 变量 vs 随机变量

  随机变量有一组取值，且取每个具体的值都有一定的概率

Example 2.1 某射手每次射击命中率为0.02，现在不断地射击，直到命中目标为止。

定义 X＝命中目标所需的射击次数

$$R_x = \{1, 2, \cdots\}$$

$$\{\text{射击到第100次命中}\} \Leftrightarrow \{X \in L_1\}，\ L_1 = \{100\}$$

$$\{\text{不多于100次命中目标}\} \Leftrightarrow \{X \in L_2\}，\ L_2 = \{1, 2, \cdots 100\}$$

填空题 1分

某公共汽车站每5分钟有一辆汽车通过，如果某人到达该车站的时刻是随机的。那么，他等车的时间X是一个随机变量，该随机变量的值域为 [填空1]（以分钟为单位）

作答

2.2.1 离散型随机变量的定义

若随机变量X可能取的值为有限个或可数个， 则称X为离散型随机变量(discrete random variable)。

可数集(也称为可列集)：是指能与自然数集N建立一一对应的集合. 即其中的元素都是可以被数到的.

如：正奇数集$\{1,3,\cdots\}$

(取其中一数为 2746489473673561, 肯定可以数到)

整数集$\{\cdots,-2,-1,0,1,2,\cdots\}$, 等等.

Example： [1] 上面例子中的射击命中次数X

[2] 在[0，1]区间上方随机抛球，观察落到有理点X上的情形。

2.2.2 概率分布律

对离散型随机变量涉及到两点：[1] 随机变量所有可能的取值；[2] 取每个值的概率。

设X所有可能的取值为x_k (k=1,2,…)，X取x_k的概率记为

$$P(X=x_k)=p_k,\qquad k=1,2,\cdots.$$

称之为离散随机变量X的概率分布或分布律。

其直观的表示是列表

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{pmatrix}$$

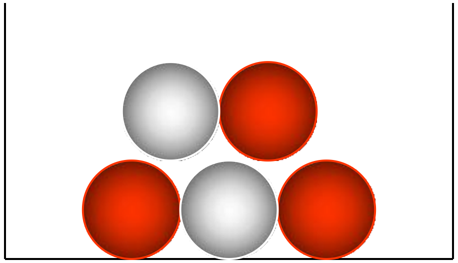
Example 2.5 如右图所示，从盒中任取3个球。取到的白球数X是一个随机变量。求X的分布律。

X可能取的值是0,1,2。取每个值的概率为:

$$P(X=0)=\frac{C_3^3}{C_5^3}=\frac{1}{10}$$

$$P(X=1)=\frac{C_3^2C_2^1}{C_5^3}=\frac{6}{10}$$

$$P(X=2)=\frac{C_3^1C_2^2}{C_5^3}=\frac{3}{10}$$



其分布律为

X	0	1	2
p_i	0.1	0.6	0.3

填空题 1分

设随机变量X的分布列为

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \dfrac{a-1}{4} & \dfrac{a+1}{4} & 0.1 & 0.2 & 0.1 \end{pmatrix}$$

$P\{-1 \le X \le 2\}$是 [填空1]

作答

Example 2.6 设随机变量X的分布列为

$$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \dfrac{a-1}{4} & \dfrac{a+1}{4} & 0.1 & 0.2 & 0.1 \end{pmatrix}$$

求$P\{-1 \le X \le 2\}$

此时我们跳过了样本空间这件事！

解 (1) 由 $$\frac{a-1}{4}+\frac{a+1}{4}+0.1+0.2+0.1=1$$

解得 $a=1.2$

故分布列为 $$\begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0.05 & 0.55 & 0.1 & 0.2 & 0.1 \end{pmatrix}$$

(2) $$P\{-1 \le X \le 2\} = \sum_{-1 \le X \le 2} p(x_i)$$

$$= 0.55 + 0.1 + 0.2 + 0.1$$

$$= 0.95$$

主观题 10分

在伯努里试验中，每次成功的概率为p。 记直至得到第r次成功时的试验次数为X，求X的分布。

作答

Example 巴斯卡分布 在伯努里试验中，每次成功的概率为p。记直至得到第r次成功时的试验次数为X，求X的分布。

解：P{X＝k}＝P{前k-1次试验中有r-1次成功，有k-r次不成功，且第k次成功}

$$= C_{k-1}^{r-1} p^{r-1} (1-p)^{k-r} p$$

$$= C_{k-1}^{r-1} p^{r} (1-p)^{k-r} \qquad (k = r, r+1, r+2, \cdots\cdots)$$

2.2.3 常见的概率分布

（一）退化分布（单点分布）

设随机变量X只取一个常数值c， 即

$$P(X=c)=1$$

称它为退化(degenerate)分布，又称为单点分布。

事实上，X可以看作一个常数，但有时我们宁愿把它看作（退化的）随机变量。

（二）0-1分布（两点分布/Bernoulli分布）

随机变量X只取两个值0和1，满足分布：

$$P(X=0)=1-p \qquad p>0$$

$$P(X=1)=p$$

称X服从参数为p的 0-1分布。

Example 2.7 2000件产品，有1990件合格品，10 件次品。从中随机抽一件，规定e_1＝{取得的是合格品}，e_2＝{取得的是次品}

定义随机变量：X表示取得次品的个数

$$X(e_1)=0, X(e_2)=1$$

$$p=10/2000=0.05$$

X服从参数0.05的0-1分布。

[1] 伯努里试验具有广泛性： 电路‘断’与‘不断’，产品‘合格’与‘不合格’，种子‘发芽’与‘不发芽’，掷硬币得‘正面’与‘反面’，…

[2] 任一伯努里试验（具有广泛性）的结果都可用伯努里分布描述

2.2.3 常见的概率分布

（三）二项分布

对于n重伯努里（Bernoulli）试验，假设每次成功的概率是p，定义随机变量X为n次试验中事件A可能发生的次数

$$P\{X=k\}=b(k,n,p)$$

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

称X服从参数为n和p的二项分布(Binomial distribution)，记为X~b(n,p)或X~B(n,p).

（三）二项分布

二项分布是概率论中最重要的分布之一，应用很广

(1)考察某地n个人是否患某种非流行性疾病，患病人数X服从二项分布.
检查一人是否患某种非流行性疾病是一次伯努里试验，各人是否生这病可认为相互独立，并可近似认为患病的概率p相等。

(2) 保险公司对某种灾害 (自行车被盗，火灾，…) 保险
各人发生此种灾害与否可认为相互独立，并假定概率相等。设一年间一人发生此种灾害的概率为p， 则在参加此种保险的n人中发生此种灾害的人数X服从二项分布。

（三）二项分布

(3) 碰运气能否通过英语四级考试

$$P\{X=k\}=b(k,85,0.25)$$

$$P\{X\geq 51\}=b(51,85,0.25)+b(52,85,0.25)+\cdots+b(85,85,0.25)$$

$$\approx 8.74\times 10^{-12}$$

(4) 机房内n台同类计算机，在一年内每台损坏的概率为p， 则在一年时间内损坏的机器数X服从二项分布。

如此，我们要问这样的分布具有什么样的性质呢？

比如： 最有可能会坏多少台机器？

为保证工作，至少备多少台机器？

二项分布的重要性质

[1] $$b(k,n,p)=b(n-k,n,1-p)$$

这从二项分布的概率以及 $C_n^k=C_n^{n-k}$ 立即可得。

也可以这样理解：n次试验中，事件 {k次成功} 与事件 {n-k次不成功} 是同样的，而 {n-k次不成功} 的概率即为

$b(n-k,n,1-p)$

很多情况下二项分布的计算很复杂，有时备有相应的计算表格，但只限于$p\le 0.5$的情况，当$p>0.5$时就可利用上式来计算。

[2] 增减性以及最可能成功（发生）次数

对固定的n、p， 由于

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

$$=\frac{(n-k+1)p}{k(1-p)}$$

$$=1+\frac{(n+1)p-k}{k(1-p)}$$

上式是否大于1，主要看(n＋1)p－k的正负，或者说看(n＋1)p与k的大小问题。

当 $k<(n+1)p$ 时，$\frac{b(k,n,p)}{b(k-1,n,p)}>1$，$b(k,n,p)$单调递增；

当 $k>(n+1)p$ 时，$\frac{b(k,n,p)}{b(k-1,n,p)}<1$，$b(k,n,p)$单调递减；

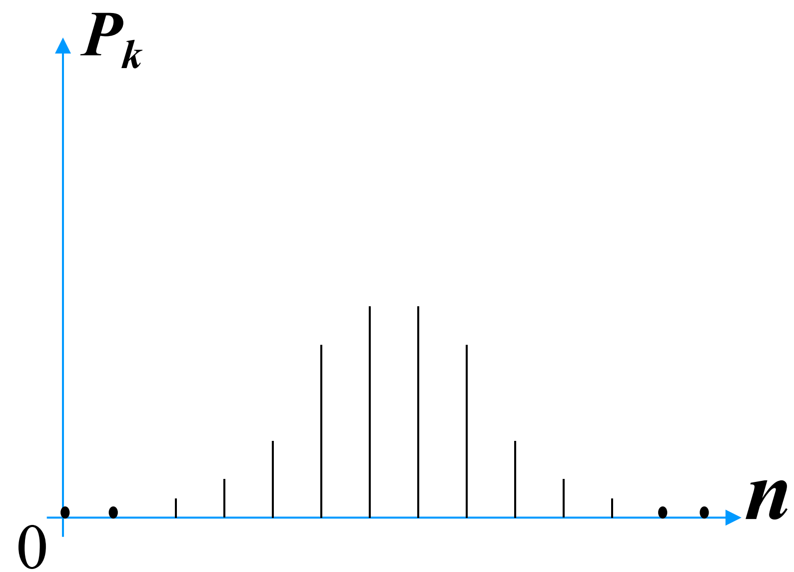
[2] 增减性以及最可能成功（发生）次数

如果(n＋1)p是整数，k＝(n＋1)p，$b(k,n,p)=b(k-1,n,p)$达到最大值;

我们称m＝(n＋1)p或(n＋1)p－1为最可能成功次数；

如果(n＋1)p不是整数，最可能成功次数为 $m=[(n+1)p]$

直观推断：概率p是统计得到的，现在做n次试验，最可能的成功次数应该是np附近。



n=13, p=0.5

[3] $n\to\infty$ 时的渐近性质

假定p与n有关，记作p_n。 考虑$n\to\infty$的情况，有下面的定理：

[泊松(Poisson)定理] 如果存在正常数λ，当 $n\to\infty$ 时有 $np_n\to\lambda$，则

$$\lim_{n\to\infty} b(k,n,p)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,2,\cdots$$

[3] n →∞ 时的渐近性质

Proof: 记$\lambda_n = np_n$

$$b(k,n,p) = C_n^k (p_n)^k (1-p_n)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k \left(1-\frac{\lambda_n}{n}\right)^{(n-k)}$$

$$= \frac{\lambda_n^k}{k!}\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\left(1-\frac{\lambda_n}{n}\right)^n \Big/ \left(1-\frac{\lambda_n}{n}\right)^k$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow$$

$$\boxed{\frac{\lambda^k}{k!}} \qquad \boxed{1} \qquad \boxed{e^{-\lambda}} \qquad \boxed{1}$$

$$\to \frac{\lambda^k}{k!}e^{-\lambda} \qquad (n \to \infty)$$

[3] n →∞ 时的渐近性质

$$\lim_{n\to\infty} b(k,n,p)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,2,\cdots$$

定理的条件 $np_n\to\lambda$ 意味着当 n很大时，p_n 必定很小.

因此，利用泊松定理，对于二项分布，当 n 很大，p 很小时有以下近似式：

$$C_n^k\, p^k(1-p)^{n-k}\approx\frac{\lambda^k e^{-\lambda}}{k!}$$

其中 $\boldsymbol{\lambda = np}$

实际计算中，$n\geq 100$，$np\leq 10$ 时近似效果就很好.

请同学们利用编程，绘图观察两者之间差距的变化
（绘图是以n作为变量，而不是b(k,n,p)中的k）

思考题： 食堂开设多少窗口合理

某学校有500人，拟建一个食堂，开设m个窗口。窗口数m太小，则经常派长队；窗口数m太大，则不经济。

假定在每一个指定时刻，500个人中每人是否去食堂是独立的，每人去食堂的概率都是0.1。

问题：” 在营业中任意时刻，保证每个窗口的排队人数（包括正在打菜的那个人）不超过10”这个事件的概率不小于0.9, 则至少需开设多少个窗口?