

# 软件工程统计方法

## 随机变量及其分布（一）

陈振宇

南京大学软件学院

Email: [zychen@software.nju.edu.cn](mailto:zychen@software.nju.edu.cn)

Homepage: [software.nju.edu.cn/zychen](http://software.nju.edu.cn/zychen)



本节内容

随机变量

分布函数

数学期望

方差

伯努利分布

二项分布

泊松分布

几何分布

超几何分布



第 1 页 共 100 页

返回

全屏

关闭

# 内容提纲

- 统计学导论
- 描述统计
- 概率计算基础
- 随机变量及其分布
- 统计量及其抽样分布
- 参数估计
- 参数假设检验
- 非参数假设检验
- 方差分析
- 回归分析



本节内容

随机变量

分布函数

数学期望

方差

伯努利分布

二项分布

泊松分布

几何分布

超几何分布



第 2 页 共 100 页

返回

全屏

关闭



# 1 本节内容

- 随机变量
- 分布函数
- 数学期望
- 方差
- 常用分布
- 二维随机向量

本节内容

随机变量

分布函数

数学期望

方差

伯努利分布

二项分布

泊松分布

几何分布

超几何分布



第 3 页 共 100 页

返回

全屏

关闭



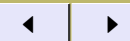
本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

## 2 随机变量

引入随机变量是研究随机现象统计规律性的需要. 为了便于数学推理和计算, 有必要将随机试验的结果数量化, 使得可以用高等数学课程中的理论与方法来研究随机试验, 研究和分析其结果的规律性, 因此, 随机变量是研究随机试验的重要而有效的工具。

**Definition 1 (随机变量)** 设 $E$ 为随机试验, 它的样本空间为 $\Omega = \{e\}$ . 若对于每一个样本点 $e \in \Omega$ , 都有唯一确定的实数 $X(e)$ 与之对应, 则称 $X(e)$ 是一个随机变量, 可简记为 $X$ 。

随机变量常用使用大写字母 $X, Y, Z$ 等表示。



第 4 页 共 100 页

返回

全屏

关闭



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

# 概率分布

**Example 1** 将一枚硬币抛掷3次, 我们感兴趣的是三次投掷中, 出现 $H$ 的总次数, 而对 $H, T$ 出现的次序不关心. 以 $X$ 记三次投掷中出现 $H$ 的总次数, 那么对于样本空间 $S = \{e\}$ 中的每一个样本点 $e$ ,  $X$ 都有一个值与之对应, 即有

样本点	$HHH$	$HHT$	$HTH$	$THH$	$TTH$	$THT$	$HTT$	$TTT$
$X$ 值	3	2	2	2	1	1	1	0

我们注意到, 随机变量的取值随试验的结果而定, 而试验的各个结果出现有一定的概率, 因而随机变量的取值有一定的概率. 如, 当且仅当事件 $A = \{HHT, HTH, THH\}$ 发生时有 $\{x = 2\}$ , 而且 $P(A) = \frac{3}{8}$ , 则 $P\{x = 2\} = \frac{3}{8}$ .

将3个球随机地放入三个框中, 如何描述随机事件和定义随机变量?



第 5 页 共 100 页

返回

全屏

关闭



# 概率分布

**Definition 2 (概率分布)** 若 $L$ 是一个实数集合, 将 $X$ 在 $L$ 上取值写成 $\{X \in L\}$ , 它表示事件 $A = \{e | X(e) \in L\}$ , 即 $A$ 是由 $S$ 中使得 $X(e) \in L$ 的所有样本点 $e$ 所组成的事件, 此时有 $P\{X \in L\} = P(A) = P\{e | X(e) \in L\}$ .  $P\{X \in L\}$ 称为随机变量 $X$ 的概率分布。

随机变量的分类:

- 离散型随机变量, 连续型随机变量, 混合型随机变量。
- 一元随机变量, 多元随机变量(向量)。

本节内容

随机变量

分布函数

数学期望

方差

伯努利分布

二项分布

泊松分布

几何分布

超几何分布



第 6 页 共 100 页

返回

全屏

关闭



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

### 3 分布函数

**Definition 3 (分布函数)** 对于随机变量 $X$ 和任意实数 $x$ , 称函数 $F(x) = P(X \leq x)$ 为随机变量 $X$ 的分布函数. 它在点 $x$ 处的值是事件 $\{X \leq x\}$ 的概率。

分布函数 $F(x)$ 的性质:

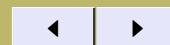
1. 归一性:  $0 \leq F(x) \leq 1, \forall x \in R$ , 且

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$$

2. 单调不减性: 若 $x_1 \leq x_2$ , 则有 $F(x_1) \leq F(x_2)$ ;

3. 右连续性: 对任意 $x_0 \in R$ , 有

$$F(x_0) = \lim_{x \rightarrow x_0^+} F(x)$$



第 7 页 共 100 页

返回

全屏

关闭

# 离散随机变量

**Definition 4 (离散型随机变量)** 一个随机变量的一切可能的取值为有限个或可列无穷多个, 则称它为离散型随机变量。

**Definition 5 (概率分布(分布律))**  $X$  是一个离散型随机变量, 其一切可能值为  $\{x_1, x_2, \dots, x_n, \dots\}$ , 且  $X$  取各值时的概率为

$$P(X = x_i) = p_i, i = 1, 2, \dots, n, \dots$$

其中  $P(x_k) \geq 0, (k = 1, 2, \dots)$ ,

$$\sum_{i=1}^{\infty} p_i = 1$$

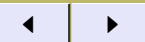
**Definition 6** 有了离散随机变量的分布律, 可以通过下式求得分布函数:

$$F(X) = P(X \leq x) = \sum_{x_i \leq x} P\{X = x_i\} = \sum_{x_i \leq x} p_i$$

显然这时  $F(x)$  是一个跳跃函数。我们可以用分布律或分布函数来描述离散型随机变量。



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 8 页 共 100 页

返回

全屏

关闭





本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

## 例子

**Example 2** 设一汽车在开往目的地的道路上需经过四组信号灯, 每组信号灯以 $\frac{1}{2}$ 的概率允许或禁止汽车通过. 以 $X$ 表示汽车首次停下时, 它已通过的信号灯的组数(设各组信号灯的工作是相互独立的), 求 $X$ 的分布律。

解: 以 $p$ 表示每组信号灯禁止汽车通过的概率, 易知 $X$ 的分布律为

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

以 $p = \frac{1}{2}$ 代入得

$X$	0	1	2	3	4
$p_k$	0.5	0.25	0.125	0.0625	0.0625

试写出相应的分布函数和画出相应的分布函数图.



第 9 页 共 100 页

返回

全屏

关闭

## 4 数学期望

问题的提出: 在一些实际问题中, 我们需要了解随机变量的分布函数外, 更关心的是随机变量的某些特征。

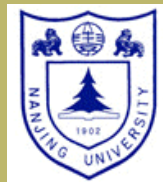
例如:

- 在评定某地区粮食产量的水平时, 最关心的是平均产量;
- 在检查一批棉花的质量时, 既需要注意纤维的平均长度, 又需要注意纤维长度与平均长度的偏离程度;
- 考察南京市居民的家庭收入情况, 我们既知家庭的年平均收入, 又要研究贫富之间的差异程度。

### 离散随机变量数学期望

设离散型随机变量 $X$ 的概率分布为 $P(X = x_k) = p_k, k = 1, 2, \dots, n, \dots$ , 如果无穷级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 则称此级数的和为随机变量 $X$ 的数学期望(或均值), 记作 $E(X)$ 或 $EX$ , 即

$$E(X) = \sum_{k=1}^{\infty} x_k p_k \quad (1)$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 10 页 共 100 页

返回

全屏

关闭

# 数学期望

**Example 3** 甲乙射击比赛, 各射击100次, 其中甲乙的成绩如下:

环数	8	9	10	环数	8	9	10
甲次数	10	80	10	乙次数	20	65	15

计算他们的平均环数.

甲平均环数计算:

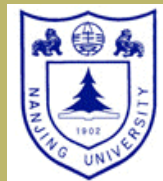
$$\frac{8 * 10 + 9 * 80 + 10 * 10}{100} = 8 * \frac{10}{100} + 9 * \frac{80}{100} + 10 * \frac{10}{100} = 9$$

乙平均环数计算:

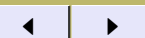
$$\frac{8 * 20 + 9 * 65 + 10 * 15}{100} = 8 * \frac{20}{100} + 9 * \frac{65}{100} + 10 * \frac{15}{100} = 8.95$$

所以甲的成绩好于乙的成绩.

对于甲来说  $\frac{10}{100}, \frac{80}{100}, \frac{10}{100}$  分别为8, 9, 10环的命中概率; 对于乙来说  $\frac{20}{100}, \frac{65}{100}, \frac{15}{100}$  分别为8, 9, 10环的命中概率. 因此我们通过概率得到数学期望的概念.



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 11 页 共 100 页

返回

全屏

关闭



# 数学期望的性质

数学期望的性质:

1. 对于常数 $C$ , 有 $E(C) = C$ ;
2. 对于常数 $C$ 及随机变量 $X$ , 有 $E(CX) = CE(X)$ ;
3. 设 $X$ 和 $Y$ 为两个随机变量, 则 $E(X + Y) = E(X) + E(Y)$ ;
4. 设随机变量 $X$ 和 $Y$ 独立, 则 $E(XY) = E(X)E(Y)$ 。

后两个性质是关于二维随机变量, 我们在稍后证明。

**Theorem 1** 设 $Y = g(X)$ 是随机变量 $X$ 的函数,  $g$ 是连续函数. 若 $X$ 为离散型随机变量, 分布律为 $P(X = x_k) = p_k, k = 1, 2, \dots, n, \dots$ , 如果级数 $\sum_{k=1}^{\infty} g(x_k)p_k$ 绝对收敛, 则

$$E(Y) = E(g(X)) = \sum_{k=1}^{\infty} g(x_k)p_k$$

例如 $Y = X + 2, Y = X^2$ 等等。



## 5 方差

设有一批灯泡寿命为: 一半约950小时, 另一半约1050小时, 平均寿命为1000小时; 另一批灯泡寿命为: 一半约1300小时, 另一半约700小时, 平均寿命为1000小时.

单从平均寿命这一指标无法判断, 进一步考察灯泡寿命 $X$ 与均值1000小时的偏离程度.

**Definition 7 (方差)** 设 $X$ 是一随机变量, 如果数学期望 $E[X - E(X)]^2$ 存在, 则称之为 $X$ 的方差, 记作 $Var(X)$ 或 $D(X)$ .  $\sqrt{Var(X)}$ 或 $\sqrt{D(X)}$ 称为 $X$ 的标准差.

方差的计算式是:  $Var(X) = EX^2 - (EX)^2$ .

方差的性质:



1. 设 $C$ 是常数, 则 $Var(C)=0$ ;
2. 设 $X$ 是随机变量,  $C$ 是常数, 则 $Var(X + C) = Var(X)$ ,  $Var(CX) = C^2Var(X)$ .
3. 设随机变量 $X$ 与 $Y$ 相互独立, 则 $Var(X \pm Y) = Var(X) + Var(Y)$ .

第三个性质关于二维变量, 我们在稍后证明。



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 13 页 共 100 页

返回

全屏

关闭

## 6 伯努利分布

伯努利分布 (Bernoulli distribution), 又名两点分布或0-1分布, 是一个离散型概率分布, 为纪念瑞士科学家雅各布·伯努利而命名。伯努利试验是只有两种可能结果的单次随机试验。

**Definition 8 (伯努利分布)** 对于一个随机试验, 如果它的样本空间只包含两个元素, 即  $\Omega = \{e_1, e_2\}$ , 我们总能在  $\Omega$  上定义一个服从伯努利分布(又称0-1分布)的随机变量. 它的概率分布为:

$$P(X = 1) = p, P(X = 0) = q, p + q = 1$$

来描述这个随机试验的结果。

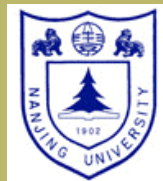
例如, 对新生婴儿的性别进行登记, 检查产品的质量是否合格, 某车间的电力消耗是否超过负荷以及前面多次讨论过的“抛硬币”试验等都可以用伯努利分布的随机变量来描述. 伯努利分布是经常遇到的一种分布。

数字特征:

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$E(X^2) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

$$Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p)$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 14 页 共 100 页

返回

全屏

关闭

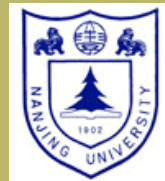
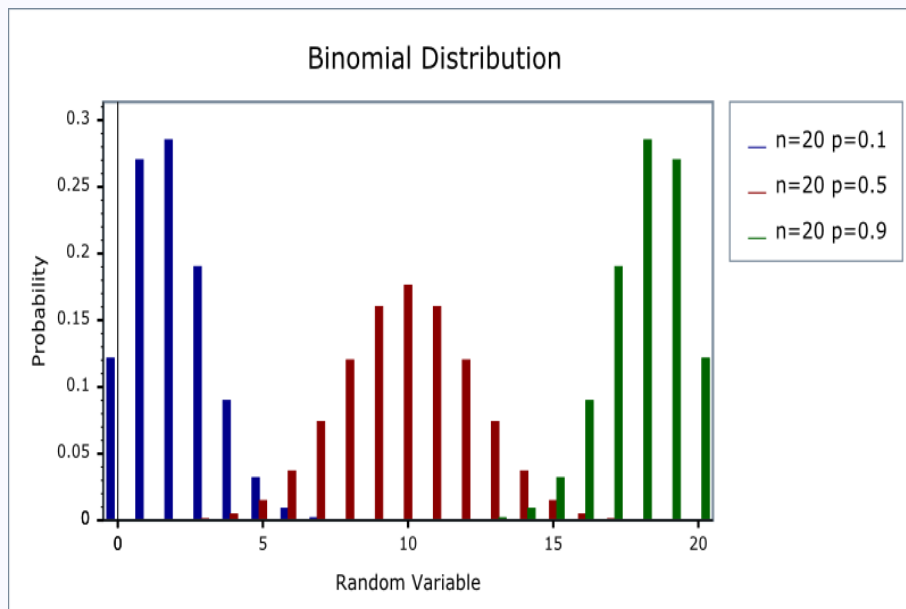
# 7 二项分布

二项分布 (Binomial Distribution) 是 $n$ 个独立的伯努利试验中成功的次数的离散概率分布。

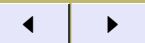
**Definition 9** 设事件 $A$ 在任一次试验中出现的概率为 $p$ , 则在 $n$ 重伯努利试验中事件 $A$ 发生的次数 $k$ 的取值为 $1, \dots, n$ 且它的概率分布为:

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

则称 $X$ 服从参数为 $n, p$ 的二项分布, 记为 $X \sim \mathbb{B}(n, p)$ .



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 15 页 共 100 页

返回

全屏

关闭



# 二项分布

数字特征:  $X = \sum_{i=1}^n X_i$ , 其中  $X_i$  服从0-1分布.

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = np$$

$$Var(X) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) = np(1-p)$$

**Example 4** 按规定, 某种型号电子元件的使用寿命超过1500小时的为一级品. 已知某一大批产品的一级品率为0.2, 现在从中随机地抽查20只. 问20只元件中恰有 $k$ 只( $k = 0, \dots, 20$ )为一级品的概率是多少?

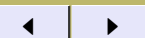
这是不放回抽样. 但由于这批元件的总数很大, 且抽查的元件的数量相对于元件的总数来说又很小, 因而可以当作放回抽样来处理, 这样做会有一些误差, 但误差不大. 我们将检查一只元件看它是否为一级品看成是一次试验, 检查20只元件相当于做20重伯努利试验.

以 $X$ 记20只元件中一级品的只数, 那么,  $X$ 是一个随机变量, 且有  $X \sim \mathbb{B}(20, 0.2)$ . 即得所求概率为

$$P\{X = k\} = C_{20}^k (0.2)^k (0.8)^{20-k}$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 16 页 共 100 页

返回

全屏

关闭



## 二项分布

**Example 5** 某人进行射击, 设每次射击的命中率为0.02, 独立射击400次, 试求至少击中两次的概率.

将一次射击看成是一次试验. 设击中的次数为 $X$ , 则 $X \sim \mathbb{B}(400, 0.02)$ .  $X$ 的分布律为

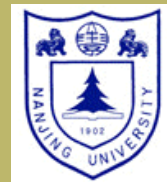
$$P\{X = k\} = C_{400}^k (0.02)^k (0.98)^{400-k}$$

即得所求概率为

$$\begin{aligned} P\{X \geq 2\} &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - (0.98)^{400} - 400(0.02)(0.98)^{399} = 0.9972 \end{aligned}$$

小概率事件是指一次试验中发生的概率很小的事件. 但从理论上讲, 一个事件发生的概率不论多小, 只要不断重复试验下去, 事件迟早会出现的, 概率是1.

其实际意义是, 我们可以借助它判断事情的真实性. 因为根据实际推断原理, 小概率事件在一次试验中几乎是不可能发生的. 而某一认为概率很小的事件, 居然在一次试验中发生了, 人们就有理由怀疑其正确性.



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 17 页 共 100 页

返回

全屏

关闭

## 二项分布

**Example 6** 设有80台同类型设备, 各台工作是相互独立的, 发生故障的概率都是0.01, 且一台设备的故障能由一个人处理. 考虑两种配备维修工人的方法, 其一是由4人维护, 每人负责20台; 其二是由3人共同维护80台. 试比较这两种方法在设备发生故障时不能及时维修的概率的大小.

按第一种方法. 以 $X$ 记“第1人维护的20台中同一时刻发生故障的台数”, 以 $A_i (i = 1, 2, 3, 4)$ 表示事件“第 $i$ 人维护的20台中发生故障不能及时维修”, 则知80台中发生故障而不能及时维修的概率为

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

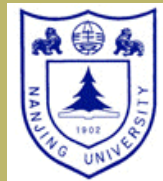
而 $X \sim \mathbb{B}(20, 0.01)$ , 故有

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} = 0.0169$$

按第二种方法. 以 $Y$ 记80台中同一时刻发生故障的台数. 此时,  $Y \sim \mathbb{B}(80, 0.01)$ , 故80台中发生故障而不能及时维修的概率为

$$P\{Y \geq 4\} = 1 - \sum_{i=0}^3 C_{80}^i (0.01)^i (0.99)^{80-i} = 0.0087$$

我们发现, 在后一种情况尽管任务重了(每人平均维护约27台), 但工作效率不仅没有降低, 反而提高了.



本节内容

随机变量

分布函数

数学期望

方差

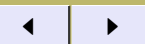
伯努利分布

二项分布

泊松分布

几何分布

超几何分布



第 18 页 共 100 页

返回

全屏

关闭



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

## 8 泊松分布

泊松分布 (Poisson Distribution) 是法国数学家泊松于1837年引入的。泊松分布适合于描述单位时间内随机事件发生的次数的概率分布。如某一服务设施在一定时间内受到的服务请求的次数, 电话交换机接到呼叫的次数、汽车站台的候客人数、机器出现的故障数、自然灾害发生的次数、DNA序列的变异数、放射性原子核的衰变数等等。

**Definition 10** 如果随机变量 $X$  的概率分布为:

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

则称 $X$ 服从参数为 $\lambda$ 的泊松分布, 记为 $\pi(\lambda)$ .



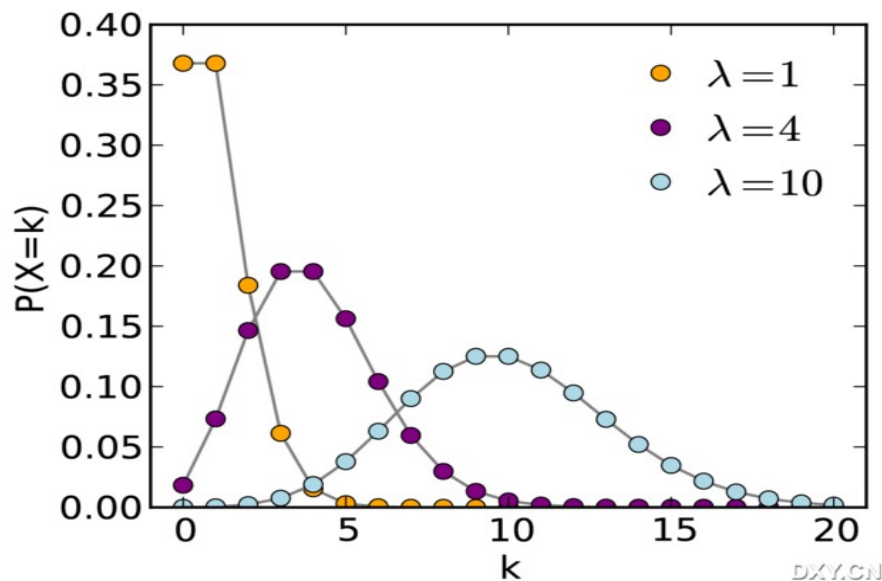
第 19 页 共 100 页

返回

全屏

关闭

# 泊松分布



**Example 7** 设某汽车停靠站候车的人数为  $X \sim \pi(\lambda)$ ,  $\lambda = 4.5$ . 求

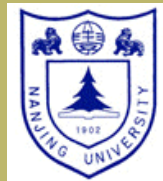
(1) 求至少有两人等车的概率;

(2) 已知至少有两人候车; 求恰有两人候车的概率.

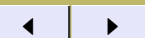
解:  $P(X = k) = \frac{4.5^k}{k!} e^{-4.5}$ , 则

$$(1) P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-4.5}(1 + 4.5) = 0.9389$$

$$(2) P(X = 2 | X \geq 2) = \frac{P(X = 2)}{P(X \geq 2)} = 0.1198$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 20 页 共 100 页

返回

全屏

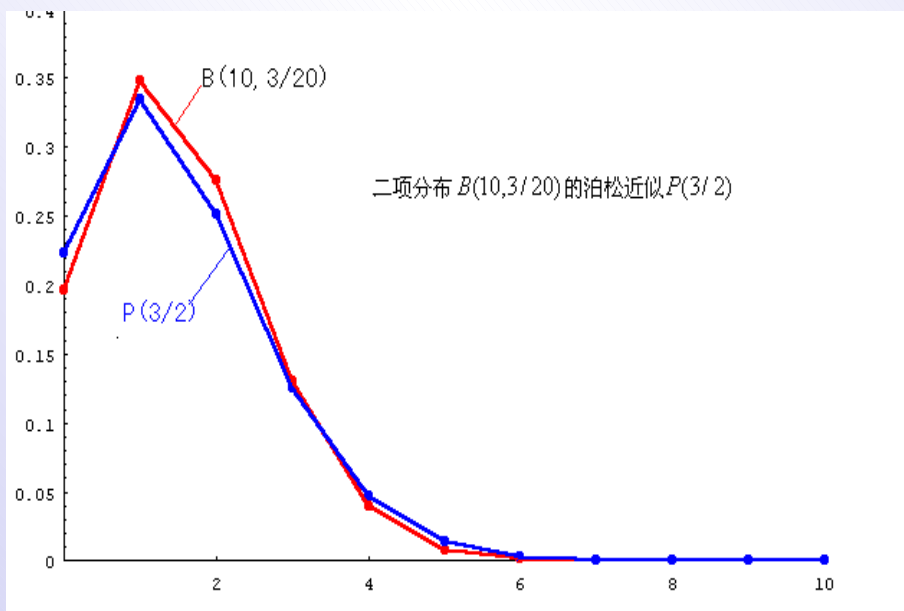
关闭

# 泊松逼近

在很多应用问题中, 我们常常这样的伯努利试验, 其中, 相对地说,  $n$ 大,  $p$ 小, 而乘积 $\lambda = np$ 大小适中. 在这种情况下, 有一个便于使用的近似公式.

**Theorem 2 (泊松逼近)** 在伯努利试验中, 以 $p_n$ 代表事件 $A$ 在试验中出现的概率, 如果 $np_n \rightarrow \lambda$ , 则当 $n \rightarrow \infty$ 时,

$$\lim_{n \rightarrow \infty} \mathbb{B}(n, p_n) = \pi(\lambda)$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 21 页 共 100 页

返回

全屏

关闭

# 泊松逼近

回顾二项分布定义：

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

$$\begin{aligned}\lim_{n \rightarrow \infty} P(X = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1 - p)^{n-k} \\&= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\&= \lim_{n \rightarrow \infty} \underbrace{\left[\frac{n!}{n^k (n-k)!}\right]}_F \underbrace{\left(\frac{\lambda^k}{k!}\right)}_{\rightarrow \exp(-\lambda)} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \\&= \lim_{n \rightarrow \infty} \underbrace{\left[\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right]}_{\rightarrow 1} \underbrace{\left(\frac{\lambda^k}{k!}\right)}_{\rightarrow \exp(-\lambda)} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \\&= \left(\frac{\lambda^k}{k!}\right) \exp(-\lambda)\end{aligned}$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 22 页 共 100 页

返回

全屏

关闭

# 泊松逼近

**Example 8** 假如生三胞胎的概率为 $10^{-4}$ , 求在100000次生育中, 有0, 1, 2次生三胞胎的概率.

解: 这可以看作伯努利试验;  $n = 100000$ ,  $p = 0.0001$ , 所求的概率直接计算为

$$\mathbb{B}(0; 100000, 0.0001) = 0.000045378$$

$$\mathbb{B}(1; 100000, 0.0001) = 0.00045382$$

$$\mathbb{B}(2; 100000, 0.0001) = 0.0022693$$

这时也可用泊松逼近,  $\lambda = np = 10$ , 而

$$\pi(0; 10) = 0.00004540$$

$$\pi(1; 10) = 0.0004540$$

$$\pi(2; 10) = 0.002270$$



本节内容

随机变量

分布函数

数学期望

方差

伯努利分布

二项分布

泊松分布

几何分布

超几何分布



第 23 页 共 100 页

返回

全屏

关闭

# 泊松分布数字特征

**Theorem 3** 设  $X \sim \pi(\lambda)$ , 则  $E(X) = \lambda$ .

证:

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

**Theorem 4** 设  $X \sim \pi(\lambda)$ , 则  $Var(X) = \lambda$ .

证:

$$\begin{aligned} E(X^2) &= E[X(X-1) + X] = E[X(X-1)] + E(X) \\ &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \lambda = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \\ &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda \\ Var(X) &= E(X^2) - E(X)^2 = \lambda \end{aligned}$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 24 页 共 100 页

返回

全屏

关闭





本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

## 9 几何分布

几何分布 (Geometric distribution) 在伯努利试验中得到一次成功所需要的试验次数  $X$ .

**Example 9** 从生产线上随机抽产品进行检测, 设产品的次品率为  $p$ ,  $0 < p < 1$ , 若查到一只次品就得停机检修, 设停机时已检测到  $X$  只产品, 试写出  $X$  的概率分布律。

设  $A_i$  为第  $i$  个抽到正品事件,  $A_i$  相互独立, 则

$$P\{X = n\} = P\{A_1, \dots, A_{n-1}, \overline{A_n}\} = (1-p)^{n-1}p$$

**Definition 11 (几何分布)** 如果随机变量  $X$  的概率分布为:

$$P(X = n) = (1-p)^{n-1}p, n = 1, 2, \dots, 0 < p < 1$$

则称  $X$  服从几何分布, 记为  $X \sim \mathbb{G}(p)$



第 25 页 共 100 页

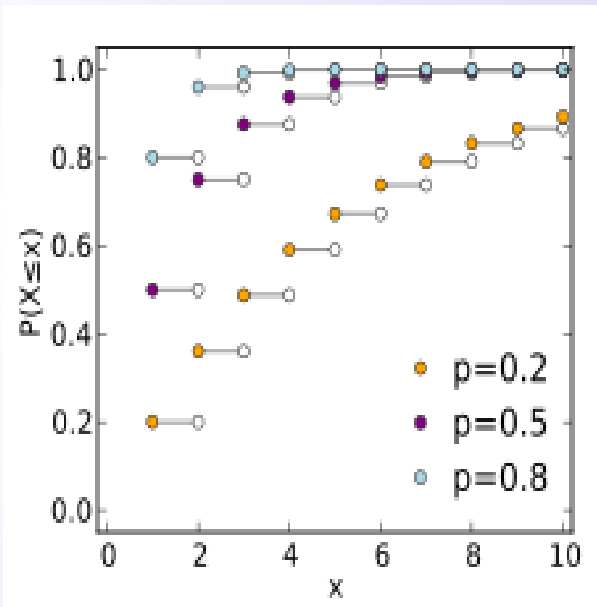
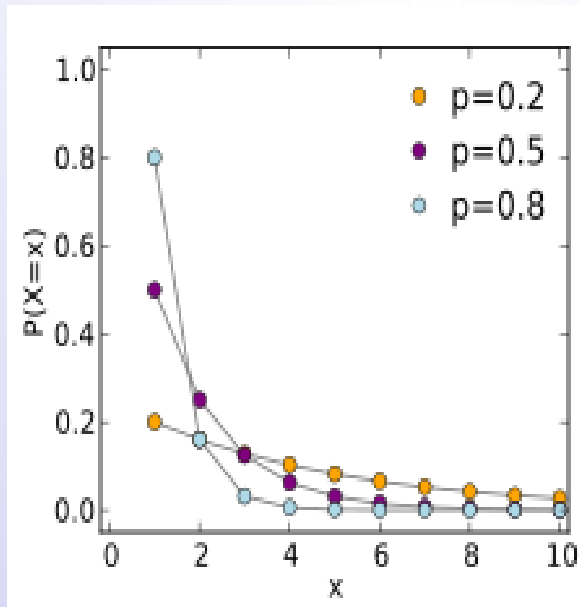
返回

全屏

关闭



# 几何分布



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

« »

« »

第 26 页 共 100 页

返回

全屏

关闭



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

# 几何分布

几何分布的数学期望( $q = 1 - p$ ):

$$E(X) = \sum_{i=1}^{\infty} ipq^{i-1}$$

$$= 1p + 2pq + 3pq^2 + \cdots + kpq^{k-1} + \cdots$$

$$= p(1 + 2q + 3q^2 + \cdots + kq^{k-1} + \cdots)$$

$$\text{令 } S = 1 + 2q + 3q^2 + \cdots + kq^{k-1} + \cdots, \quad qS = 1q + 2q^2 + 3q^3 + \cdots + kq^k + \cdots$$

$$S - qS = (1 - q)S = 1 + q + q^2 + \cdots + q^k + \cdots = \frac{1}{1 - q}$$

$$\text{所以 } S = \frac{1}{(1 - q)^2},$$

$$E(X) = pS = \frac{p}{(1 - q)^2} = \frac{1}{p}$$



第 27 页 共 100 页

返回

全屏

关闭

# 几何分布

几何分布的方差( $q = 1 - p$ ):

$$E(X^2) = \sum_{i=1}^{\infty} i^2 p q^{i-1}$$

$$= p(1 + 2^2 q + 3^2 q^2 + \cdots + k^2 q^{k-1} + \cdots)$$

令  $S = 1 + 2^2 q + 3^2 q^2 + \cdots + k^2 q^{k-1} + \cdots$ , 对  $T' = S$ , 即  $T$  关于  $q$  求一阶导得  $S$

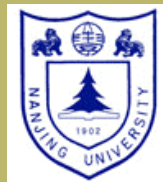
$$T = q + 2q^2 + 3q^3 + \cdots + kq^k + \cdots = \frac{q}{(1-q)^2}$$

则  $S = T'$ ,

$$S = T' = \frac{(1-q)^2 + 2(1-q)q}{(1-q)^4} = \frac{1-q^2}{(1-q)^4} = \frac{1+q}{(1-q)^3}$$

所以  $E(X^2) = pS = \frac{2-p}{p^2}$

$$D(X) = E(X^2) - E^2(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$



本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布



第 28 页 共 100 页

返回

全屏

关闭



本节内容
随机变量
分布函数
数学期望
方差
伯努利分布
二项分布
泊松分布
几何分布
超几何分布

# 几何分布

在伯努利试验中, 等待首次成功的时间 $t$ 服从几何分布。现在假定已知在前 $m$ 次试验中没有出现成功, 那么为了达到首次成功所再需要的等待时间 $t'$ 也还是服从几何分布, 与前面的失败次数 $m$ 无关, 形象化地说, 就是把过去的经历完全忘记了。

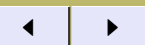
**Definition 12 (无记忆)** 对于一个非负随机变量 $X$ , 如果对于任意 $s, t \geq 0$ 有

$$P\{X > s + t | X > t\} = P\{X > s\} \quad (2)$$

则我们称 $X$ 是无记忆的。

$$P\{X \leq n\} = 1 - q^n, P\{X > n\} = q^n$$
$$P\{X > n\} = \sum_{i=n+1}^{\infty} pq^{i-1} = pq^n \sum_{i=0}^{\infty} q^i = \frac{pq^n}{1-q} = q^n$$

因此无记忆性是几何分布所具有的一个有趣的性质。





本节内容  
随机变量  
分布函数  
数学期望  
方差  
伯努利分布  
二项分布  
泊松分布  
几何分布  
超几何分布

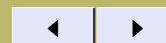
## 10 超几何分布

一批产品共 $N$ 件，含 $M$ 件是次品，随机地从这 $N$ 件产品中抽取 $n$ 件产品，求恰有 $k$ 件次品的概率。

**Definition 13 (超几何分布)** 如果随机变量 $X$ 的概率分布为：

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 1, 2, \dots, n$$

其中 $N, M, n$ 均为正整数，且 $M \leq N, n \leq N$ ，则称 $X$ 服从参数为 $N, M, n$ 的超几何分布，记为 $X \sim \mathbb{H}(N, M, n)$ 。



第 30 页 共 100 页

返回

全屏

关闭



本节内容
随机变量
分布函数
数学期望
方差
伯努利分布
二项分布
泊松分布
几何分布
超几何分布

# 超几何分布

我们把二项分布与超几何分布作一比较。 $N$ 件产品，有 $M$ 件次品和 $N - M$ 件正品。如果每抽一件产品放回后，再抽下一件产品，如此有放回地随机地抽取 $n$ 件，这是 $n$ 重伯努利试验，那么所抽的 $n$ 件产品的次品数 $X \sim \mathbb{B}(n, \frac{M}{N})$ 。当 $N \gg n$ 时，我们可以采用二项分布近似超几何分布。

超几何分布、二项分布和泊松分布都是重要的离散型随机变量的概率分布。有时，他们的概率计算会十分繁冗。当试验次数 $n$ 很大时，可以推导出这三个分布间有一种近似关系式

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

这里，第一个等式要求 $n$ 很大，且 $n/N$ 较小，取 $p = M/N$ 即成立。第二个等式要求 $n$ 很大时成立。实际使用时， $n \geq 20$ 即可，当 $n \geq 50$ 时，效果更好。而泊松分布可通过查表计算，比较简单。

