

Lec-5. 随机变量、离散型随机变量

主讲教师：吴利苏 (wulisu@sdust.edu.cn)

主 页：wulisu.cn

目录

1. 随机变量

2. 离散型随机变量的分布律

3. 典型离散型随机变量

- 0-1 分布
- 二项分布
- 泊松分布
- 几何分布

试验结果的量化

对样本空间 S 中的样本点 e 的描述一般有以下两种情况.

■ 量化的:

- ◆ 降雨量;
- ◆ 候车人数;
- ◆ 发生交通事故的次数;...

■ 非量化的:

- ◆ 明天天气 (晴, 多云...);
- ◆ 化验结果 (阳性, 阴性);...

试验结果的量化

对样本空间 S 中的样本点 e 的描述一般有以下两种情况.

■ 量化的:

- ◆ 降雨量;
- ◆ 候车人数;
- ◆ 发生交通事故的次数;...

■ 非量化的:

- ◆ 明天天气 (晴, 多云...);
- ◆ 化验结果 (阳性, 阴性);...

中心问题: 将实验结果数量化.

引例

例

在一袋中有红球, 白球, 任意取一个观察颜色.

$$S = \{\text{红}, \text{白}\},$$

$$\text{令 } X(e) = \begin{cases} 1 & e = \text{红}; \\ 0 & e = \text{白}. \end{cases}$$

引例

例

在一袋中有红球, 白球, 任意取一个观察颜色.

$$S = \{\text{红}, \text{白}\},$$

$$\text{令 } X(e) = \begin{cases} 1 & e = \text{红}; \\ 0 & e = \text{白}. \end{cases}$$

X 将样本空间 S 数量化

$$X: S \longrightarrow \{0, 1\}.$$

随机变量

定义

设随机实验的样本空间为 $S = \{e\}$, 若

$$X = X(e)$$

为定义在 S 上的实值单值函数, 则称 $X = X(e)$ 为随机变量.

注

- (1) 随机变量 $X: S \rightarrow \mathbb{R}$ 为一映射, 其自变量具有随机性;
- (2) 随机事件可表示为
$$A = \{e: X(e) \in I\} = \{X \in I\}, I \subset \mathbb{R};$$
- (3) 对于 $i \neq j$, 则必有 $\{X = i\} \cap \{X = j\} = \emptyset$, 一对一;
- (4) 一般用大写英文字母 X, Y, Z 或希腊字母 ξ, η 表示.

例

一枚硬币抛三次, 观察正反.

样本空间为

$$S = \{HHH, HHT, HTH, THH, HTT, TTH, THT, TTT\}.$$

若 X 表示 3 次中出现正面的次数, 则

- $A = \{\text{正面出现了一次}\} = \{HTT, TTH, THT\}$
 $= \{e : X(e) = 1\} = \{X = 1\};$
- $B = \{3 \text{ 次出现的情况相同}\} = \{X = 0 \text{ or } 3\};$
- $C = \{\text{正面至少出现一次}\} = \{X \geq 1\}.$

常见的两类随机变量:

- 离散型随机变量;
- 连续型随机变量.

离散型随机变量

定义

若随机变量 X 的取值为有限个或可数个, 则称 X 为离散型随机变量.

例

- 观察掷一个骰子出现的点数. X 的取值: 1, 2, 3, 4, 5, 6.
- 记 X 为连续射击命中时的射击次数.
 $X = 1, 2, \dots$

离散型随机变量的概率分布律 (简称分布律)

X	x_1	x_2	\cdots	x_k	\cdots
P	p_1	p_2	\cdots	p_k	\cdots

离散型随机变量的概率分布律 (简称分布律)

X	x_1	x_2	\cdots	x_k	\cdots
P	p_1	p_2	\cdots	p_k	\cdots

分布律的内容 { 随机变量的所有可能取值
取每个可能取值相应的概率

离散型随机变量的概率分布律 (简称分布律)

X	x_1	x_2	\cdots	x_k	\cdots
P	p_1	p_2	\cdots	p_k	\cdots

分布律的内容 $\left\{ \begin{array}{l} \text{随机变量的所有可能取值} \\ \text{取每个可能取值相应的概率} \end{array} \right.$

分布律的另一表示形式:

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

性质

- 非负性: $p_k \geq 0$;
- 规范性: $\sum_{k=1}^{\infty} p_k = 1$.

性质

- 非负性: $p_k \geq 0$;
- 规范性: $\sum_{k=1}^{\infty} p_k = 1$.

例

掷骰子.

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

例

设一汽车在开往目的地的道路上需要经过 4 组信号灯, 每组信号灯以 $\frac{1}{2}$ 的概率允许或禁止通过, 以 X 表示汽车首次停下时, 它已通过的信号灯的组数. (每组 2 信号灯的工作是相互独立的) 求 X 的分布律.

例

设一汽车在开往目的地的道路上需要经过 4 组信号灯, 每组信号灯以 $\frac{1}{2}$ 的概率允许或禁止通过, 以 X 表示汽车首次停下时, 它已通过的信号灯的组数. (每组 2 信号灯的工作是相互独立的) 求 X 的分布律.

解: 设 p 表示每组信号灯禁止汽车通过的概率.

X	0	1	2	3	4
P_k	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$

$$P\{X = k\} = (1-p)^k p = \frac{1}{2^{k+1}}, \quad k = 0, 1, 2, 3, 4. \quad \square$$

例

设随机变量 X 所有可能取值为 $1, 2, \dots, n$, 且 $P\{X = k\} = ak, k = 1, \dots, n$. 求 a .

例

设随机变量 X 所有可能取值为 $1, 2, \dots, n$, 且 $P\{X = k\} = ak, k = 1, \dots, n$. 求 a .

解: 由分布律的规范性,

$$\sum_{k=1}^n P\{X = k\} = \sum_{k=1}^n ak = a \frac{n(n+1)}{2} = 1,$$

得 $a = \frac{2}{n(n+1)}$.



例

设一袋中装有标号为 1, 2, 2, 3, 3, 3 数字的六个球. 现从中任取一球, 用 X 表示取到球上标有的数字. 求

(1) X 的分布律;

(2) $P\{X \leq \frac{2}{7}\}$, $P\{0.5 \leq X \leq 2\}$.

解: (1) X 的所有可能取值为 1, 2, 3.

$$P\{X = 1\} = \frac{1}{6}, P\{X = 2\} = \frac{1}{3}, P\{X = 3\} = \frac{1}{2}.$$

X	1	2	3
P_k	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

$$(2) P\{X \leq \frac{2}{7}\} = P(\emptyset) = 0,$$

$$P\{0.5 \leq X \leq 2\} = P\{X = 1\} + P\{X = 2\} = \frac{1}{2}. \quad \square$$

0-1 分布

定义

若 X 的概率分布律为

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

其中 $0 < p < 1$, 则称 X 服从**0-1 分布**或**两点分布**. 记为 $X \sim 0-1(p)$ 或者 $X \sim b(1, p)$.

0-1 分布

定义

若 X 的概率分布律为

X	0	1
P_k	$1 - p$	p

其中 $0 < p < 1$, 则称 X 服从**0-1 分布**或**两点分布**. 记为 $X \sim 0-1(p)$ 或者 $X \sim b(1, p)$.

其分布律亦可写为: $P\{X = k\} = p^k(1 - p)^{1-k}$,
 $k = 0, 1$.

若样本空间只包含两个元素 $S = \{e_1, e_2\}$, 则总能在 S 上定义一个服从 (0-1) 分布的随机变量:

$$X = \begin{cases} 0, & e = e_1; \\ 1, & e = e_2. \end{cases}$$

若样本空间只包含两个元素 $S = \{e_1, e_2\}$, 则总能在 S 上定义一个服从 (0-1) 分布的随机变量:

$$X = \begin{cases} 0, & e = e_1; \\ 1, & e = e_2. \end{cases}$$

比如:

- 质检是否合格
- 对新生婴儿的性别进行登记
- 检验种子是否发芽
- 考试是否通过
- 马路乱停车是否会受罚
- 表白是否成功

伯努利 (Bernoulli) 试验

只有两个可能结果的试验, 称为伯努利试验, 故两点分布有时也称为伯努利分布.

将一个伯努利试验独立重复进行 n 次, 称为 n 重伯努利试验.

伯努利 (Bernoulli) 试验

只有两个可能结果的试验, 称为伯努利试验, 故两点分布有时也称为伯努利分布.

将一个伯努利试验独立重复进行 n 次, 称为 n 重伯努利试验.

在 n 重伯努利试验中,

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

其中 $k = 0, 1, \dots, n$. 上式称为二项概率公式.

二项分布

定义

若 X 的概率分布律为

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

其中 $n > 0, 0 < p < 1$, 则称 X 服从参数 n, p 的**二项分布**, 记 $X \sim b(n, p)$.

二项分布

定义

若 X 的概率分布律为

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

其中 $n > 0, 0 < p < 1$, 则称 X 服从参数 n, p 的**二项分布**, 记 $X \sim b(n, p)$.

X	0	1	...	k	...	n
P	q^n	$C_n^1 p q^{n-1}$...	$C_n^k p^k q^{n-k}$...	q^n

$n = 1$ 时, 二项分布化为

$$P\{X = k\} = p^k(1 - p)^{1-k}, k = 0, 1$$

所以 (0-1) 分布为 $b(1, p)$.

例

在相同条件下相互独立地进行 5 次射击, 每次击中目标的概率为 0.6, 则击中次数 X 服从 $X \sim b(5, 0.6)$ 的二项分布.

例

按规定, 某种型号电子元件的使用寿命超过 $1500h$ 的为一级品. 已知某一大批产品的一级品概率为 0.2 , 现从中抽查 20 次, 问 20 只元件中恰有 k 只为一级品的概率.

解: 不放回抽样, 但由于总量很大, 抽查的数量相对于来说又很小, 可近似当作放回抽样.

我们把检查一只元件看它是否为一级品看成一次试验, 20 只看成 20 次伯努利试验, 以 X 记 20 只一级品的只数. $X \sim b(20, 0.2)$,

$$P\{X = k\} = C_{20}^k 0.2^k \cdot 0.8^{20-k}, \quad k = 0, 1, \dots, 20. \quad \square$$

$$P\{X = 0\} = 0.012, \quad P\{X = 1\} = 0.058, \quad P\{X = 4\} = 0.218, \\ P\{X = 5\} = 0.125,$$

当 k 增加时, $P\{X = k\}$ 先增到最大值, 随后减少.

对于固定的 n, p , 二项分布都具有这一性质.

例

某人进行射击, 假设每次命中率为 0.02, 独立射击 400 次. 求至少击中两次的概率.

例

某人进行射击, 假设每次命中率为 0.02, 独立射击 400 次. 求至少击中两次的概率.

解: 将一次射击看成一次试验. 设击中次数为 X , $X \sim b(400, 0.02)$.

$$P\{X = k\} = C_{400}^k 0.02^k \cdot 0.98^{400-k}, \quad k = 0, 1, \dots, 400.$$

$$\begin{aligned} P\{X \geq 2\} &= 1 - P\{X = 0\} - P\{X = 1\} = \\ &= 1 - 0.98^{400} - 400 \cdot 0.02 \cdot 0.98^{399} = 0.9972. \quad \square \end{aligned}$$

例

设有 80 台同类型设备, 各台工作是相互独立的, 发生故障的概率是 0.01, 且一台设备故障能由一人处理. 考虑两种配备维修工人的方案.

(1) 由 4 人维护, 每人 20 台;

(2) 3 人共同维护 80 台.

比较两种方法在设备发生故障时不能及时维护的概率.

解: (1), 以 X 记第 1 人维护的 20 台中同一时刻发生故障的台数, A_i 表示第 i 人维护的 20 台中发生故障时不能及时维修. 则 80 台中发生故障而不能及时维修的概率

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq P(A_1) = P\{X \geq 2\}$$

$$X \sim b(20, 0.01), P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - 0.99^{20} - 20 \cdot 0.01 \cdot 0.99^{19} = 0.0169.$$

(2), Y 记 80 台中同一时刻发生故障的台数.

$$Y \sim b(80, 0.01)$$

$$P\{Y \geq 4\} = 1 - \sum_{k=0}^3 C_{80}^k 0.01^k \cdot 0.99^{80-k} = 0.0087.$$

在后一种情况尽管任务重了, 但工作效率提高了. 工作方法很重要.

泊松 (Poisson) 分布

定义

若 X 的概率分布律为：

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

其中 $\lambda > 0$ 是常数, 则称 X 服从参数为 λ 的泊松分布, $X \sim \pi(\lambda)$.

- 非负性 $P\{X = k\} \geq 0$.
- 规范性

$$\begin{aligned}\sum_{k=0}^{\infty} P\{X = k\} &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \cdot e^{\lambda} \\ &= 1.\end{aligned}$$

泊松分布的用途

如果某事件以固定强度 λ , 随机且独立地出现. 该事件在单位时间内出现的次数可以看成泊松分布. 比如:

- 一本书一页中的印刷错误数;
- 某人一天收到的微信数量;
- 某医院在一天中的急诊患者人数;
- 某放射性物质射出的粒子;
- 显微镜下某区域中的白血球.

泊松定理

用泊松分布来逼近二项分布的定理

定理

设 $\lambda > 0$ 是一常数, n 是任意正整数, 设 $np_n = \lambda$, 则对于任一固定的非负整数 k , 有

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

证明: 记 $p_n = \frac{\lambda}{n}$.

$$\begin{aligned} C_n^k p_n^k (1 - p_n)^{n-k} &= C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left[1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right] \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

k 固定, $n \rightarrow \infty$ 时,

$$1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \rightarrow 1, \quad \left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}, \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow 1.$$

得证. \square

条件 $np_n = \lambda$ 意味着 n 很大时, p_n 必定很小. 二项分布的概率值可以以 $\lambda = np$ 的泊松分布值近似.

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

例

计算机硬件公司制造特殊型号的微型芯片, 次品率达 0.1%, 各芯片成为次品相互独立. 求 1000 只产品中至少有 2 只次品的概率. 以 X 记产品中的次品数, $X \sim b(1000, 0.001)$.

例

计算机硬件公司制造特殊型号的微型芯片，次品率达 0.1%，各芯片成为次品相互独立. 求 1000 只产品中至少有 2 只次品的概率. 以 X 记产品中的次品数, $X \sim b(1000, 0.001)$.

解:

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} = \\ 1 - 0.999^{1000} - 1000 \cdot 0.999^{999} \cdot 0.001 \approx 0.2642411.$$

$$\text{利用 } C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}. \quad (\lambda = np)$$

$$\lambda = 1000 \times 0.001 = 1, \quad P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - e^{-1} - e^{-1} \approx 0.2642411. \quad \square$$

一般当 $n \geq 20$, $P \leq 0.05$ (另一种说法 $n > 10$, $P < 0.1$) 时, 用泊松分布.

几何分布

定义

若 X 的概率分布律为:

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

其中 $0 < p < 1$, 则称 X 服从参数为 p 的几何分布. 记为 $X \sim \text{Geom}(p)$.

X	1	2	...	k	...
P_k	p	qp	...	$q^{k-1}p$...

用途: 在重复多次的伯努利试验中, 试验进行到某种结果出现第一次为止, 此时的试验总次数服从几何分布.

例

设某批产品的次品率为 p , 对该产品作有放回的抽样调查, 直到第一次抽到一只次品为止. 那么所抽到的产品数 X 是一个随机变量, 求 X 的分布律.

例

设某批产品的次品率为 p , 对该产品作有放回的抽样调查, 直到第一次抽到一只次品为止. 那么所抽到的产品数 X 是一个随机变量, 求 X 的分布律.

解: $X = 1, 2, \dots$, $A_i = \{\text{第 } i \text{ 个产品是正品}\}$,

$$\begin{aligned}P\{X = k\} &= P(A_1 A_2 \dots A_{k-1} \bar{A}_k) \\&= P(A_1) \dots P(A_{k-1}) P(\bar{A}_k) \\&= (1 - p)^{k-1} p.\end{aligned}$$