

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 $P(A | B) = \frac{P(AB)}{P(B)}$ 是样本空间的收缩；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 重贝努里试验可算；但假设错了，风险能低估十几个数量级。

★ 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

场景 一位基金经理，**连续 10 年跑赢基准指数**。基金公司把这写进宣传材料，管理规模翻了几番。

冲突 有人算了一笔账：如果他毫无本事，每年跑赢的概率就是抛硬币的 $1/2$ ，那么连赢 10 年的概率是 $(1/2)^{10} = 1/1024 \approx 0.1\%$ 。**千分之一的事发生了，这总不能是运气吧？**
这个算式**一点没错**。但结论错了。

悬念 **因为全市场有几千位基金经理。在几千次机会里，出现一个「千分之一」几乎是必然的。**
要看清这件事，光算「某一个事件的概率」不够——我们需要一次性掌握「跑赢年数」这个数量的整张概率表。

来源与说明

本页两个概率为教学化演算： $(1/2)^{10} = 1/1024$ ；5000 位经理中至少一人 10 战全胜的概率 $1 - (1 - 1/1024)^{5000} \approx 99.2\%$ 。数字用于说明「多重比较」这一统计现象，不指涉任何具体机构或个人。严格的处理属第 12 讲假设检验的内容。

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**，知道各自在回答什么问题
- 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

基金经理不关心导致股价波动的每一条新闻（具体事件），只关心最终的**涨跌幅**——一个数值。

人工智能

垃圾邮件过滤器最终要输出一个**是垃圾邮件的得分**——也是一个数值。

本讲的核心任务

造一座桥，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字（如 2）。**这座桥就是随机变量。**

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。

★ 三个字要咬清楚：它是个函数

X **不是**一个数，是一条规则：给定 $\omega \in \Omega$ ，返回一个实数 $X(\omega)$ 。「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 $X(\text{正正反}) = 2$ ， $X(\text{反反反}) = 0$ 。

现在事件可以用它来写

$\{X = 2\}$ 就是事件 $\{\text{正正反}, \text{正反正}, \text{反正正}\}$ 的简写。**随机变量把事件的书写从「列举结果」变成了「写不等式」。**

分布：一次性描述全部取值

定义

随机变量 X 的**分布**，指对所有实数集 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

回到开场

「跑赢年数」 X 只能取 $0, 1, \dots, 10$ ，是离散型。我们要的正是它的**整张分布表**，而不是某一个 $P(X = 10)$ 。

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 及其所有可能取值 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 (就是三条公理的翻译)

记 $p_k = f(x_k)$, 则

- $p_k \geq 0$, 且对不可能的取值 $f(x) = 0$

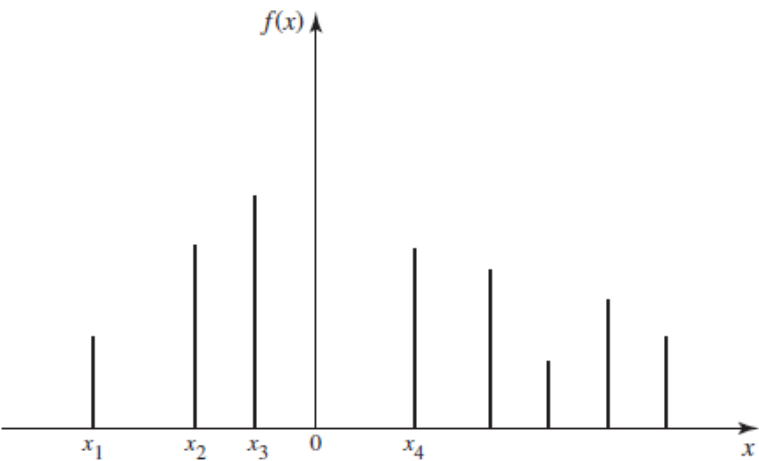
← 非负性

- $\sum_k p_k = 1$

← 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 $\{p_k\}$ 都定义了一个合法的分布。这是后面「确定常数 c 」那类题的依据。

三种等价写法，与第一个例子



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——三种写法内容完全相同，按场合选。

例：离散型均匀分布

彩票选一个三位数，开奖从 0-999 随机抽。中奖号码 X 的概率函数：

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

若 X 的 n 个取值概率都是 $1/n$ ，称 X 服从**离散型均匀分布**。

用归一化条件定常数

例

某人买彩票中奖率 p , 没中就继续买下一期, 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:

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

这个式子马上会有名字

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, 这时归一化就是一个级数收敛的条件。

分布之一：伯努利分布（0-1 分布）

定义

X 服从参数为 p 的**伯努利分布**，若 X 只取 0, 1, 且

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

它简单到看起来没用，但它是一切的原子

接下来的三个分布，全部由伯努利试验搭成：

- **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
- **信用评分模型做的全部事情，就是估计每个客户的 $p = P(X = 1)$ ——这个 p 决定贷款利率与风险拨备**

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 分类模型的输出就是一个伯努利随机变量的参数 p ——它根据输入特征尽可能准地估出这个 p 。

上一讲的贝叶斯分类器、本讲的伯努利分布，说的是同一件事的两面：**前者算 p ，后者描述 p 定下来之后的随机性。**

分布之二：几何分布

从一个试验说起

机器有 10% 概率生产次品，各件独立。重复生产**直到出现次品为止**，令 X 为结束时生产的产品数量。 X 服从什么分布？

几何分布

X 服从参数为 p 的**几何分布**，若

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

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

和二项分布的分工

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 。

两者哪个是随机的正好对调。

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为 p ，则所需月数 X 服从几何分布
- **理解摩擦性失业**：算 $P(X > 12) = (1 - p)^{12}$ ，即一年还没找到工作的比例
- **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

- 新推荐算法吸引用户点击的概率为 p
- 用户在第 X 次看到推荐时才首次点击， X 服从几何分布
- X 普遍小 $\rightarrow$ 算法有吸引力； X 普遍大 $\rightarrow$ 需要优化

注意 $P(X > k) = (1 - p)^k$ 这个式子有多好用——「**至今尚未成功**」的概率，正面算只需一步。这是上一讲「先看对立事件」习惯的又一次兑现。

几何分布的无后效性

无记忆性

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： $\frac{P(X > s + t)}{P(X > t)} = \frac{(1 - p)^{s+t}}{(1 - p)^t} = (1 - p)^s = P(X > s)$ 。完整证明见教师笔记。

这句话的分量

「已经等了 t 次还没成功」这个信息，对将来毫无价值。 再等 s 次仍不成功的概率，和从头开始等 s 次一模一样。

对照上一讲：这正是「独立」在时间维度上的表现——**过去不携带关于未来的信息。**

用在求职模型上，这个性质其实**反直觉且多半不成立**：现实中失业越久越难找到工作（技能折旧、雇主筛选）。**模型的假设失效在哪里，比模型本身更值得想。**

分布之三：二项分布

从伯努利搭起来

机器次品率 $p = 10\%$ ，检查 n 个产品。令 $X_i = 1$ 表示第 i 个是次品，则 $X_1, \dots, X_n$ **独立同分布**，都服从参数 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 $X \sim B(n, p)$

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

这正是上一讲的贝努里定理 $p_n(k)$ ——**现在它有了名字和身份：一个随机变量的分布。**

$n = 1$ 时它就是伯努利分布

所以伯努利是二项的特例。「 n 个独立同分布伯努利变量之和」是二项分布最本质的刻画——第 8 讲中心极限定理会从这里出发。

交互：离散分布实验室

怎么用

切换四个分布，拖参数看形状。二项分布下还可勾选「叠加泊松逼近」。

当堂要做的三个实验

1. 二项：把 p 拖到 0.5，看形状最对称；拖向两端看偏斜
2. 几何：注意概率**单调递减**——最可能的永远是「第一次就成」
3. 泊松：注意期望与方差**相等**，别的分布没有这性质



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

回到开场：用二项分布把账算清

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 战全胜的期望人数： $5000/1024 \approx 5$ 人
- 至少有一人 10 战全胜的概率： $1 - (1 - \frac{1}{1024})^{5000} \approx 99.2\%$
- 赢 8 年以上的期望人数： $5000 \times 5.5\% \approx 273$ 人

「连赢 10 年」在几千人的行业里几乎必然会落到某个人头上，并不能据此推出他有能力——除非把「一共有多少次机会」也算进去。这叫**多重比较**，第 12 讲细讲。

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 $X \sim B(50, 0.04)$
- 「超预期」损失的概率

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

这是设定风险拨备与对冲规模的依据

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
- 成功识别数 $X \sim B(20, 0.95)$
- 算 $P(X < 18) = \sum_{k=0}^{17} C_{20}^k (0.95)^k (0.05)^{20-k}$
- 若这个概率极低而实测确实 $X = 17$ ，就有理由**质疑它宣称的性能**

右边这个推理——「**若假设为真，观测到的结果极不可能，于是怀疑假设**」——就是假设检验的全部逻辑。第 12 讲会给它一套严格的语言。但请注意它与上一页的教训并不矛盾：**前提是你只测了这一个模型。**

顺带一条：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」这个名字正是由此而来。

分布之四：为什么还需要泊松分布

前面两个分布都有上限

二项分布的取值最多到 n 。但有些量**事先不知道上限**：

- 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，但平均次数 np 是个正常的数。

泊松分布是怎么被推出来的

店主的做法

顾客平均每小时 4.5 名。把一小时切成 3600 秒，假设每秒最多来一位顾客，每秒来人的概率 $p = 4.5/3600 = 0.00125$ 。于是一小时的顾客数 $X \sim B(3600, 0.00125)$ 。

但这个做法有两个毛病

C_{3600}^x 根本算不动；而且它无法描述顾客数超过 3600 的情况。

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

X 服从参数为 λ 的**泊松分布**，若

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

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 $B(n, p_n)$ 趋近于参数 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。完整证明见教师笔记。

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。回到上一页的交互器，勾上「叠加泊松逼近」，把 n 拖大、 p 拖小，两条线会重合。

泊松分布例题（一）

例：近似的精度

大样本中某事件发生率 0.01，随机抽 200 人，求至少 4 人的概率。

精确（二项 $B(200, 0.01)$ ）：

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

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 λ 处印错，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 $e^{-0.5} \approx 0.607$ ——平均每页半个错，仍有六成的页面是干净的。

泊松分布例题（二）：两个分布套在一起

例

一本书 n 页，每页平均 λ 个错误。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 座的飞机卖 200 张票，平均 1% 的乘客不出现。所有到场乘客都有座，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 张就有四成概率坐不下。**

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**：求最小的 k 使 $P(X \leq k) \geq 99.9\%$ ——这决定要备多少理赔金
- **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 $P(X \geq 3000 \mid \lambda = 500)$ 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」**。这就是所有异常检测系统的骨架。

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。「等车时间恰好是 3.000000... 分钟」的概率是 0——每一点的概率都是 0， $f(x) = P(X = x)$ 全等于 0，什么信息都没带。

需要一个对两类都管用的描述方式。

分布函数：对两类都管用的描述

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**。

为什么它能通吃

它问的不是「恰好等于 x 的概率」，而是「**不超过 x 的概率**」。后者对连续型也是个有意义的正数。

例

X 服从 0-1 分布， $P(X = 0) = p$ 、 $P(X = 1) = 1 - p$ 。其分布函数为

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

——一个**阶梯函数**，在每个取值点向上跳一格，跳的高度就是该点的概率。

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

前三条由概率的单调性与规范性直接得到 (第 1 讲性质 3、性质 5)。第四条: 不减函数的间断点必为跳跃间断, 而互不相交的跳跃区间至多可列个; 右连续来自 $F(x) = P(X \leq x)$ 中取的是闭区间。完整证明见教师笔记。

★ 「右连续」这一条不是技术细节

它来自定义里那个 $\leq$ 。若把定义改成 $F(x) = P(X < x)$, 得到的就是左连续函数。 **约定不同, 公式全变**——所以下一页那组公式必须和这个约定配套记。

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**)

最后一条把两类随机变量统一了

离散型： F 在取值点跳跃，跳高就是 $P(X = x)$ 。

连续型： F 处处连续，于是 $P(X = x) = 0$ —— 每一点的概率都是 0，但落在区间里的概率不是 0。

这解释了上一页那个困惑：连续型不是「没有概率」，是**概率不集中在点上，而是摊在区间上**。

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 $0.4, 0.1, 0.3, 0.2$ 。画 $F(x)$ 。

阶梯：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，高度和为 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，永远够不到 1 但趋于 1。

预告：连续型的分布函数长什么样

例

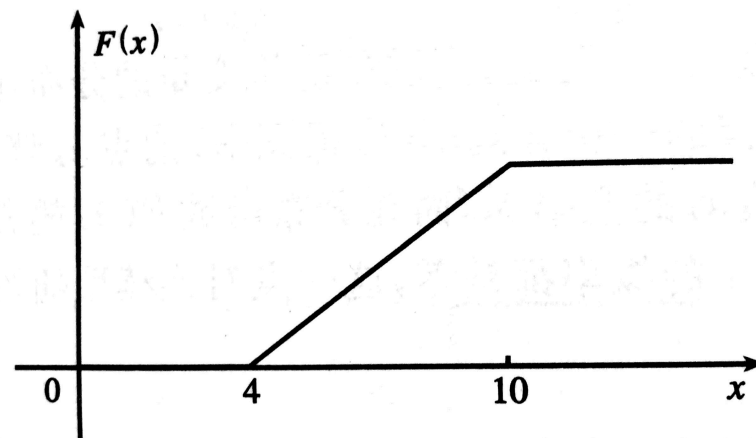
在区间 $[4, 10]$ 上任意抛一个质点， X 为质点与原点的距离。若落在任一子区间的概率与该区间长度成正比，求 X 的分布函数。

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——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 年以上的期望人数是 273 人。

只盯住一个事件的概率会看错；**掌握「跑赢年数」的整张分布，才能把账算全。**这就是从「事件」升级到「随机变量」的意义。

② 核心结论

- **随机变量是函数**： $X : \Omega \rightarrow \mathbb{R}$ ；事件 $\{X \in C\}$ 是它的原像
- 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 次数成功） $\rightarrow$ 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
- **几何与二项别混：** 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

本讲最后一页的 $F(x) = \frac{x-4}{6}$ 是一条**光滑上升的曲线**，没有台阶——所以 $P(X = x) = 0$ ，概率函数彻底失效了。

但那条曲线的**斜率**还在：斜率大的地方，概率就密集。 **下一讲：用「密度」代替「概率」，把 F 的导数请出来。**