概率论与数理统计

第二章: 随机变量 (random variables)

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

古典概型中的几个问题

设 $(\Omega, \mathcal{A}, \mathrm{P})$ 为随机试验 E 的概率空间.

问题一

样本空间 Ω 中的元素与试验有关, 从数学角度看, 希望 Ω 是**抽象的集合**.

问题二

非等可能事件 的概率怎么计算?

问题三

在概率论如何使用**微积分理论**?

样本空间 Ω 抽象化

例

抛一枚硬币，考察正、反面的情况，则 $\Omega = \{H, T\}$ (有具体含意的空间). 令

$$X(\omega) = \begin{cases} 1, & \omega = H \\ 0, & \omega = T. \end{cases}$$

在上述映射下，新的“样本空间”为 $\tilde{\Omega} = \{0, 1\}$. 我们有对应关系

$$\{X = 1\} \leftrightarrow \{H\}, \quad \{X = 0\} \leftrightarrow \{T\}.$$

随机变量的定义

一个随机变量是样本空间 Ω 到实数 $\mathbb{R}$ 的映射.

例

将一枚硬币连抛三次，观察正、反面出现的情况，则样本空间为

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

定义随机变量 $X=$ 正面出现的次数，则

$$\Omega \xrightarrow{X} \Omega' = \{0, 1, 2, 3\}.$$

我们有事件:

$$\begin{aligned}
\{X=0\} &= \{TTT\}, \\
\{X=1\} &= \{HTT, THT, TTH\} \\
\{X \le 2\} &= \{TTT, HTT, THT, TTH, HHT, HTH, THH\} \\
&= \{X \le 3\} - \{X=3\} \\
&= \Omega - \{HHH\}.
\end{aligned}$$

随机变量

很多实验产生的结果本身就是随机变量

例

连续型随机变量 (continuous r.v.):

- 某地区的日平均气温 X, 日平均降水量 Y.
- 电子产品的寿命 Y.
- 某城市的日耗电量 W 是一随机变量.

离散型随机变量 (discrete r.v.):

- 一人连续对目标射击 n 次, 击中目标次数 X
- 从一大批产品中随机抽取 n 件进行测试, 其测得的次品数 N.

注

通常用大写字母 X, Y, Z, W 等表示随机变量, 用小写字母 x, y, z, w 表示实数 (非随机).

离散型随机变量

定义

若随机变量仅取有限个或可列个值，则称 X 为**离散型随机变量**.

例

- 将一枚硬币连抛三次，观察正、反面出现的情况，定义 X 为出现正面的次数. X 的取值为 $0, 1, 2, 3$，故 X 为离散型随机变量.
- 用同一支枪对目标进行射击，直到击中目标为止，则射击次数 X 是离散型随机变量.
- 114 查号台一天接到的呼叫次数 X 是离散型随机变量.

问

电子产品的寿命 X 是否是离散型随机变量? **否**.

质量函数

设 X 为离散型随机变量, 设 X 的所有可能取值为 $x_1, x_2, \ldots, x_k, \ldots$, 且

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

X 的统计规律完全由数列 $\{x_k\}$, $\{p_k\}$ 确定.

定义

称 $P(X = x_k) = p(x_k)$, $k = 1, 2, \ldots$ 为离散型随机变量 X 的**概率质量函数 (probability mass function)** 或 **频率函数**.

注

离散型随机变量的概率质量函数包括两方面:

1. 随机变量的所有取值
2. 随机变量取各个值的概率

例题

将一枚硬币连抛三次，观察正、反面出现的情况，记 X 为正面出现的次数．求 X 的概率质量函数．

解

X 的取值为 $0,1,2,3$，其样本空间为

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

故 X 的概率密度函数为

$$P(X=0)=\frac{1}{8},\quad P(X=1)=\frac{3}{8}$$
$$P(X=2)=\frac{3}{8},\quad P(X=3)=\frac{1}{8}.$$

问

频率函数有什么特点？**求和为 1**．

频率函数的性质

频率函数的基本性质

1. $p(x_k) \geq 0,\ k = 1, 2, \ldots$
2. $\sum_{k=1}^{\infty} p(x_k) = 1.$

第二点的证明

$$\sum_{k=1}^{\infty} p(x_k) = \sum_{k=1}^{\infty} \mathrm{P}(X = x_k) = \mathrm{P}\left(\bigcup_{k=1}^{\infty}\{X = x_k\}\right) = \mathrm{P}(\Omega) = 1.$$

两个方向理解基本特征

- 离散型随机变量的频率函数一定满足两个性质
- 满足两个性质的数列 $\{p_k\}$ 一定是某离散型随机变量的频率函数.

频率函数的几种表示方法

解析式法

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

列表法

分布列:

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

矩阵法

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{bmatrix}$$

离散型随机变量的概率分布规律相当于向位于 x_k 处的“盒子”中扔球.

例

一球队要经过四轮比赛才能出线. 设球队每轮被淘汰的概率为 $p = 0.5$. 记 X 表示球队结束比赛时的比赛场数, 求 X 的频率函数.

解

随机变量 X 的可能取值为 $1, 2, 3, 4$. 记 A_k 为球队通过第 k 轮比赛的事件, $k = 1, 2, 3, 4$. 则由独立性,

$$
\begin{aligned}
\mathrm{P}(X=1) &= \mathrm{P}(\bar{A}_1) = p \\
\mathrm{P}(X=2) &= \mathrm{P}(A_1\bar{A}_2) = \mathrm{P}(A_1)\mathrm{P}(\bar{A}_2) = (1-p)p \\
\mathrm{P}(X=3) &= \mathrm{P}(A_1A_2\bar{A}_3) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(\bar{A}_3) = (1-p)^2p \\
\mathrm{P}(X=4) &= \mathrm{P}(A_1A_2A_3) = (1-p)^3.
\end{aligned}
$$

代入 $p = 0.5$, 所求的 X 之频率函数为

X	1	2	3	4
p_k	0.5	0.25	0.125	0.125

例

掷两个均匀的骰子，观测其点数，令 X 为两骰子点数之和. 求

a) X 的频率函数

b) X 为奇数的概率.

解

a)

$X=k$	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b)

$$p_{odd} = \frac{2+4+6+4+2}{36} = \frac{1}{2}.$$

分布函数

思考

随机变量 $X = X(\omega)$ 是样本点 ω 的函数，是“随机函数”，不方便应用微积分工具. 怎样将“随机函数”化为“普通函数”?

观察

对于随机变量 X, $\forall x \in (-\infty, \infty)$, 集合

$$\{X \leq x\} = \{\omega \mid X(\omega) \leq x\} \in \mathcal{A}$$

是事件.

定义

称函数 $F(x) = \mathrm{P}(X \leq x)$, $-\infty < x < \infty$, 为随机变量 X 的**累积分布函数 (CDF, cumulative distribution function)**, 简称**分布函数**.

注

随机变量的分布函数是关于自变量 x 的普通函数，它不再是随机的!

例题

设 X 的频率函数为

X	1	2	3
p_k	0.3	0.2	0.5

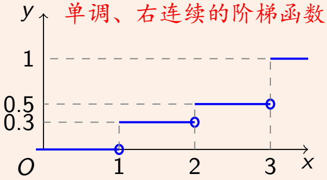
求 X 的分布函数.

解

由分布函数的定义有

$$
\begin{aligned}
F(x) &= \mathrm{P}(X \le x) = \mathrm{P}(\varnothing) = 0, \quad x < 1 \\
F(x) &= \mathrm{P}(X \le x) = \mathrm{P}(X = 1) = 0.3, \quad 1 \le x < 2 \\
F(x) &= \mathrm{P}(X \le x) = \mathrm{P}(X = 1,\ 2) = 0.5, \quad 2 \le x < 3 \\
F(x) &= \mathrm{P}(X \le x) = \mathrm{P}(\Omega) = 1, \quad 3 \le x.
\end{aligned}
$$

于是，$F(x) = \begin{cases} 0, & x < 1, \\ 0.3, & 1 \le x < 2, \\ 0.5, & 2 \le x < 3, \\ 1, & 3 \le x. \end{cases}$



分布函数的基本性质

$$F(x) = \mathrm{P}(X \le x), \quad -\infty < x < \infty.$$

分布函数的基本性质

1. $F(x)$ 是单调不减函数
2. $0 \le F(x) \le 1$ 且
$$F(-\infty) := \lim_{x\to-\infty} F(x) = 0, \quad F(\infty) := \lim_{x\to\infty} F(x) = 1.$$
3. $F(x)$ 为右连续函数 (或右边左极, càdlàg),
$$F(x+0) := \lim_{t\to x+0} F(t) = F(x).$$

注

以上性质是分布函数的本质特征:

- 随机变量的分布函数必满足这三条性质
- 满足这三条性质的 $F(x)$ 必是某随机变量的分布函数.

问

怎样利用分布函数计算概率

$$P(a < X \leq b), \quad a < b?$$

分析

$$\{a < X \leq b\} = \{X \leq b\} - \{X \leq a\}$$
$$\Rightarrow P(\{a < X \leq b\}) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

问

怎样计算概率 $\mathrm{P}(X=c)$ (c 为常数)?

分析

由可列可加性, 对于单调递减的可列集合列 $A_1 \supset A_2 \supset A_3 \supset \cdots$, 有

$$\mathrm{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathrm{P}(A_n).$$

故

$$\begin{aligned}
\mathrm{P}(X=c) = \mathrm{P}\left(\bigcap_{n=1}^{\infty}\{c-\frac{1}{n}<x\le c\}\right) &= \lim_{n\to\infty} \mathrm{P}(c-1/n<x\le c) \\
&= \lim_{n\to\infty}\Big[F(c)-F(c-1/n)\Big] \\
&= F(c)-F(c-0).
\end{aligned}$$

注

若 $F(x)$ 在 $x=c$ 处连续, 则 $\mathrm{P}(X=c)=0$. 否则 $\mathrm{P}(X=c)>0$.

单点分布

定义

对常数 c, 如果随机变量 X 的频率函数为

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

则称随机变量 X 服从单点分布.

注

- 严格说单点分布不具有 "随机性", 视为随机变量是为了理论上的需要.
- 单点分布有时也称为 "退化分布".

定义

某事件发生的概率为 1, 则称该事件 "几乎处处" (a.e.) 或 "几乎必然" (a.s.) 发生. 如

$$\begin{aligned} P(X=c)=1 \quad &\Rightarrow \quad X \overset{\text{a.e.}}{=} c, \text{ or } X=c, \text{ a.e.} \\ P(X=Y)=1 \quad &\Rightarrow \quad X \overset{\text{a.e.}}{=} Y, \text{ or } X=Y, \text{ a.e.} \end{aligned}$$

两点分布 (伯努利随机变量)

定义

如果随机变量 X 的频率函数为

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

则称随机变量 X 服从**0-1 两点分布** 或 **Bernoulli 分布**, 其中 $0<p<1$ 为常数.

注

若一个试验只产生两个结果, 则可以用服从 Bernoulli 分布的随机变量来描述.

例子

- 一门课的考试是“及格”还是“不及格”.
- 刚出生的新生儿是“男”还是“女”.
- 产品检验的结果是“合格”还是“不合格”.
- 射击结果是“击中目标”还是“没击中目标”.

例: 收入分布

我国 2012 年家庭人均收入 R (千元) 分布如下:

R	1	2	4.5	9	15.9	25.8	34.3
收入低于 R 的家庭比例	0.05	0.1	0.25	0.5	0.75	0.9	0.95

若规定家庭人均收入 2 千元为贫困线, 又令

$$X = \begin{cases} 1, & R \leq 2, \\ 0, & R > 2. \end{cases}$$

则 X 服从参数为 0.1 的两点分布.

注

X 也可以看作某家庭是否是贫困家庭的 "示性函数" (indicator function):

$$X = \mathbf{1}_{\{R \text{ 在贫困以下}\}}.$$

伯努利试验与二项分布

伯努利试验

只产生两个结果 $A, \bar{A}$ 的试验.

伯努利实验产生的两点分布.

n 重伯努利实验

将伯努利实验独立并重复进行 n 次的试验.

例

- 某战士用步枪对目标进行射击, 记 A 为击中目标的事件. 每射击就是一个伯努利试验. 如果对目标进行 n 次射击, 则是一个 n 重伯努利实验.
- 从一批产品中随机抽取一个产品进行检验, 记 A 为产品合格的事件, 则每检验一个产品就是一个伯努利试验.

伯努利实验分析

在伯努利试验中，令

$$P(A) = p, \quad P(\bar{A}) = 1 - p$$

注

重复 每次试验中概率 $P(A)$ 保持不变

独立 各次试验的结果互不影响.

记 A_i 为第 i 次试验的结果，$i = 1, 2, \ldots, n$. 由独立性，有

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}), \quad 1 \le i_1 < i_2 < \cdots < i_k \le n.$$

定义

令 X 为 n 重伯努利试验中事件 A 发生的次数，则 X 是一个离散型随机变量，称为**二项分布**.

二项分布的频率函数

记 X 为 n 重伯努利试验中事件 A 发生的次数.

X 的取值为 $0, 1, 2, \ldots, n$.

二项分布的频率函数

$\{X = k\}$ 发生意味着 A 发生 k 次, $\bar{A}$ 发生 $(n-k)$ 次, 故由有限可加性

$$\begin{aligned}
\mathrm{P}(X = k) &= \mathrm{P}\Big(\bigcup_{1 \le i_1 < \cdots < i_k \le n} A_{i_1} \cdots A_{i_k} \bar{A}_{j_1} \cdots \bar{A}_{j_{n-k}} \Big) \\
&= \sum_{1 \le i_1 < \cdots < i_k \le n} \mathrm{P}\Big(A_{i_1} \cdots A_{i_k} \bar{A}_{j_1} \cdots \bar{A}_{j_{n-k}} \Big) \\
&= \sum_{1 \le i_1 < \cdots < i_k \le n} p^k (1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^k, \quad k = 0, 1, \ldots, n.
\end{aligned}$$

定义

若随机变量 X 的频率函数为

$$\mathrm{P}(X=k)=\binom{n}{k}p^k q^{n-k},\quad k=0,1,\ldots,n,\ q=1-p,$$

则称 X 服从参数为 (n,p) 的**二项分布**，记为 $X\sim \mathrm{Bin}(n,p)$. 特别地，当 $n=1$ 时，$\mathrm{Bin}(1,p)$ 就是两点分布，即

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

注

由牛顿二项式定理，易见

$$\sum_{k=0}^{n}\mathrm{P}(X=k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=\left[p+(1-p)\right]^n=1.$$

最可能出现次数

问题

令

$$b(k; n, p) := \binom{n}{k} p^k (1-p)^{n-k}.$$

考察当 n, p 固定时, $b(k; n, p)$ 随 k 的变化情况.

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{n!/k!(n-k)!}{n!/(k-1)!(n-k+1)!} \cdot \frac{p}{q} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p - k}{q}.$$

- 当 $k < (n+1)p$ 时, $b(k; n, p)$ 随 k 增加而增加.
- 当 $k = (n+1)p$ 时, $b(k; n, p) = b(k-1; n, p)$.

定义

$m = [(n+1)p]$ ($[x]$ 表示不超过 x 的最大整数) 为**最可能出现次数**, $b(m; n, p)$ 为**中心项**.

例题

一大批电子元件有 10% 已经损坏. 若从这批元件中随机选取 20 只来组成一个线路, 问此线路能正常工作的概率是多少?

解

因为元件的数量很大, 所以取 20 只元件可看作是有放回抽样, 可以用 n 重伯努利试验刻画.
记 X 为 20 只元件中合格品的数量, 则 $X \sim \mathrm{Bin}(20, 0.9)$, 因此

$$\mathrm{P}(\text{线路正常}) \approx \mathrm{P}(X = 20) = 0.9^{20} \times 0.1^{0} = 0.9^{20} \approx 0.1216.$$

例题

已知 100 个产品中有 5 个次品，现从中有放回地取 3 次，每次任取 1 个. 求所取的 3 个中恰有 2 个次品的概率.

解

有放回地抽样是条件相同且独立的试验，它是伯努利试验. 设 X 为所取的 3 个中次品数，则由题意 $X \sim \mathrm{Bin}(3, 0.05)$，因此所求概率为

$$P(X=2) = \binom{3}{2}(0.05)^2 0.95 = 0.007125.$$

注

若本例改为“无放回”，则此试验不是伯努利试验，只能用更一般的古典概型求解:

$$P(X=2) = \frac{\binom{95}{1}\binom{5}{2}}{\binom{100}{3}} \approx 0.00618.$$

例: 加州抢劫案 (60 年代)

据报道, 加州某地发生一起抢劫案. 目击嫌疑人有两个: 一个男的理平头黑人, 一个女的黑发梳马尾型. 不久抓到一对具有上述特征的夫妇 (情侣), 能否判断他们有罪?

有罪的论点

数学家通过计算机模拟, 得出一对夫妇具有一述特征的概率为 $p = 8.3 \times 10^{-8}$. 这是一个小概率事件. 陪审团在无其它证据的情况下, 裁决他们有罪.

无罪的论点

加州高院推翻了该裁决. 高院认为犯罪的认定应当具有唯一性. 用上列算式: 设 X 为具有上述特征的夫妇数, 则 $X \sim \mathrm{Bin}(n, p)$. 要计算

$$\mathrm{P}\left(X > 1 \mid X \geq 1\right) = \frac{\mathrm{P}(X > 1)}{\mathrm{P}(X \geq 1)} = \frac{\mathrm{P}(X \geq 1) - \mathrm{P}(X = 1)}{\mathrm{P}(X \geq 1)} = \frac{1 - (1-p)^n - np(1-p)^{n-1}}{1 - (1-p)^n}.$$

n 的估计为 8.3×10^6, 而上式约为 0.2966, 不算小.