

概率论与数理统计

Lecture 3: 随机变量及其分布 (一)

曾靖

中国科学技术大学统计与金融系



目录

2.1 随机变量的概念

2.2 离散型随机变量

2.1 随机变量的概念

为什么要引入随机变量

样本空间的样本点可能是一些实体，并没有量化。对这些样本点进行计算，除了集合运算工具外，无法使用其他数学分析工具。

例 1

从某厂的产品中随机抽取 3 件进行检验，以 A 表示 {抽出的该件产品是合格品}，则检验的可能结果有 8 种：

$$\bar{A}\bar{A}\bar{A}, \bar{A}\bar{A}A, \bar{A}A\bar{A}, A\bar{A}\bar{A}, AA\bar{A}, A\bar{A}A, \bar{A}AA, AAA.$$

按顺序记为 $\omega_i, i = 1, 2, \dots, 8$ 。

随机变量的引入

如果我们关心的是 3 件产品中的合格品件数，可以作映射 $X: \Omega \rightarrow \mathbb{R}$ ，使

$$X(\omega_1) = 0, \quad X(\omega_2) = X(\omega_3) = X(\omega_4) = 1,$$

$$X(\omega_5) = X(\omega_6) = X(\omega_7) = 2, \quad X(\omega_8) = 3.$$

那么，我们可以利用映射 X 简单表达出事件 {3 件产品中恰有两件合格品}：

$$\{\omega : X(\omega) = 2\} = \{\omega_5, \omega_6, \omega_7\},$$

可以看出，映射 X 把样本空间的样本点与实数 $\{0, 1, 2, 3\}$ 建立对应关系。这种映射称为**随机变量**。

随机变量的定义

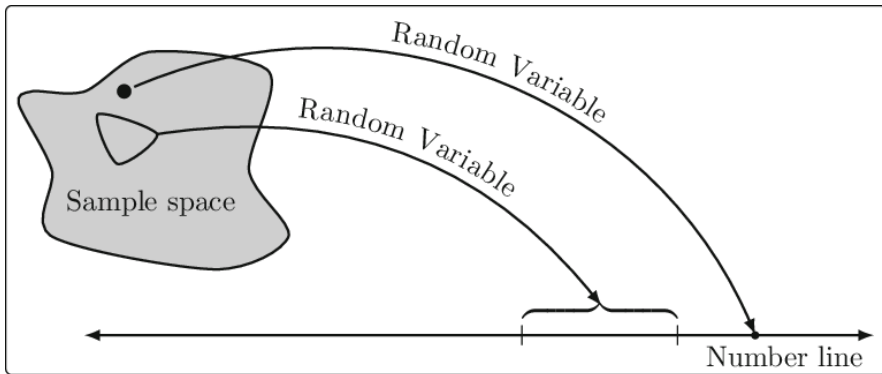
定义 2.1 随机变量

令 Ω 为一个样本空间。随机变量 X 是一个映射 $X: \Omega \rightarrow \mathbb{R}$, 对每个 $A \subset \mathbb{R}$, $\{X \in A\}$ 是一个随机事件, 且

$$P(X \in A) = P(\omega : X(\omega) \in A).$$

直观上, 随机变量是取值随试验结果而定, 且有一定概率分布的变量。

随机变量示意图



随机变量的分类

常见的随机变量可以分为两大类：

- ▶ 只取有限个或可数个值的随机变量称为**离散型随机变量**；
- ▶ 取连续的值且密度存在的随机变量称为**连续型随机变量**。

当然，存在既非离散型也非连续型的随机变量，但它们在实际中并不常见，也不是我们这里研究的主要对象。

随机变量的分布

引入随机变量后，我们需要研究它的两个方面：

1. 随机变量可能取哪些值？
2. 随机变量取这些值的概率分别是多少？

随机变量的分布

随机变量的可能取值及其对应的概率，合在一起称为随机变量的分布。

2.2 离散型随机变量

离散型随机变量的定义

定义 2.2 离散型随机变量和分布律

如果随机变量 X 只取有限多个或可数多个值, 那么称 X 为**离散型随机变量**。设 X 取的一切可能值为 $x_1, x_2, \dots$, 则

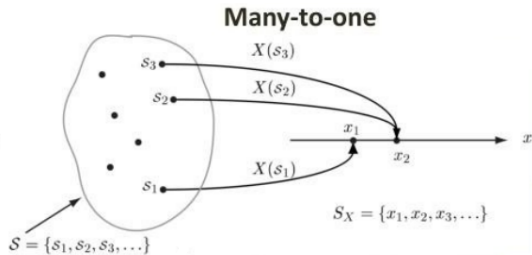
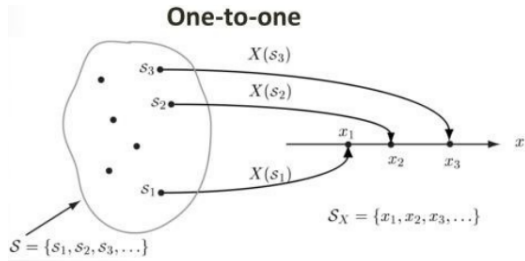
$$P(X = x_k) = p_k, \quad k = 1, 2, \dots,$$

其中

$$p_k \geq 0, \quad \sum_k p_k = 1.$$

称 $P(X = x_k) = p_k$ 为 X 的**分布律**或**概率质量函数** (probability mass function, or pmf)。

离散型随机变量示意图



分布律的表示

概率质量函数指出了全部概率 1 是如何在 X 的所有可能值之间分配的。它也就是离散型随机变量的**分布律**，可以列表形式给出：

$$\begin{array}{c|ccccc} X & a_1 & a_2 & \cdots & a_n & \cdots \\ \hline P & p_1 & p_2 & \cdots & p_n & \cdots \end{array}$$

或

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

设 A 为 $\{a_1, a_2, \dots\}$ 的任意子集，则

$$P(X \in A) = \sum_{x_k \in A} P(X = x_k).$$

Bernoulli 试验

定义 Bernoulli 试验

设 A 是随机试验中的一个事件，且 $0 < p = P(A) < 1$ 。若每次试验只有“ A 发生”和“ A 不发生”两种互斥且完备的结果，则称该试验为 **Bernoulli 试验**； A 发生称为成功， $\bar{A}$ 发生称为失败。

下面的 **0-1 分布** 和 **二项分布** 都是以 **Bernoulli 试验** 为基础的。

0-1 分布

定义 2.3 0-1 分布

设随机变量 X 只取 0,1 两值,

$$P(X=1) = p, \quad P(X=0) = 1 - p,$$

则称 X 服从**0-1 分布**或**Bernoulli 分布**。也可以写为

$$P(X=x) = p^x(1-p)^{1-x}, \quad x=0 \text{或} 1.$$

离散均匀分布

定义 2.4 离散均匀分布

设随机变量 X 取值 $a_1, a_2, \dots, a_n$, 且有

$$P(X = a_k) = \frac{1}{n}, \quad k = 1, \dots, n.$$

则称 X 服从离散均匀分布。

可以看出, **离散均匀分布**正是**古典概型**的抽象。

离散均匀分布：应用实例

例 2

假设一个池塘中有 N 条鱼，从池中任意捞 M 条鱼并进行标记后再放回池塘， $M \ll N$ 。现从中捞出 n 条鱼，求其中恰有 m 条鱼有标记的概率。

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

此分布称为参数为 (N, M, n) 的超几何分布。

二项分布

定义 2.5 二项分布

设某事件 A 在一次试验中发生的概率为 p 。现把试验独立地重复 n 次。以 X 记 A 在这 n 次试验中发生的次数，则 X 取值 $0, 1, \dots, n$ ，且有

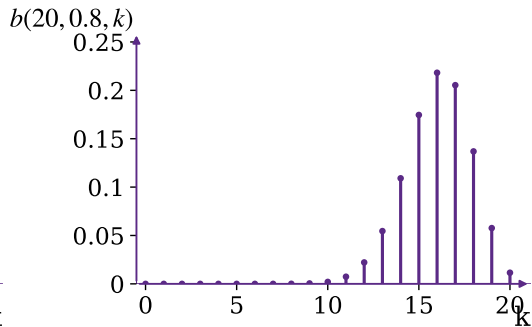
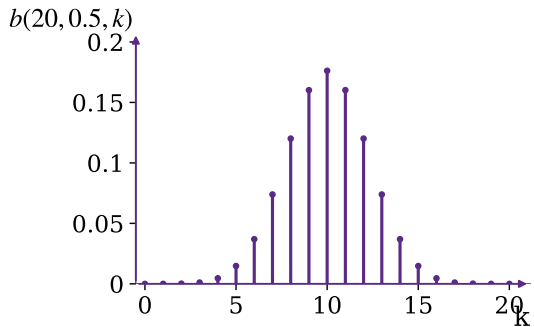
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

称 X 服从**二项分布**，记为 $X \sim B(n, p)$ 。从

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

知道这确实是一个概率函数。

二项分布分布律



例：猜牌实验

例 3

历史上，人们曾研究一些人是否具有超感知力。猜牌实验是其中一种：实验者在一个房间里注视 4 张红桃和 4 张黑桃，另一房间中的同卵双胞胎在收到公平观测者给出的信号后，猜测每张牌的颜色。兄弟知道牌堆中有 4 张红桃和 4 张黑桃。

问题：如果兄弟猜对了 6 张牌的颜色，你认为这对兄弟之间是否有超感知力？如果实验独立重复 10 次，其中 5 次猜对了至少 6 张牌的颜色，你有什么结论？

几何分布

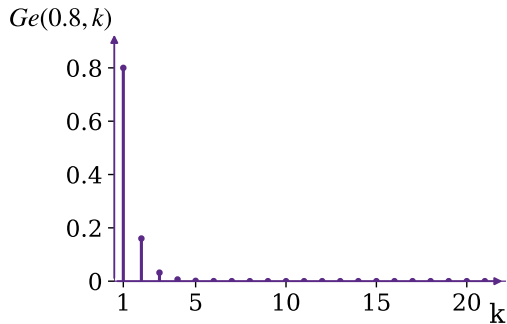
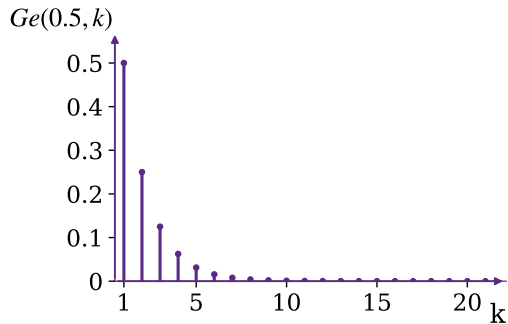
定义 2.6 几何分布（负二项分布的特例）

将 Bernoulli 试验不断独立重复下去，用 X 表示试验首次成功时需要的实验次数，则

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

称此分布为**几何分布**，记为 $X \sim G(p)$ 。

几何分布分布律



几何分布：应用实例

例 4

一个人要开门，他共有 n 把钥匙，其中仅有一把可以打开门。现随机地有放回地从中选取一把开门，若不成功再放回去重新随机选取一把开门，问这人在第 s 次才首次试开成功的概率。

几何分布的无记忆性

定理 2.1 无记忆性

以所有正整数为取值集合的随机变量 ξ 服从**几何分布** $G(p)$, 当且仅当对任何正整数 m 和 n , 都有

$$P(\xi > m + n \mid \xi > m) = P(\xi > n).$$

Pascal 分布 (负二项分布)

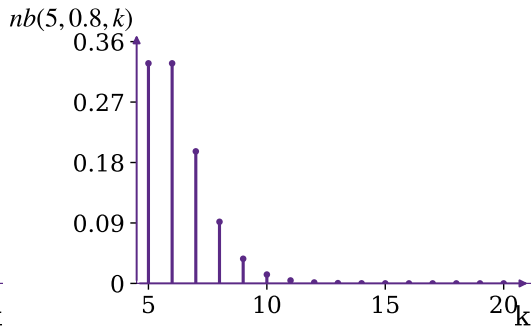
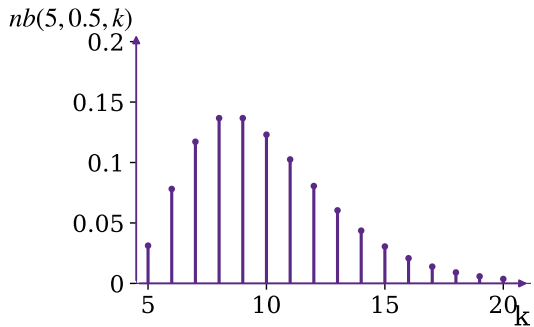
定义 2.7 负二项分布 (Pascal 分布)

将 Bernoulli 试验不断独立重复下去, 用 X_r 表示第 r 次成功发生时的试验次数, 则 X_r 的分布律为

$$\begin{aligned} P(X_r = k) &= \binom{k-1}{r-1} p^{r-1} q^{k-r} \cdot p \\ &= \binom{k-1}{r-1} p^r q^{k-r}, \quad k = r, r+1, \dots \end{aligned}$$

称此概率分布为 **Pascal 分布**, 也称为 **负二项分布**, 记为 $X_r \sim NB(r, p)$ 。

负二项分布分布律



Pascal 分布：应用实例

例 5

某人口袋里放有两盒火柴，每盒装有火柴 n 根。他每次随机取出一盒，并从中拿出一根火柴使用。试求他取出一盒，发现已空，而此时另一盒中尚余 r 根火柴的概率。

Poisson 分布

例 6

从体积为 V 的液体中随机取出体积为 D 的液体样本 ($0 < D < V$)。假设原液体中均匀、独立地分布着 N 个微生物，且每个微生物以与体积成比例的概率进入样本。在显微镜下观察样本，问发现 x 个微生物的概率是多少？

记 X 为样本中微生物的个数。由于每个微生物独立地以概率 D/V 落入样本，故

$$X \sim B\left(N, \frac{D}{V}\right).$$

于是，令 $V, N \rightarrow \infty$ ，且 $N/V = d$ 保持常数，有

$$P(X = x) = \binom{N}{x} \left(\frac{D}{V}\right)^x \left(1 - \frac{D}{V}\right)^{N-x} \rightarrow e^{-Dd} \frac{(Dd)^x}{x!}.$$

Poisson 分布的定义

定义 2.8 Poisson 分布

设随机变量 X 的概率分布为

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

则称 X 服从参数为 λ 的 **Poisson 分布**，并记 $X \sim P(\lambda)$ 。

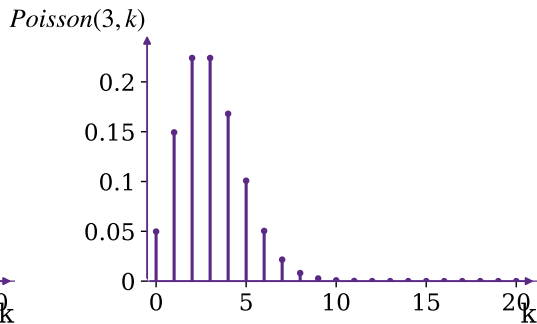
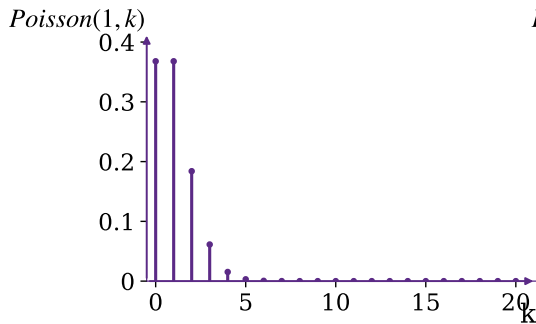
由于 $e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \dots$ ，所以 $\sum_{k=0}^{\infty} P(X = k) = 1$ 。

Poisson 分布的应用

- ▶ 电话总机在单位时间内接到的转接电话数；
- ▶ 纺纱机在单位时间内的纺纱断头数；
- ▶ 交通路口在固定时间段内发生的交通事故数；
- ▶ 放射性物质发射的射线到达计数器的 α 粒子数；
- ▶ 固定时间段内发生的地震次数。

Poisson 分布常用来描述稀有事件发生的次数。

Poisson 分布分布律



Poisson 逼近定理

定理 2.2 Poisson 逼近定理

记 X 为 n 独立 Bernoulli 试验中事件 A 发生的次数, 且记 p_n 为事件 A 发生的概率, 则 $X \sim B(n, p_n)$ 。如果 $np_n \rightarrow \lambda$, 则当 $n \rightarrow \infty$ 时,

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

建议

在实际应用中, 通常当 $n \geq 30$ 且 $np_n \leq 5$ 时即可使用 Poisson 近似。

二项分布与 Poisson 分布

区别

- ▶ **二项分布**：在固定的 n 次独立试验中，计数某类事件发生了多少次，取值范围为 $0, 1, \dots, n$ 。
- ▶ **Poisson 分布**：在固定的时间、空间或其他区间内，计数某类事件发生了多少次，取值范围为 $0, 1, 2, \dots$ 。

联系

当 n 很大、 p 很小且 $np = \lambda$ 保持稳定时，

$$B(n, p) \approx P(\lambda).$$

Poisson 逼近定理：应用实例

例 7

同类型的设备 200 台，独立工作，每台设备发生故障的概率为 0.01，在通常情况下，一台设备的故障可由一个人来处理。问至少需要配备多少工人才能保证设备发生故障而得不到维修的概率小于 0.01？

Pop quiz

问题 1 设离散型随机变量 X 的分布律为

$P(X=0) = \frac{1}{4}$, $P(X=1) = \frac{1}{2}$, $P(X=2) = \frac{1}{4}$ 。定义 $Y = \mathbf{1}_{\{X \geq 1\}}$, 则 Y 的分布律为:

(A) $P(Y=0) = \frac{3}{4}$, $P(Y=1) = \frac{1}{4}$

(B) $P(Y=0) = \frac{1}{2}$, $P(Y=1) = \frac{1}{2}$

(C) $P(Y=0) = \frac{1}{4}$, $P(Y=1) = \frac{3}{4}$

问题 2 若 $X \sim P(3)$, 则 $P(X=0)$ 等于:

(A) e^{-3}

(B) $3e^{-3}$

(C) $1 - e^{-3}$

Pop quiz

问题 1 设离散型随机变量 X 的分布律为

$P(X=0) = \frac{1}{4}$, $P(X=1) = \frac{1}{2}$, $P(X=2) = \frac{1}{4}$ 。定义 $Y = \mathbf{1}_{\{X \geq 1\}}$, 则 Y 的分布律为:

(A) $P(Y=0) = \frac{3}{4}$, $P(Y=1) = \frac{1}{4}$

(B) $P(Y=0) = \frac{1}{2}$, $P(Y=1) = \frac{1}{2}$

☒ (C) $P(Y=0) = \frac{1}{4}$, $P(Y=1) = \frac{3}{4}$

问题 2 若 $X \sim P(3)$, 则 $P(X=0)$ 等于:

☒ (A) e^{-3}

(B) $3e^{-3}$

(C) $1 - e^{-3}$