
离散型随机变量与分布函数

第 4 讲 · 概率论与数理统计

复旦大学经济学院 ECON130001

上一讲我们做了什么

让概率随信息更新

条件概率 $P(A | B) = \frac{P(AB)}{P(B)}$ 即把样本空间缩到 B ；由因算果用全概率公式，由果推因用贝叶斯定理。

也看清了独立性的两面

独立 $\iff P(AB) = P(A)P(B)$ ，它让 n 重伯努利试验可算；但这一假设不成立时，风险可能被低估十几个数量级。

★ 留下的问题

上一讲最后那个式子 $p_n(k) = C_n^k p^k (1-p)^{n-k}$ ，把「恰好成功 k 次」的概率写成了 k 的一个函数。

与其每次都问「某个事件的概率是多少」，不如**一次性描述一个数量的全部可能取值及其概率**。这个「数量」该怎么定义？

为什么需要这一讲

背景 一位基金经理，**连续 10 年跑赢基准指数**。基金公司把这一纪录写进宣传材料，他管理的资金规模翻了几番。

转折 有人算了一笔账：如果他毫无选股能力，每年能否跑赢就像抛一次硬币，概率为 $1/2$ ，且各年互不影响。那么连赢 10 年的概率是 $(1/2)^{10} = 1/1024 \approx 0.1\%$ 。**概率只有千分之一的事发生了，这似乎不能用运气解释。**这个算式**没有错**，但结论错了。

问题 **全市场有几千位基金经理。在几千次机会里，出现一件概率只有千分之一的事，几乎是必然的。**要把这一点算清楚，只算「某一个事件的概率」不够，需要一次性掌握「跑赢年数」这个数量的整张概率表。

来源与说明

本页的概率为教学化演算： $(1/2)^{10} = 1/1024$ ；若按 5000 位经理计，至少一人 10 年全胜的概率为 $1 - (1 - 1/1024)^{5000} \approx 99.2\%$ 。这些数字用于说明「多重比较」这一统计现象，不涉及任何具体机构或个人。第 11 讲（假设检验）还会回到这个问题。

本讲学习目标

- ① 说清**随机变量**是什么——它是一个函数，不是一个数
- ② 掌握四个基本离散分布：**伯努利、几何、二项、泊松**，知道各自描述哪一类随机现象
- ③ 会用**分布函数** $F(x) = P(X \leq x)$ 统一描述离散与连续两类随机变量

为什么要引入「随机变量」

此前我们关注的是具体事件

抛 3 次硬币, $A = \{\text{恰好 2 次正面}\} = \{\text{正正反, 正反正, 反正正}\}.$

经济与金融

影响股价的新闻（具体事件）有很多条，基金经理最终关注的是**涨跌幅**——一个数值。

人工智能

垃圾邮件过滤器最终输出的是一**封邮件属于垃圾邮件的得分**——也是一个数值。

本讲的核心任务

把样本空间里的每个结果（如「正正反」）映射到一个我们关心的数（如 2）。**这样的映射（函数）就是随机变量。**

随机变量的定义

定义

定义在样本空间上的**实值函数**，称为**随机变量**。

常用大写字母 X, Y, Z 或希腊字母 η, ζ, ξ 表示。

严格地说，随机变量还需满足可测性要求（对任意实数 x ， $\{X \leq x\}$ 都是事件）；本课程涉及的函数均满足。

★ 要点：随机变量是函数

X **不是**一个数，是一条规则：给定 $\omega \in \Omega$ ，得到一个实数 $X(\omega)$ 。「随机」体现在 ω 是随机抽到的， X 这条规则本身**完全确定**。

例：抛硬币 3 次， X = 正面次数，则 $X(\text{正正反}) = 2$ ， $X(\text{反反反}) = 0$ 。

事件可以用随机变量来写

$\{X = 2\}$ 就是事件 {正正反, 正反正, 反正正} 的简写。**有了随机变量，事件不必再列举结果，写成等式或不等式即可。**

分布：一次性描述全部取值

定义

随机变量 X 的**分布**，指对实数的任意子集 C ，事件 $\{X \in C\}$ 的概率 $P(X \in C)$ 的全体。

离散型

只能取**有限个或可列无穷个**值。

例：抛硬币的正面次数、一天的索赔件数。

连续型

在一个**区间**上可以取所有值。

例：股票收益率、等车时间。

第 5 讲专门讨论

回到开场

「跑赢年数」 X 只能取 $0, 1, \dots, 10$ ，是离散型。我们需要的是它的**整张分布表**，而不只是 $P(X = 10)$ 这一个数。

离散型的分布怎么写：概率函数

定义 (probability function, p.f.)

对离散型随机变量 X 和任意实数 x , 定义

$$f(x) = P(X = x)$$

称 f 为 X 的**概率函数** (也叫分布律)。

两条性质 (由概率的公理得到)

记 $p_k = f(x_k)$, 则

- $p_k \geq 0$, 且对不可能的取值 $f(x) = 0$

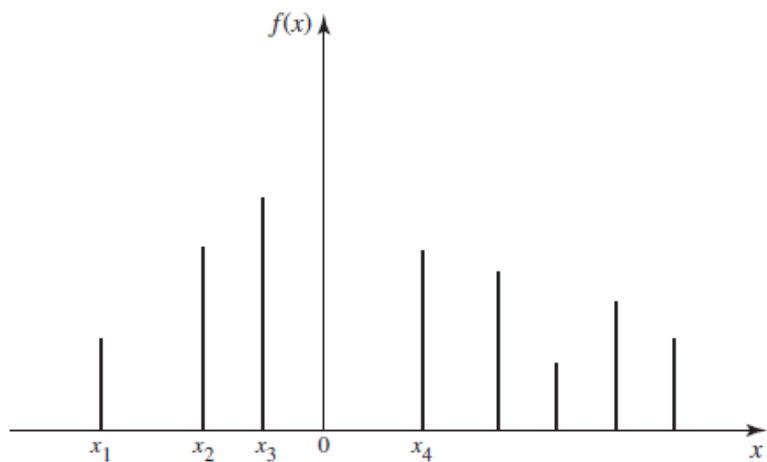
← 非负性

- $\sum_k p_k = 1$

← 规范性 $P(\Omega) = 1$ + 可列可加性

反过来也成立：任何满足这两条的 $\{p_k\}$ 都是某个随机变量的概率函数。

概率函数的三种写法：离散型均匀分布



概率函数图

概率分布表

X	x_1	x_2	$\dots$	x_k
p	p_1	p_2	$\dots$	p_k

函数式、图、表三种写法包含的信息相同，视场合选用。

例：离散型均匀分布

某彩票从 000–999 中等可能地开出一个号码 X ，其概率函数为

$$f(x) = \frac{1}{1000}, \quad x = 0, 1, \dots, 999$$

若 X 只取 n 个值，且取每个值的概率都是 $1/n$ ，则称 X 服从**离散型均匀分布**。

两道例题：写出概率函数，确定常数

例：写出概率函数

某人每期买一张彩票，每期中奖的概率都是 p ，各期相互独立。没中就继续买，直到中奖为止。求购买次数 ξ 的概率函数。

前 $k - 1$ 次都没中、第 k 次中：

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

这就是稍后要讲的几何分布

例：定常数 c

(1) $f(x) = cx, x = 1, \dots, 5$:

$$c(1 + 2 + 3 + 4 + 5) = 1 \Rightarrow c = \frac{1}{15}$$

(2) $f(x) = c/2^x, x = 0, 1, 2, \dots$:

$$c \sum_{x=0}^{\infty} \frac{1}{2^x} = 2c = 1 \Rightarrow c = \frac{1}{2}$$

(2) 用到了等比级数求和。**离散型的取值可以有可列无穷个**，这时归一化条件要求一个无穷级数的和等于 1。

分布之一：伯努利分布（0-1 分布）

定义

若 X 只取 0, 1 两个值，且取值的概率如下，则称 X 服从参数为 p 的**伯努利分布**：

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

它是后面三个分布的基本单元

接下来的三个分布都从伯努利试验出发：

- **几何分布**：一直试到第一次成功，试了几次
- **二项分布**：固定试 n 次，成功几次
- **泊松分布**： n 很大、 p 很小时二项分布的极限

伯努利分布：从信用违约到 AI 二分类

信用风险评估

- 银行最关心的问题：这个客户未来会违约吗？
- 令 $X = 1$ 表示违约、 $X = 0$ 表示不违约， X 就是伯努利随机变量
- **信用评分模型的任务是估计每个客户的**
 $p = P(X = 1)$ ，利率与风险拨备都据此确定

AI 的二分类

- 邮件是垃圾 ($X = 1$) 还是正常 ($X = 0$) ?
- 影像是否显示异常？
- 用户下个月会流失还是留存？

AI 分类模型输出的是伯努利参数 p 的估计值：根据输入特征，把这个 p 估得尽可能准。

上一讲的贝叶斯分类器，对一封邮件算出的是 $p = P(\text{垃圾邮件} \mid \text{邮件中的词})$ 。有了这个 p ，「这封邮件是不是垃圾邮件」就是一个参数为 p 的伯努利随机变量。

分布之二：几何分布

从一个试验说起

机器生产的每件产品是次品的概率为 10%，各件独立。一直生产到**出现第一件次品为止**，令 X 为已生产的件数。 X 服从什么分布？

几何分布

若 X 的概率函数如下，则称 X 服从参数为 p 的**几何分布**：

$$f(x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, \dots \\ 0, & \text{其他} \end{cases}$$

含义：前 $x - 1$ 次都失败（每次概率 $1 - p$ ，独立相乘），第 x 次成功。

几何分布：从求职时间到 A/B 测试

求职时间的建模

- 毕业生每月找到满意工作的概率都是 p ，各月独立，则所需月数 X 服从几何分布
- **理解摩擦性失业**：算 $P(X > 12) = (1 - p)^{12}$ ，即一年后仍未找到工作的比例
- **评估政策**：若某政策把 p 从 0.2 提到 0.3， $P(X > 12)$ 从 6.9% 降到 1.4%

A/B 测试与用户行为

- 用户每次看到推荐时点击的概率为 p
- 用户在第 X 次看到推荐时才首次点击， X 服从几何分布
- A/B 测试：新旧算法各分一组用户，比较两组 X 的分布， X 整体偏小的一组，推荐更有吸引力

注意 $P(X > k) = (1 - p)^k$ 这个式子： $\{X > k\}$ 就是「前 k 次都失败」，各次概率直接相乘即得，不必把 $x > k$ 的无穷多项加起来。下一页要用到它。

几何分布的无记忆性

无记忆性（也称无后效性）

若 X 服从几何分布，则对任意 $s, t \in \mathbb{N}$

$$P(X > s + t \mid X > t) = P(X > s)$$

证明思路

$P(X > k) = (1 - p)^k$ 。代入条件概率定义：
$$\frac{P(X > s + t)}{P(X > t)} = \frac{(1 - p)^{s+t}}{(1 - p)^t} = (1 - p)^s = P(X > s)。$$

这条性质的含义

「已经等了 t 次还没成功」这一信息，对预测今后没有帮助。再等 s 次仍不成功的概率，与从头开始等 s 次相同。

对照第 3 讲：各次试验相互独立，**之前失败多少次，都不改变以后每次成功的概率 p** 。这比牙膏那题的马尔可夫性质更强：那里下一次买哪种牙膏的概率取决于这一次买的是哪种，这里则与已经等了多久无关。

用在求职模型上，这个性质**多半不成立**：现实中失业越久越难找到工作（技能折旧、雇主筛选）， p 并非不变。**用几何分布建模之前，要先检验「 p 不变」这条假设。**

分布之三：二项分布

由伯努利分布构造

机器次品率 $p = 10\%$ ，各件独立，检查 n 件产品。令 $X_i = 1$ 表示第 i 件是次品（否则 $X_i = 0$ ），则 $X_1, \dots, X_n$ **独立同分布**，都服从参数为 p 的伯努利分布。令 $X = X_1 + \dots + X_n$ 为次品总数， X 服从什么分布？

二项分布 $X \sim B(n, p)$

$$f(x) = \begin{cases} C_n^x p^x (1-p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{其他} \end{cases}$$

它就是上一讲伯努利定理中的 $p_n(k)$ ，**现在作为随机变量的分布，有了专门的名称。**

$n = 1$ 时它就是伯努利分布。一般地，**二项分布可以看作「 n 个独立同分布的伯努利变量之和」。**

与几何分布的区别：几何固定目标（第一次成功），数次数， X 无上限；**二项固定次数 n ，数成功， X 最多是 n 。**哪个量固定、哪个量随机，两者正好相反。

回到开场：用二项分布重新计算

设定

假设基金经理毫无选股能力，每年跑赢基准的概率 $p = 0.5$ ，各年独立。10 年中跑赢的年数 $X \sim B(10, 0.5)$ 。

单看一个人

$$P(X = 10) = C_{10}^{10}(0.5)^{10} = \frac{1}{1024} \approx 0.1\%$$
$$P(X \geq 8) = \frac{C_{10}^8 + C_{10}^9 + C_{10}^{10}}{1024} = \frac{56}{1024} \approx 5.5\%$$

确实罕见。

但市场上有几千个人

设有 5000 位经理，**纯靠运气**：

- 10 年全胜的平均人数： $5000/1024 \approx 5$ 人
- 至少有一人 10 年全胜的概率：
 $1 - (1 - \frac{1}{1024})^{5000} \approx 99.2\%$
- 赢 8 年及以上的平均人数：
 $5000 \times 56/1024 \approx 273$ 人

在几千人的行业里，「连赢 10 年」几乎必然出现在某个人身上，所以不能仅凭这一纪录断定他有选股能力：**判断一件事是否罕见，要看一共有多少次机会让它发生。**

二项分布：风险管理与模型评估

信用组合违约风险

- 持有 $n = 50$ 家公司的债券组合，各家一年内违约概率 $p = 0.04$ ，相互独立
- 违约家数 $X \sim B(50, 0.04)$
- 违约超过 2 家的概率

$$P(X \geq 3) = 1 - \sum_{k=0}^2 C_{50}^k (0.04)^k (0.96)^{50-k}$$

是设定风险拨备与对冲规模的依据之一

评估 AI 分类器

- 反欺诈模型宣称准确率 $p = 0.95$ ，用 $n = 20$ 个已知样本测试，只识别对 16 个
- 若宣称属实，识别数 $X \sim B(20, 0.95)$,

$$P(X \leq 16) = \sum_{k=0}^{16} C_{20}^k (0.95)^k (0.05)^{20-k} \approx 1.6\%$$

- 这么差的结果很少出现，就有理由**质疑它宣称的性能**

右边的推理是：**若宣称属实，测出这么差的结果的概率只有 1.6%，所以怀疑宣称。**上一页却说，「连赢 10 年」这样的小概率事件不能说明能力。两者的区别在于机会的多少：上一页是在 5000 人里挑出表现最好的一个，几千次机会里出现一次千分之一的事并不奇怪；这里只检验了一个模型、只测了一次。

归一化为什么自动成立

二项式展开定理

对任意 p, q 与正整数 n

$$(p + q)^n = \sum_{k=0}^n C_n^k p^k q^{n-k}$$

取 $q = 1 - p$

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

二项分布的归一化条件是二项式定理的直接推论。 二项分布的各个概率恰好是 $(p + q)^n$ 展开式中的各项，这一分布因此得名。

分布之四：为什么还需要泊松分布

二项分布的取值有上限

$X \sim B(n, p)$ 最多取到 n 。但有些计数**事先给不出上限**：

- 一小时内到店的顾客数
- 一天内通过某路口的车辆数
- 一年内发生自然灾害的次数

我们需要一种新的离散分布，来描述**大量重复试验中「稀有事件」发生的次数**——试验次数 n 极大、单次概率 p 极小，而平均次数 np 保持在一个适中的量级。

泊松分布是怎么被推出来的

店主的问题与做法

店主估计顾客平均每小时来 4.5 名，**想知道一小时内顾客数 X 的分布**。他把一小时切成 3600 秒，假设每秒最多来一位顾客、各秒独立，每秒来人的概率 $p = 4.5/3600 = 0.00125$ ，于是 $X \sim B(3600, 0.00125)$ 。

但这个做法有两个问题

C_{3600}^x 难以计算；而且按这个模型，顾客数不可能超过 3600。

观察相邻两项之比：
$$\frac{f(x+1)}{f(x)} = \frac{(n-x)p}{(x+1)(1-p)} \approx \frac{np}{x+1}$$
—— n 与 p 只以乘积 np 出现。

令 $\lambda = np$ ，递推

$f(1) = \lambda f(0)$, $f(2) = \frac{\lambda^2}{2} f(0)$, $f(3) = \frac{\lambda^3}{6} f(0)$,一般地 $f(x) = \frac{\lambda^x}{x!} f(0)$ 。

由归一化与 $\sum_{x \geq 0} \lambda^x / x! = e^\lambda$ 得 $f(0) = e^{-\lambda}$ ，于是 $f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ 。

泊松分布及其与二项分布的关系

定义

若 X 的概率函数如下 ($\lambda > 0$) , 则称 X 服从参数为 λ 的**泊松分布**:

$$f(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{其他} \end{cases}$$

泊松定理 (二项分布的极限)

令 $\lambda_n = np_n$ 。若 $n \rightarrow \infty$ 且 $\lambda_n \rightarrow \lambda$, 则 $B(n, p_n)$ 趋于参数为 λ 的泊松分布。

证明思路

$$C_n^x p_n^x (1 - p_n)^{n-x} = \frac{\lambda_n^x}{x!} \cdot \underbrace{\frac{n}{n} \frac{n-1}{n} \dots \frac{n-x+1}{n}}_{\rightarrow 1} \cdot \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^n}_{\rightarrow e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{-x}}_{\rightarrow 1} \rightarrow \frac{e^{-\lambda} \lambda^x}{x!}$$

三个花括号分别趋于 1、 $e^{-\lambda}$ 、1。

实用判据: $n \geq 20$ 且 $p \leq 0.05$ 时, 用 $\lambda = np$ 的泊松分布近似二项分布, 误差已经很小。

交互：离散分布实验室

怎么用

切换分布，拖动参数看形状变化。

当堂要做的四个实验

- 1. 二项： p 拖到 0.5，左右对称；拖向两端，变得偏斜
- 2. 几何：概率**单调递减**，最可能的取值总是 1
- 3. 泊松： λ 调大，峰**右移**，形状趋于对称
- 4. 二项：勾选「叠加泊松逼近」， n 拖大、 p 拖小，柱形与折线**逐渐重合**



交互演示：四个离散分布参数联动，含泊松逼近二项

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/discrete_lab.html

泊松分布例题（一）

例：近似的精度

某事件在人群中的发生率为 0.01。随机抽 200 人，求其中至少 4 人发生该事件的概率。

精确（二项 $B(200, 0.01)$ ）：

$$P(X \geq 4) = 1 - \sum_{k=0}^3 C_{200}^k (0.01)^k (0.99)^{200-k} \approx 0.1420$$

近似（泊松 $\lambda = np = 2$ ）： $P(X \geq 4) \approx 0.1429$ 。 **差 0.0009——完全够用，而计算量小得多。**

例：一页书的错印

一页书平均有 λ 处错印，错印数服从泊松分布。求某页**没有**错印的概率。

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}$$

若 $\lambda = 0.5$ ，则 $e^{-0.5} \approx 0.607$ ：平均每页 0.5 处错印，仍有约六成的页面没有错印。

泊松分布例题（二）：两个分布套在一起

例

一本书 n 页，各页错误数相互独立，都服从参数为 λ 的泊松分布。求**至少有 m 页**错误多于 k 个的概率。

解：分两层

第一层（页内，泊松）：某一页错误多于 k 个的概率

$$p = \sum_{i=k+1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!}$$

第二层（页间，二项）：整本书里「错误多于 k 个」的页数 $X \sim B(n, p)$,

$$P(X \geq m) = \sum_{x=m}^n C_n^x p^x (1-p)^{n-x}$$

例：航空公司超售

198 座的航班卖出 200 张票，每位乘客不出现的概率为 1%，相互独立。要让到场乘客都有座位，需**至少 2 人不出现**： $X \sim B(200, 0.01) \approx \text{Poisson}(2)$ ， $P(X \geq 2) = 1 - 3e^{-2} \approx 0.5940$ ——**只超售 2 张，就有约四成的概率有人没有座位。**

泊松分布：流量与事件模型

保险精算

- 车险公司平均每天 $\lambda = 10$ 起索赔，每天索赔数 $X \sim \text{Poisson}(10)$
- **资本充足**：确定需要准备多少理赔资金，即求最小的 k 使 $P(X \leq k) \geq 99.9\%$
- **异常预警**：某天接到 30 起，而 $P(X \geq 30)$ 极小 $\rightarrow$ 判为异常，可能预示突发灾害天气

网站流量与攻击检测

- 服务器平均每分钟 $\lambda = 500$ 次请求
- **容量规划**：算 $P(X > 600)$ ，指导带宽与机器数
- **攻击检测**：某分钟突增到 3000 次，而 $\lambda = 500$ 时 $P(X \geq 3000)$ 几乎为 0 $\rightarrow$ 自动判定为攻击并告警

两个场景是同一个模式：**用泊松分布刻画「正常水平」，再用尾概率定义「异常」**。许多异常检测规则都采用这种结构。

接下来

到这里，离散型随机变量已经能用概率函数 $f(x) = P(X = x)$ 完整描述了。但这套写法对连续型行不通。

例：等车时间

公交车每 10 分钟一班，乘客随机到站，等车时间 X 落在 $[0, 10]$ 任一子区间的概率与区间长度成正比（第 2 讲的几何概型）。 X 恰好等于 3 分钟的概率是多少？

对任意 $0 < h < 7$, $\{X = 3\} \subset \{3 \leq X \leq 3 + h\}$, 由第 1 讲的单调性,

$$P(X = 3) \leq P(3 \leq X \leq 3 + h) = \frac{h}{10}$$

h 可以任意小，所以 $P(X = 3) = 0$ 。同样的推理对每一点都成立。

每一点的概率都是 0, $f(x) = P(X = x)$ 全等于 0, 描述不了这个分布；可 X 落在 $[3, 5]$ 内的概率是 $\frac{2}{10}$, 并不是 0。我们需要一种对离散型与连续型都适用的描述方式。

分布函数：离散型与连续型通用

定义 (cumulative distribution function, c.d.f.)

对随机变量 X 和任意实数 x , 定义

$$F(x) = P(X \leq x)$$

称 F 为 X 的 **(累积) 分布函数**。

为什么两类都适用

$F(x)$ 是「**不超过** x 」的概率，而不是「恰好等于 x 」的概率。对连续型，前者随 x 变化，不会处处为 0。

例

X 服从参数为 p 的 0-1 分布: $P(X = 1) = p$, $P(X = 0) = 1 - p$ 。其分布函数为

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

F 是**阶梯函数**：在每个取值点向上跳跃，跳跃高度等于该点的概率。

分布函数的四条性质

- $0 \leq F(x) \leq 1$, 对一切 $x \in (-\infty, +\infty)$ 成立
- F 是 x 的**不减**函数
- $F(-\infty) = 0$, $F(+\infty) = 1$
- F 至多有**可列个**间断点, 且在间断点处**右连续**

证明思路

第一条因为 $F(x)$ 是概率; 第二条由第 1 讲的单调性得到; 第三条与右连续要用到可列可加性。第四条: 不减函数的间断点都是跳跃间断, 各跳跃区间互不相交, 所以至多可列个; 右连续来自 $\{X \leq x\}$ 包含端点 x 。

★ 「右连续」这一条不是技术细节

它来自定义中的「 $\leq$ 」。若把定义改为 $F(x) = P(X < x)$, F 就是左连续的, **下一页的几个公式也要相应改写**。因此, 下一页的公式都以 $F(x) = P(X \leq x)$ 这一定义为前提。

由分布函数确定概率

对随机变量 X ：

- $P(X > x) = 1 - F(x)$
- $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
- $P(X < x) = F(x^-)$ （左极限）
- $P(X = x) = F(x) - F(x^-)$ （该点的跳跃高度）

用最后一条统一看待两类随机变量

离散型： F 在取值点跳跃，跳跃高度就是 $P(X = x)$ 。

连续型： F 处处连续，于是 $P(X = x) = 0$ ——每一点的概率都是 0，但落在一个区间里的概率可以不是 0。

第 27 页的等车时间就是这样： $P(X = 3) = 0$ ，而 $P(3 < X \leq 5) = \frac{2}{10}$ 。连续型并不是「没有概率」，而是**概率不集中在单个点上，分布在区间上**。

离散型的分布函数：画两个图

对离散型随机变量 X 及其概率函数 p_k ,

$$F(x) = \sum_{k: x_k \leq x} p_k$$

例

X 取 $-2, 0, 1, 4$, 概率分别为 $0.4, 0.1, 0.3, 0.2$ 。画 $F(x)$ 。

F 在 -2 跳到 0.4 , 在 0 跳到 0.5 , 在 1 跳到 0.8 , 在 4 跳到 1 。四级台阶, 四次跳跃的高度之和为 1 。

例

反复抛硬币直到出现正面, $X =$ 抛的次数。画 $F(x)$ 。

X 服从 $p = 1/2$ 的几何分布, $P(X = k) = (1/2)^k$ 。

$$F(k) = 1 - (1/2)^k$$

无穷多级台阶, 越来越矮; F 始终小于 1 , 但趋于 1 。

预告：连续型的分布函数是什么形状

例

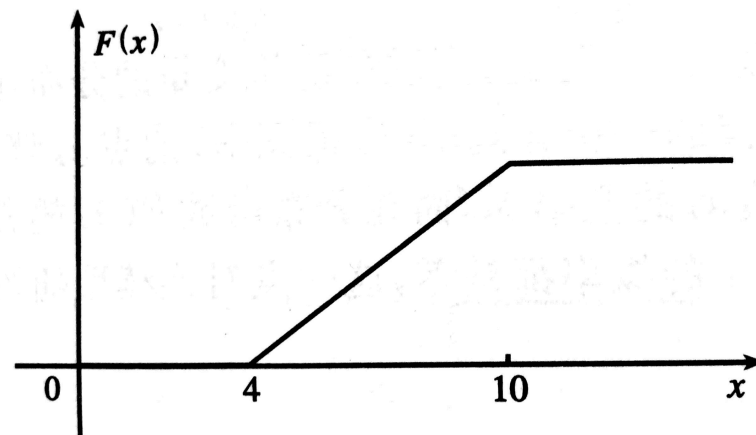
在区间 $[4, 10]$ 上任意抛一个质点， X 为质点与原点的距离。若落在任一子区间的概率与该区间长度成正比，求 X 的分布函数。

解

$$F(x) = \begin{cases} 0, & x < 4 \\ \frac{x-4}{6}, & 4 \leq x < 10 \\ 1, & x \geq 10 \end{cases}$$

没有台阶了，是一条连续上升的折线。任何一点的概率都是 0，但

$$P(5 < X \leq 7) = \frac{7-4}{6} - \frac{5-4}{6} = \frac{1}{3}。$$



连续型： F 处处连续

本讲总结

① 回到开场那个问题

连赢 10 年的概率确实只有 $1/1024$ 。但在 5000 位经理里，**至少出现一个这样的人的概率是 99.2%**，赢 8 年及以上的平均有 273 人。

只看一个事件的概率会得出错误结论；**掌握「跑赢年数」的整张分布，才能把全部情形算进去。**这是从「事件」转向「随机变量」的原因。

② 核心结论

- **随机变量是函数**： $X : \Omega \rightarrow \mathbb{R}$ ；事件 $\{X \in C\} = \{\omega : X(\omega) \in C\}$
- 离散型用**概率函数** $f(x) = P(X = x)$ ，两条性质 $p_k \geq 0$ 、 $\sum p_k = 1$
- 四个分布：伯努利（单次） $\rightarrow$ 几何（试到成功） $\rightarrow$ 二项（ n 次中成功几次） $\rightarrow$ 泊松（ n 大 p 小的极限）
- **分布函数** $F(x) = P(X \leq x)$ 对离散与连续通用； $P(X = x) = F(x) - F(x^-)$

本讲总结 (续)

③ 易错提醒

- **随机变量是函数，不是数。** 随机的是它的输入 ω ，规则本身是确定的
- **几何与二项不要混淆：** 几何固定「目标」数次数，二项固定「次数」数成功
- 泊松近似的条件是 n **大**、 p **小**、 np **适中**，三者缺一不可
- F 是**右**连续的，这来自定义里的 $\leq$ ； $P(X < x) = F(x^-)$ 而不是 $F(x)$

④ 下一讲

第 32 页的 $F(x) = \frac{x-4}{6}$ 是一条**连续上升的折线**，没有台阶，所以每一点的 $P(X = x)$ 都是 0，概率函数无法描述这个分布。

但那条折线的**斜率**还在：斜率越大的地方，概率越密集。**下一讲：对 F 求导得到「密度函数」，用它代替概率函数。**