《初等概率论》第 5 讲

邓 婉 璐

清华大学
统计学研究中心

September 29, 2018

一、随机变量

♣ 为什么引进“随机变量”

1. (复杂性) ‘事件’是用来描述随机试验的某些现象是否出现的，要说明比较复杂的试验结果，需要定义许多事件；

2. (方便性) 在许多概率模型中，试验结果是数值化的，比如仪器的仪表板的读数、股票的价格、气温等；还有其它的试验结果虽不是数值化的，但这些试验结果与某些数值紧密相连.

3. (全面性) 对随机试验来说，不仅关心试验出现什么结果，更重要的是要知道这些结果将以什么概率出现.

$(\mathbb{R}, \mathcal{B})$ $f^{-1}(B) = A$

Ω $(\Omega, \mathcal{F}, P)$ $f: \Omega \to \mathbb{R}$ $f(x) =$

一、随机变量

例 1.1

设 Ω 是样本空间，对于事件 A，**示性函数** (*indicator function*) I_A 是 Ω 上的函数，即：

$$I_A(\omega) = \begin{cases} 1, & \text{当 } \omega \in A, \\ 0, & \text{当 } \omega \notin A. \end{cases}$$

对任意的 $x \in R$，函数 $I_A(\omega)$ 满足：

$$\{\omega | I_A(\omega) \le x\} = \begin{cases} \emptyset, & \text{当 } x < 0, \\ A^c, & \text{当 } x \in [0,1), \\ \Omega, & \text{当 } x \ge 1. \end{cases}$$

于是无论 x 取何值，$\{I_A \le x\} \overset{\text{def}}{=} \{\omega | I_A(\omega) \le x\}$ 都是事件.
将 Ω 上的具有这种性质的函数称为随机变量.

一、随机变量

由于可以对 I_A 进行数学运算，所以随机变量的引入会带来许多方便。比如：

命题 1.1

证明： $(A \triangle B) \triangle C = A \triangle (B \triangle C)$, 其中 A, B, C 为事件, $\triangle$ 表示事件的对称差运算.

证明. 注意到：

$$I_{(A\triangle B)\triangle C} = |I_{A\triangle B} - I_C| = ||I_A - I_B| - I_C|,$$
$$I_{A\triangle(B\triangle C)} = |I_A - I_{B\triangle C}| = |I_A - |I_B - I_C||.$$

如果 $I_B = 0$, 则 $I_{(A\triangle B)\triangle C} = |I_A - I_C| = I_{A\triangle(B\triangle C)}$.
如果 $I_B = 1$, 则 $I_{(A\triangle B)\triangle C} = |1 - I_A - I_C| = I_{A\triangle(B\triangle C)}$.

一、随机变量

例 1.2

在一副扑克的 52 张中任取 1 张，样本空间的每个点表示 1 张扑克．用 X 表示所取扑克的大小．如：$X = 3$ 表示所取到的扑克是 3，如将 X 视为样本空间上的函数，即 $X : \Omega \mapsto \mathbb{N}$，则

$$\{X = 3\} \stackrel{\text{def}}{=} \{\omega | X(\omega) = 3\} = \{\clubsuit 3,\ \diamondsuit 3,\ \heartsuit 3,\ \spadesuit 3\}.$$

显然，对 $\forall x \in R$，$\{X \le x\}$ 是一个事件．称 X 是**随机变量**．由于 X 的取值是数，所以可对 X 进行数学运算．

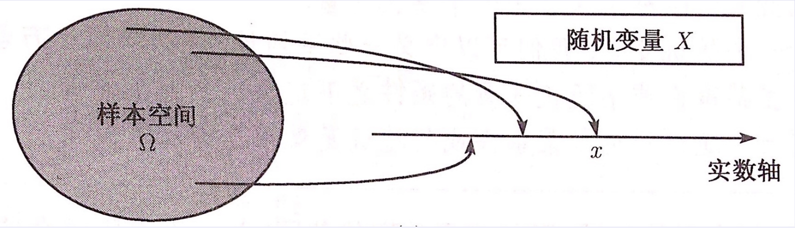
例 1.3

人的血压总在不断的变化之中．设甲的收缩压的变化范围是 $90 - 180$mmHg．用 X 表示甲的收缩压的测量结果，称 X 是**随机变量**．$X = 125$ 表示甲的收缩压是 125mmHg．这里试验的样本空间是 $\Omega = [90, 180]$，随机变量 X 是 Ω 上的函数，满足 $X(\omega) = \omega$．

一、随机变量

♣ “随机变量”是什么？

- 给每个可能的试验结果分配一个数. 即，一个从样本空间 Ω 到实数轴 $\mathbb{R}$ 的函数. 离散、连续.



讨论：

- 对同一个样本空间，可以定义不同的随机变量.
- 随机性来自哪里？
- 本质上，是一个试验某一方面的数值上的总结.
- 解决问题？例，学生、身高、$[165, 170]$ 的概率.

一、随机变量

定义：随机变量 (random variable, r.v., rv)

设 $(\Omega, \mathscr{F})$ 是可测空间，如果 Ω 上的函数 $X(\omega)$ 满足：对 $\forall x \in R$, $\{\omega | X(\omega) \le x\} \in \mathscr{F}$，就称 $X(\omega)$ 是 $(\Omega, \mathscr{F})$ 上的随机变量，简称随机变量. 通常将随机变量 $X(\omega)$ 简记为 X.

习惯上用大写的 $X, Y, Z, \varepsilon, \xi, \eta$ 等表示随机变量，不够用的时候可以使用下标，比如 $X_1, X_2, \cdots$.

以后用 $\{X \le x\}$ 表示事件 $\{\omega | X(\omega) \le x\}$.

♣ 显然，

1. $\{X > x\} = \{X \le x\}^c \in \mathscr{F}$;
2. $\{X < x\} = \cup_{n=1}^{\infty} \{X \le x - \frac{1}{n}\} \in \mathscr{F}$;
3. $\{X \ge x\} = \{X < x\}^c \in \mathscr{F}$.

一、随机变量

用 R 表示全体实数，用 $\mathcal{C}$ 表示 R 中左开右闭的子区间的全体，即 $\mathcal{C}=\{(a,b]\}$. 令 $\mathcal{B}=\sigma(\mathcal{C})$，通常称 $\mathcal{B}$ 称为 **Borel** 域，称 $\mathcal{B}$ 的元素为 **Borel** 集.

当 $(\Omega,\mathcal{F},\mathbb{P})$ 是概率空间，下面的定理表明对任何 Borel 集 A，$\{X\in A\}$ 都是事件，于是可以计算概率 $\mathbb{P}(X\in A)$.

定理 1.1

设 X 是可测空间 $(\Omega,\mathscr{F})$ 上的随机变量，则对任意的 Borel 集 A，有

$$\{X\in A\}\overset{\text{def}}{=}\{\omega|X(\omega)\in A\}\in\mathscr{F}.$$

证明. 对 $A\subset R$，令

$$X^{-1}(A)=\{\omega|X(\omega)\in A\},\quad \mathscr{A}=\{A|X^{-1}(A)\in\mathscr{F},A\subset R\}.$$

下面证明： $\mathscr{A}$ 是 σ-域，且 $\mathcal{B}\subset\mathscr{A}$.

(a). $X^{-1}(R)=\{\omega|X(\omega)\in R\}=\Omega\in\mathscr{F}$，故 $R\in\mathscr{A}$;

一、随机变量

(b). 对任意的 $A \in \mathscr{A}$, 有 $X^{-1}(A) \in \mathscr{F}$, 于是

$$X^{-1}(A^c) = \{\omega | X(\omega) \in A^c\} = \{\omega | X(\omega) \in A\}^c = [X^{-1}(A)]^c \in \mathscr{F}.$$

故 $A^c \in \mathscr{F}$.

(c). 对 $A_j \in \mathscr{A}$, 有 $X^{-1}(A_j) \in \mathscr{F}$, $j = 1, 2, \ldots$, 从而

$$\begin{aligned} X^{-1}(\cup_{i=1}^{\infty} A_i) &= \{\omega | X(\omega) \in \cup_{i=1}^{\infty} A_i\} \\ &= \cup_{i=1}^{\infty} \{\omega | X(\omega) \in A_i\} \\ &= \cup_{i=1}^{\infty} X^{-1}(A_i) \in \mathscr{F}. \end{aligned}$$

故 $\cup_{i=1}^{\infty} A_i \in \mathscr{A}$. 因此 $\mathscr{A}$ 是 σ-域.

对 $(a, b] \in \mathcal{C}$, $X^{-1}((a, b]) = \{a < X \le b\} = \{X \le b\} - \{X \le a\} \in \mathscr{F}$. 所以 $(a, b] \in \mathscr{A}$, 于是 $\mathcal{C} \subset \mathscr{A}$. 因此, $\mathcal{B} = \sigma(\mathcal{C}) \subset \mathscr{A}$. ■

一、随机变量

♣ 注：在定义中，“$\{X \le x\} \in \mathscr{F}$”可用其它形式代替，如：

1. $\{X > x\} \in \mathscr{F}$;
2. $\{X < x\} \in \mathscr{F}$;
3. $\{X \ge x\} \in \mathscr{F}$.

$(-\infty, x]$

$\mathcal{B}$

一、随机变量

例 1.4

在 52 张扑克中任取 13 张，求这 13 张牌中恰好有 5 张梅花的概率.

解. 用 X 表示"这 13 张牌中梅花的张数"，则 $X=5$ 是关心的事件，易得 $\mathbb{P}(X=5)=\binom{13}{5}\binom{39}{8}/\binom{52}{13}$. 为了弄清 $X(\omega)$，设试验的样本空间 $\Omega=\{\omega\}$，即 Ω 由 $\binom{52}{13}$ 个样本点组成，每个样本点 ω 是不分次序的 13 张牌. $X(\omega)$ 是定义在 Ω 上的函数：

$$X(\omega)=\text{“}\omega\text{中的梅花数”},\quad \omega\in\Omega.$$
$$\{X=5\}=\{\omega|\omega\text{中有 5 张梅花}\}.$$

尽管将 X 的函数关系表示出来了，但对问题 $\mathbb{P}(X=5)=?$ 的解决并没有很多的帮助. 因此，人们并不十分看重函数 $X=X(\omega)$ 在 Ω 上是如何具体定义的，只是在必要的时候才将自变元 ω 写出来.

一、随机变量

♣ 概念

1. n 维 Borel 域
2. Borel 可测函数（简称可测函数）

连续函数、阶梯函数、单调函数以及这些函数的线性组合都是可测函数

一、随机变量

定理 1.2

如果 X 是可测空间 $(\Omega,\mathscr{F})$ 上的随机变量，$g(x)$ 是可测函数，则 $Y=g(X)$ 是 $(\Omega,\mathscr{F})$ 上的随机变量.

证明. 只要证对任意的 $a\in R$，对 $\{Y\le a\}\in\mathscr{F}$. 对任意给定的 a，定义

$$B=\{x|g(x)\le a\}$$

则 $B\in\mathcal{B}$. 注意到 $Y(\omega)=g(X(\omega))$ 是 Ω 上的函数，再利用定理1.1得到

$$\begin{aligned}\{Y\le a\}&=\{\omega|Y(\omega)\le a\}=\{\omega|g(X(\omega))\le a\}\\&=\{\omega|X(\omega)\in B\}=\{X\in B\}\in\mathscr{F}.\end{aligned}$$

■

一、随机变量

完全类似的可以证明，如果 $X_1, X_2, ..., X_n$ 都是可测空间 $(\Omega, \mathscr{F})$ 上的随机变量，$g(x_1, x_2, ..., x_n)$ 是 n 元可测函数，则 $Y = g(X_1, X_2, ..., X_n)$ 是 $(\Omega, \mathscr{F})$ 上的随机变量.

无特殊说明，本课以后用到的一元和多元实函数都是指可测函数。

♣ 构造随机变量的方式

1. 四则运算：$aX + bY$, XY, X/Y, 对 $a, b \in R$;
2. 极限：$\sup\limits_n X_n$; $\inf\limits_n X_n$; $\limsup\limits_n X_n$; $\liminf\limits_n X_n$; $\lim\limits_n X_n$(如果存在);
3. 变换：$\max\{X, Y\}$, $\min\{X, Y\}$, $\sin(X)$, e^X, X^+, X^-, $|X|^p$ 等;
4. 函数复合.

二、随机变量的独立性

定义：随机变量的相互独立性

设 $X_1, \ldots, X_n$ 是 $(\Omega, \mathscr{F})$ 上的随机变量，如果对任意的实数 $x_1, \ldots, x_n$ 都有

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \cdots \mathbb{P}(X_n \le x_n),$$

称随机变量 $X_1, \ldots, X_n$ 相互独立.

$A_1 \times \cdots \times A_n \quad \mathcal{B}$

定义：独立序列

如果对任意的 n, $X_1, \ldots, X_n$ 相互独立，则称随机变量序列 $\{X_i\}$ 相互独立，此时称 $\{X_i\}$ 为独立序列.

$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n)$

$= P(X_1 \in A_1) P(X_2 \in A_2) \cdots P(X_n \in A_n)$

二、随机变量的独立性

定理 2.1

设 $X_1, \ldots, X_n$ 相互独立，则对任何 Borel 集 $A_1, \ldots, A_n$，事件

$$\{X_1 \in A_1\}, \quad \{X_2 \in A_2\}, \quad \ldots, \quad \{X_n \in A_n\}$$

相互独立.

定理 2.2

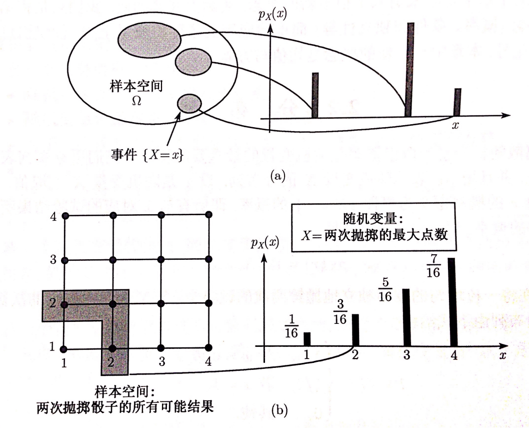
设随机变量 $X_1, \ldots, X_n$ 相互独立，$g_1(x), \ldots, g_n(x)$ 是一元实可测函数，$\varphi(x_1, x_2, \ldots, x_k)$ 是 k 元实可测函数，则

1. 随机变量 $g_1(X_1), \ldots, g_n(X_n)$ 相互独立；
2. 随机变量 $\varphi(X_1, X_2, \ldots, X_k), X_{k+1}, \ldots, X_n$ 相互独立.

三、离散型随机变量

定义：离散型随机变量 (discrete random variable)

如果随机变量 X 只取有限个值 $x_1, \dots, x_m$ 或者可列个值 $x_1, x_2, \cdots$，则称 X 是离散型随机变量，简称离散随机变量.



三、离散型随机变量

定义：概率分布、概率分布列

设 X 是离散型随机变量，称

$$\mathbb{P}(X = x_k) = p_k, \quad k \geq 1$$

为 X 的概率分布，称 $\{p_k\}$ 是概率分布列，简称为分布列 (Probability mass function, PMF).

当分布列 $\{p_k\}$ 的规律性不够明显时，概率分布常写为：

X	x_1	x_2	x_3	$\cdots$
$\mathbb{P}$	p_1	p_2	p_3	$\cdots$

♣ 分布列 $\{p_k\}$ 满足以下性质：

1. $p_k \geq 0$；
2. $\sum_{k=1}^{\infty} p_k = 1$.

三、离散型随机变量

♣ 常用的几种离散型随机变量及其概率分布

1. 两点分布 $B(1,p)$【Bernoulli 分布】

任何试验，当只考虑成功 (质量合格、健康状态、政治态度、电话机待机状态等) 与否时，就可以用如下随机变量描述：

$$X = \begin{cases} 1, & \text{试验成功}, \\ 0, & \text{试验不成功}. \end{cases}$$

定义：Bernoulli 分布

如果 X 只取值 0 或 1，概率分布是 $\mathbb{P}(X=1)=p=1-\mathbb{P}(X=0)$，则称 X 服从两点分布 (Bernoulli 分布)，记作 $X \sim B(1,p)$.

其分布列为：

$$\begin{array}{c|cc} X & 0 & 1 \\ \hline \mathbb{P} & 1-p & p \end{array}.$$

2. 二项分布 $B(n,p)$

设试验 S 成功的概率是 p, 将试验 S 重复 n, 用 X 表示成功的次数, 求 $P(X=k)$.

解：用 A_j 表示第 j 次试验成功，则 $A_1, A_2, ..., A_n$ 相互独立，且 $\mathbb{P}(A_j)=p$. 从 n 次试验中选定 k 次试验的方法共有 $\binom{n}{k}$ 种. 对第 i 种取法，设其对应着 $\{j_1, j_2, ..., j_k\}$, 用

$$B_i = A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_k} \cap A^c_{j_{k+1}} \cap \cdots \cap A^c_{j_n}$$

表示第 $\{j_1, j_2, ..., j_k\}$ 次试验成功，其余不成功，则 $\{B_i\}$ 互不相容，并且

$$\{X=k\} = \bigcup_{i=1}^{\binom{n}{k}} B_i, \quad \mathbb{P}(B_i) = p^k(1-p)^{n-k}.$$

$\omega: X(\omega)=k.$

利用概率的有限可加性，可得

$$\mathbb{P}(X=k)=\sum_{i=1}^{\binom{n}{k}}\mathbb{P}(B_i)=\binom{n}{k}p^k(1-p)^{n-k}.$$

发现：$\binom{n}{k}p^k(1-p)^{n-k}$ 为二项展开式

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

的第 $k+1$ 项.

三、离散型随机变量

定义：Binomial 分布

如果随机变量 X 的概率分布如下：

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\dots,n,$$

则称 X 服从二项分布，其中 $p\in(0,1)$，记作 $X\sim B(n,p)$. B 是 Binomial 的首字母.

♣ 如果随机变量 $X_1,\dots,X_n$ 相互独立，都服从两点分布 $B(1,p)$，则 $S=X_1+\dots+X_n\sim B(n,p)$.

♣ 如果随机变量 X,Y 相互独立，且 $X\sim B(n,p)$, $Y\sim B(m,p)$，则 $X+Y\sim B(m+n,p)$.

三、离散型随机变量

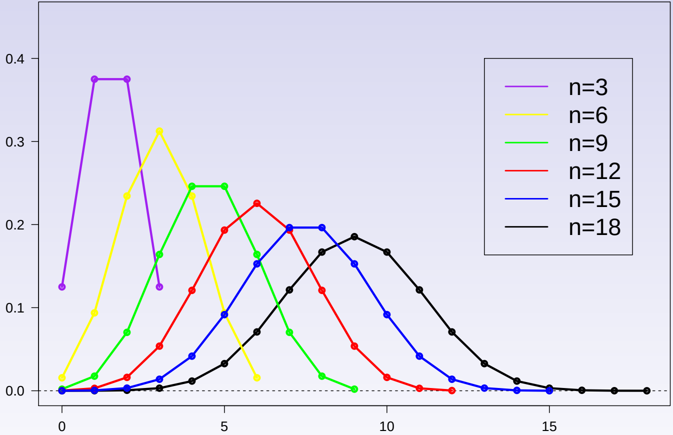
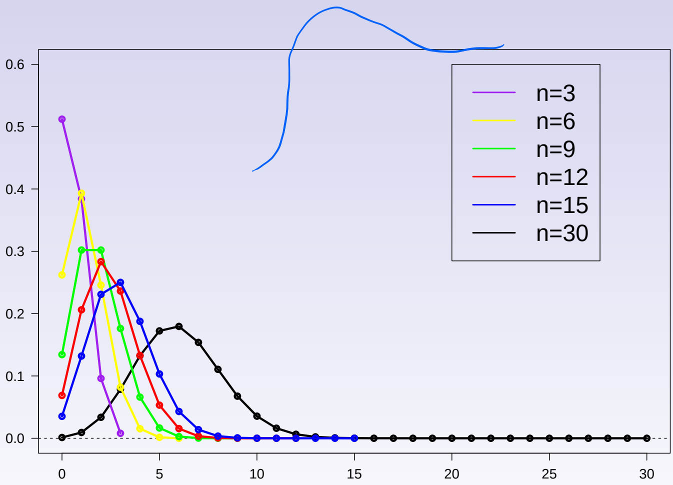


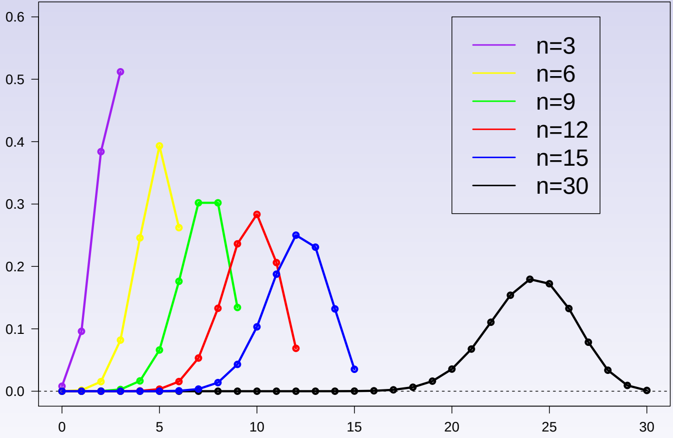
图: 二项分布的折线图，其中 $p = 0.5$.

三、离散型随机变量



图：二项分布的折线图，其中 $p = 0.2$.

三、离散型随机变量



图：二项分布的折线图，其中 $p = 0.8$.

三、离散型随机变量

♣ 令 $b(k,n,p)=\mathbb{P}(X=k)$. 则

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

故

1. 当 $k<(n+1)p$ 时，则 $b(k,n,p)>b(k-1,n,p)$，此时后项大于前项，$b(k,n,p)$ 随 k 的增大而增大;
2. 当 $k>(n+1)p$ 时，则 $b(k,n,p)<b(k-1,n,p)$，此时后项小于前项，$b(k,n,p)$ 随 k 的增大而减小;
3. 当 $k=(n+1)p$ 时，则 $b(k,n,p)=b(k-1,n,p)$，此时 $b((n+1)p,n,p)$ 与 $b((n+1)p-1,n,p)$ 两项相等且为最大.

三、离散型随机变量

定理 3.1

二项分布的最大可能值 k_0 存在，即满足

$$b(k_0, n, p) = \max_{0\le k\le n} b(k, n, p)$$

的整数 k_0 存在，且

$$k_0 = \begin{cases} (n+1)p,\ (n+1)p-1, & \text{如果 } (n+1)p \text{ 为整数}, \\ [(n+1)p], & \text{如果 } (n+1)p \text{ 不是整数}, \end{cases}$$

即，(a) 当 $(n+1)p$ 为整数时，$b((n+1)p, n, p)$ 与 $b((n+1)p-1, n, p)$ 均为最大项，(b) 当 $(n+1)p$ 不为整数时，$b([(n+1)p], n, p)$ 为唯一的最大项.

一般称 $b(k_0, n, p)$ 为二项分布的中心项.

三、离散型随机变量

例 3.1

9 人同时向同一目标各打一枪，如果每个人射击是相互独立的且每人射击一次击中的概率均为 0.3，求有两人以上击中目标的概率以及最可能击中目标的人数.

解. 设 X 为击中目标的人数，则 $X \sim B(9, 0.3)$. 所求概率为

$$\begin{aligned}\mathbb{P}(X > 2) &= \sum_{k=3}^{9} \binom{9}{k} (0.3)^k (1-0.3)^{9-k} \\ &= 1 - \sum_{k=0}^{2} \binom{9}{k} (0.3)^k (1-0.3)^{9-k} \\ &= 1 - 0.4624 = 0.5376.\end{aligned}$$

又因为 $(n+1)p = 10 \times 0.3 = 3$ 为整数，故最可能击中的目标人数是 3 或者 2.

三、离散型随机变量

③ **几何分布** $G(p)$

例 3.2

甲向一个目标射击，直到击中为止．用 X 表示首次击中目标时的射击次数．如果甲每次击中的概率是 $p>0$，求 $P(X=k)$．

解．用 A_j 表示甲第 j 次没击中目标，由 $\{A_j\}$ 的独立性知

$$\begin{aligned}\mathbb{P}(X=k) &= \mathbb{P}(A_1\cap\cdots\cap A_{k-1}\cap A_k^c)\\ &= (1-p)^{k-1}p, \quad k=1,2,\cdots.\end{aligned}$$

♣ 注意到 $\mathbb{P}(X<\infty)=\sum_{k=1}^{\infty}(1-p)^{j-1}p=1$，可知甲一直射击下去总有击中目标的时候．这相当于独立重复试验一直做下去，总有成功的时候．

三、离散型随机变量

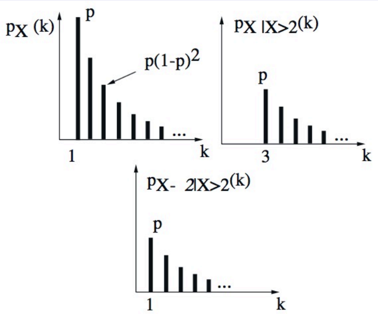
定义：Geometric 分布

如果随机变量 X 的概率分布如下：

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

则称 X 服从参数为 p 的几何分布. $X \sim G(p)$.

♣ 名字来源：几何级数



定理 3.2

取正整数值的随机变量 $X \sim G(p)$ 的**充要条件**是 X 有无记忆性，即对每个 $k \geq 1$,

$$\mathbb{P}(X = k+1 | X > k) = \mathbb{P}(X = 1).$$

证明. $(\Longrightarrow)$. 由条件概率公式得

$$\begin{aligned}\mathbb{P}(X = k+1 | X > k) &= \frac{\mathbb{P}(X = k+1, X > k)}{\mathbb{P}(X > k)} \\ &= \frac{\mathbb{P}(X = k+1)}{\mathbb{P}\left(\bigcup_{j=k+1}^{\infty}\{X = j\}\right)} \\ &= \frac{(1-p)^k p}{\sum\limits_{j=k+1}^{\infty}(1-p)^{j-1}p} = p \\ &= \mathbb{P}(X = 1).\end{aligned}$$

$(\Longleftarrow)$. 设 $r_k = \mathbb{P}(X > k)$, 利用 $\{X = k+1\} = \{X > k\} - \{X > k+1\}$, 得

$$p = \mathbb{P}(X = k+1 | X > k) = \frac{\mathbb{P}(X > k) - \mathbb{P}(X > k+1)}{\mathbb{P}(X > k)}$$
$$= 1 - \frac{r_{k+1}}{r_k}.$$

利用 $r_0 = \mathbb{P}(X > 0) = 1$, 可得

$$r_{k+1} = (1-p) r_k = \ldots = (1-p)^{k+1}.$$

因此

$$\begin{aligned}\mathbb{P}(X = k) &= \mathbb{P}(X > k-1) - \mathbb{P}(X > k) \\ &= r_{k-1} - r_k \\ &= (1-p)^{k-1} p.\end{aligned}$$

三、离散型随机变量

4. 帕斯卡分布 (Pascal distribution)

例 3.3

甲向一个目标射击，直到击中 r 次为止. 用 X 表示射击停止时的射击次数. 如果甲每次击中的概率是 $p \in (0,1)$，求 $P(X=k)$.

定义：Pascal 分布

如果随机变量 X 的概率分布如下：

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

则称 X 服从帕斯卡分布.

当 $r=1$ 时，帕斯卡分布就是几何分布.

三、离散型随机变量

5. 负二项分布 $\mathrm{NB}(r,p)$

♣ 在上述例子3.3中，令 $Y=X-r$，则 Y 是射击停止时，射击失败的次数.

定义：负二项分布

如果随机变量 Y 的概率分布如下：

$$\mathbb{P}(Y=k)=\binom{k+r-1}{r-1}(1-p)^k p^r, \quad k=0,1,\dots, \tag{1}$$

则称 $Y\sim \mathrm{NB}(r,p)$.

三、离散型随机变量

♣ 名称来源

(a). “负指数二项展开式”

$$
\begin{aligned}
(1-x)^{-r} &= \sum_{k=0}^{\infty} \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!}(-x)^k \\
&= \sum_{k=0}^{\infty} \frac{(r+k-1)(r+k-2)\cdots(r+1)r}{k!}x^k \\
&= \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} x^k, \quad |x|<1.
\end{aligned}
$$

令 $x=1-p$ 并两边乘以 p^r，得

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

这验证了 (1) 确实是分布列.

三、离散型随机变量

(b). 试验方式不同.

与二项分布是“反其道而行之”: 二项分布是定下总抽样个数 n 而把废品个数 X 作为随机变量; 负二项分布则相反, 它定下废品个数 r 而把总抽样次数减去 r 作为随机变量.

例 3.4

为了检查某厂产品的废品率 p 大小, 有两个试验方案可采取: 一是从该厂产品中有放回地抽取若干个, 检查其中的废品数 X, 这一方案导致二项分布.
另一个方案是先指定一个自然数 r, 一个一个有放回地从该厂产品中抽样检查, 直到发现第 r 个废品为止. 以 X 记到当时为止已检出的合格品个数. 显然, 若废品率 p 小, 则 X 倾向于取较大的值, 反之, 则 X 倾向于取小值. 故 X 可用于研究 p 的目的.

三、离散型随机变量

6. 超几何分布 $H(n, M, N)$

例：在包含 N 个元素的总体中，M 个是红的，$N-M$ 个是黑的．任意选取 n 个元素组成一组．试求所取出的这一组中，恰有 k 个红元素的概率．

定义：超几何分布

如果随机变量 X 的概率分布如下：

$$\mathbb{P}(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k=0,1,\dots,\min\{n,M\},$$

则称 X 服从超几何分布 (Hypergeometric distribution)，记作 $X \sim H(n, M, N)$.

♣ 应用

- 质量检测
- 捕获-再捕获模型
- $\cdots$

例 3.5

N 件产品中恰有 M 件次品，从中任取 n 件，用 X 表示这 n 件中的次品数，则 X 服从超几何分布.

如果这批产品充分多，无放回的抽取和有放回的抽取就没有本质的差异. 无放回抽取时，$X \sim B(n, p_N)$，其中 $p_N = M/N$ 是次品率. 实际上，当 $\lim_{N\to\infty} M/N = p$ 时，有

$$
\begin{aligned}
\mathbb{P}(X = m) &= \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \\
&= \frac{M!}{m!(M-m)!}\,\frac{(N-M)!}{(n-m)!(N-M-n+m)!}\,\frac{n!(N-n)!}{N!} \\
&= \binom{n}{m}\frac{M(M-1)\cdots(M-m+1)}{N^m} \\
&\quad \times \frac{(N-M)\cdots(N-M-n+m+1)}{N^{n-m}}
\end{aligned}
$$

$$\times\frac{N^n}{N(N-1)\cdots(N-n+1)}\to\binom{n}{m}p^m(1-p)^{n-m}.$$

即，超几何分布可以用二项分布近似.

在实际问题中，对于较大的 N，计算 $\binom{N}{n},\binom{M}{m},\binom{N-M}{n-m}$ 要比计算 $\binom{n}{m}$ 费时多了. 因为抽样的个数 n 一般不会很大，所以对较大的 N 和 M，采用近似公式

$$\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}\approx\binom{n}{m}p_N^m(1-p_N)^{n-m}$$

往往会更方便.

三、离散型随机变量

7. 负超几何分布 $NH(n, M, N)$

例：在包含 N 个元素的总体中，M 个是红的，$N-M$ 个是黑的. 每次无放回地抽取一个元素，直到抽到 r 个红元素为止，此时共抽取了 k 个黑元素的概率.

定义：负超几何分布

如果随机变量 X 的概率分布如下：

$$\mathbb{P}(X=k)=\frac{\binom{k+r-1}{k}\binom{N-k-r}{M-r}}{\binom{N}{M}},\quad k=0,1,\dots,M,$$

则称 X 服从负超几何分布 (Negative hypergeometric distribution)，记作 $X\sim NH(M,N,r)$.

♣ 应用

- 质量检测
- 捕获-再捕获模型
- $\cdots$

三、离散型随机变量

8. Poisson 分布 $\mathcal{P}(\lambda)$

定义：Poisson 分布

如果随机变量 X 的概率分布如下：

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

则称 X 服从参数为 λ 的 Poisson 分布，记作 $X \sim \mathcal{P}(\lambda)$，其中 λ 是正常数.

♣ 实际应用：

- 某段高速公路上一年内的交通事故数；
- 某市场一天中到达的顾客次数；
- 某办公室一天中收到的电话数；
- 某大学一天中上课迟到的总人数；
- $\cdots$.

三、离散型随机变量

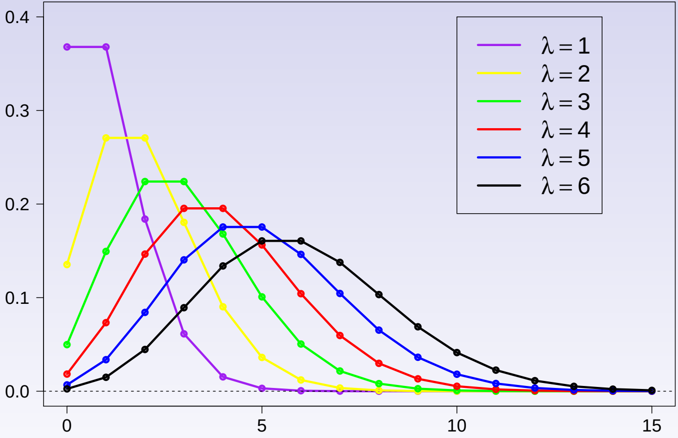


图: Poisson 分布的折线图.

三、离散型随机变量

《初等概率论》
第 5 讲

邓婉璐

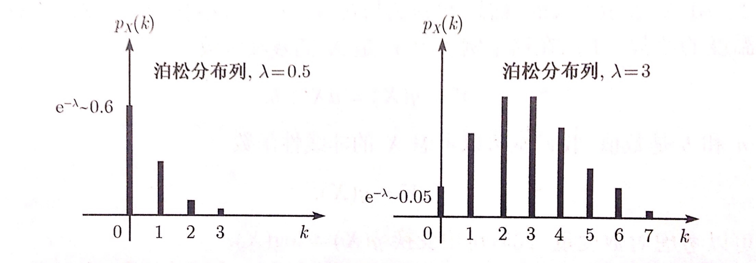
随机变量

随机变量的独立性

离散型随机变量

小结

作业



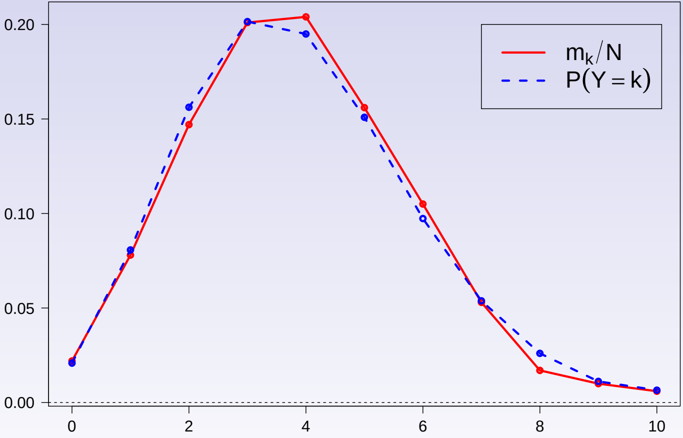
例 3.6

1910 年,著名科学家 Rutherford 和 Geiger 观察了放射性物质钋 (Polonium) 放射 α 粒子的情况，他们进行了 $N = 2608$ 次观测，每次观测 7.5 秒，一共观察到 10094 个 α 粒子放出. 下面是观测记录. 最后一列中的随机变量 $Y \sim \mathcal{P}(3.87)$, 其中 $3.87 = 10094/2608$ 是 7.5 秒中放射出 α 粒子的平均数. 用 X 表示这块放射性钋在 7.5 秒内放射出的 α 粒子数，下表的最后两列表明事件 $\{X = k\}$ 在 $N = 2608$ 次重复观测中发生的频率和 $\mathbb{P}(Y = k)$ 基本相同.

三、离散型随机变量

表：放射性物质钋放射 α 粒子数

观察到的 α 粒子数 k	观察到 k 个粒子的次数 m_k	发生的频率 m_k/N	$\mathbb{P}(Y=k)$
0	57	0.022	0.021
1	203	0.078	0.081
2	383	0.147	0.156
3	525	0.201	0.201
4	532	0.204	0.195
5	408	0.156	0.151
6	273	0.105	0.097
7	139	0.053	0.054
8	45	0.017	0.026
9	27	0.010	0.011
10+	16	0.006	0.007
总计	2608	0.999	1.00



图：例子中的频率 m_k/N 和概率 $\mathbb{P}(Y=k)$.

三、离散型随机变量

♣ 下面证明 $X \sim \mathcal{P}(\lambda)$. 设想将 $t = 7.5$ 秒等分成 n 段，每段是 $\delta_n = t/n$ 秒. 对充分大的 n，假定:

1. 在 δ_n 内最多只有一个 α 粒子放出，并且放出一个粒子的概率是 $p_n = \mu\delta_n = \mu t/n$, 这里 μ 是正常数;
2. 各个小时间段内是否放射出 α 粒子数相互独立.

在上述假定下，这块放射性物质放射出的粒子数 $X \sim B(n, p_n)$. 于是

$$\begin{aligned}
\mathbb{P}(X = k) &= \lim_{n\to\infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
&= \lim_{n\to\infty} \frac{(\mu t)^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\mu t}{n}\right)^{n-k} \\
&= \frac{(\mu t)^k}{k!} \mathrm{e}^{-\mu t}.
\end{aligned}$$

取 $\lambda = \mu t$, 得 $X \sim \mathcal{P}(\lambda)$.

三、离散型随机变量

♣ 上述例子中，我们验证了二项分布可以用 Poisson 分布近似的事实：如果 n 很大，p 很小，且 $np \approx \lambda$，其中 λ 是大小适当的固定常数，则可用 $\mathcal{P}(\lambda)$ 近似二项分布 $B(n, p)$，特别是当 n 无法确定时，就应该使用 Poisson 分布了.

小结

知识点

- 随机变量：随机性来自哪里？
- 随机变量的独立性
- 离散随机变量、分布列

技巧

- 类比法掌握新概念（e.g. 事件的独立性、随机变量的独立性）
- 利用实际情景理解不同分布列之间的关系

作业

课后习题（第二章）：3，6，7，10，Borel-Cantelli 引理的证明（选作）

祝大家国庆节快乐！