

# 随机变量的数字特征

## 第 7 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

从「一个变量」跨到「一群变量」

联合  $f(x, y)$  看全貌、边缘  $f_1 = \sum_y f$  看单个、条件  $g_2(y | x) = f / f_1$  看联动；独立  $\iff f = f_1 f_2$ 。

也会求和的分布了

离散沿斜线求和、连续用卷积；二项、泊松、正态都可加，正态是**方差**相加——但要求独立。

### ★ 留下的问题

不独立时方差怎么算？「一起动」的程度，能不能压缩成一个数？

而且——整张联合分布表信息量很大，但汇报时没人要看一张表，大家要的是「平均多少」「波动多大」这样的一**两个数**。

## 为什么需要这一讲

**场景** 一个赌局：反复抛硬币，直到第一次出现正面。若是在第  $k$  次出现的，庄家付你  $2^k$  元。你愿意花多少钱买这个参赌资格？

**冲突** 算一下期望：第  $k$  次才出正面的概率是  $\frac{1}{2^k}$ ，收益  $2^k$  元，于是

$$E = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot 2^k = 1 + 1 + 1 + \cdots = \infty$$

**期望是无穷大。**按「期望值最大化」这条原则，你应该愿意付任何价格——倾家荡产也该买。可实际上，没有人愿意为它出超过二三十块。

**悬念五年。**这不是外行的困惑。这道题 1713 年由尼古拉·伯努利提出，困住了当时整个欧洲的数学家二十年。「期望」这一个数，既不足以描述一个分布，也不足以指导决策。还差什么？

来源

圣彼得堡悖论。Nicolas Bernoulli 于 1713 年 9 月 9 日致 Pierre Rémond de Montmort 的信中提出；Daniel Bernoulli 于 1738 年在圣彼得堡科学院的 *Commentarii* 上发表 *Specimen Theoriae Novae de Mensura Sortis*，以「效用随财富递减（对数效用）」给出解答，这也是**期望效用理论**的起点。[plato.stanford.edu/entries/paradox-stpetersburg](https://plato.stanford.edu/entries/paradox-stpetersburg) | 访问日期 2026-08-02。

## 本讲学习目标

- ① 会算**期望**——分布的「价值中心」在哪，回答「平均能赚多少」
- ② 会算**方差**——分布的「离散程度」如何，回答「风险有多大」
- ③ 会算**协方差与相关系数**——两个变量「联动」多强，回答上一讲留下的问题

一个复杂的分布如何变成可比较的决策指标？投资者 A 的收益服从  $f_A$ 、B 服从  $f_B$ ，谁更好？AI 模型 A 的误差服从  $g_A$ 、B 服从  $g_B$ ，哪个更优？**本讲用几个数概括整个分布。**

# 为什么要「数字特征」

分布包含全部信息，但常常求不出、也没必要

- **平均程度**：乘坐这个航班平均延误多久
- **偏离平均的程度**：标准长 1 米的产品，实际偏离多少
- **两个量的相关性**：睡眠质量与学习成绩有没有关系

先从「平均」开始

某航班过去四天分别延误 0, 10, 20, 30 分钟，平均延误多久？——15 分钟。

换个问法：过去四天里**两天**延误 10 分钟、一天 20 分钟、一天没延误呢？

$$\frac{10 + 10 + 20 + 0}{4} = 10 \times \frac{1}{2} + 20 \times \frac{1}{4} + 0 \times \frac{1}{4} = 10$$

**右边这个写法就是期望**——把「平均」重写成「取值 × 概率再求和」。

# 期望的定义

## 定义

**离散型**：若级数  $\sum_k x_k f(x_k)$  **绝对收敛**，则

$$EX = \sum_{k=1}^{\infty} x_k f(x_k)$$

**连续型**：若积分  $\int_{-\infty}^{\infty} x f(x) dx$  **绝对收敛**，则

$$EX = \int_{-\infty}^{\infty} x f(x) dx$$

★ 「绝对收敛」这四个字不是装饰

开场那个赌局的期望就是**不收敛**的——所以严格地说，它的期望不存在，而不是「等于无穷大」。

期望不存在的分布是真实存在的（柯西分布就是一例）。 **见到期望先问一句：它收敛吗？**

## 两个最简单的例子

### 例（离散）

$X$  服从 0-1 分布,  $P(X = 1) = p$ 。

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

**伯努利分布的期望就是  $p$  本身**——「平均成功次数」等于「成功概率」。这条后面反复要用。

### 例（连续）

$X$  有密度  $f(x) = 2x$  ( $0 < x < 1$ ) 。

$$E(X) = \int_0^1 x \cdot 2x \, dx = \frac{2}{3}$$

直觉核对：密度在  $x$  大处更高，所以期望应该**大于**区间中点 0.5。  $\frac{2}{3}$  符合。**算完顺手做一次这种核对。**

## 二项分布的期望（硬算一次）

例

盒中红球比例  $p$ ，**有放回**取  $n$  个，其中  $X$  个红球。求  $EX$ 。

解：把  $k$  吸收进组合数

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

关键一步：  $k C_n^k = k \cdot \frac{n!}{k!(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} = n C_{n-1}^{k-1}$ ，于是

$$EX = np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \cdot 1 = np$$

末尾那个和式是  $B(n-1, p)$  的全部概率，等于 1。

能算，但**不好玩**。等讲完期望的性质，这题会缩成一行——见 slide 20。

## 几何分布的期望：一个错位相减的技巧

$X \sim \text{几何分布}(p)$ , 求  $EX$

$$EX = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

两边乘  $(1-p)$  并把指标平移:

$$(1-p)EX = \sum_{k=1}^{\infty} k(1-p)^k p = \sum_{k=1}^{\infty} (k-1)(1-p)^{k-1}p$$

上下两式**相减**,  $k - (k-1) = 1$ :

$$pEX = \sum_{k=1}^{\infty} (1-p)^{k-1}p = 1 \implies EX = \frac{1}{p}$$

**结果很好记:** 成功率  $p$ , 平均要试  $1/p$  次。  $p = 0.1$  就平均试 10 次。

错位相减这一招在第 10 讲还会再用一次。

## 另外三个分布的期望

### 泊松 ( $\lambda$ )

$$EX = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} = \lambda$$

$x = 0$  那项为 0, 从 1 开始; 提出一个  $\lambda$  后剩下的又是全部概率。

### 指数 ( $\lambda$ )

$$EX = \int_0^{\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}$$

分部积分。**速率  $\lambda$  的倒数就是平均等待时间**——每小时来 4 辆车, 平均等 15 分钟。

### 正态 ( $\mu, \sigma^2$ )

$$EX = \mu$$

由密度关于  $\mu$  **对称**直接看出, 不必算积分。

# 小结：五个分布的期望

| 分布                    | $EX$        | 怎么记                    |
|-----------------------|-------------|------------------------|
| 伯努利 ( $p$ )           | $p$         | 平均成功次数 = 成功概率          |
| 二项 $B(n, p)$          | $np$        | 做 $n$ 次, 每次期望 $p$      |
| 几何 ( $p$ )            | $1/p$       | 成功率的倒数                 |
| 泊松 ( $\lambda$ )      | $\lambda$   | $\lambda$ 本来就定义成「平均次数」 |
| 指数 ( $\lambda$ )      | $1/\lambda$ | 速率的倒数                  |
| 正态 $N(\mu, \sigma^2)$ | $\mu$       | 对称中心                   |

这张表要背下来，第 8 讲起每一讲都在用。

# 随机变量函数的期望：不必先求分布

## 定理

令  $Y = r(X)$ , 则

$$E[r(X)] = \sum_{\text{所有 } x} r(x)f(x) \quad \text{或} \quad E[r(X)] = \int_{-\infty}^{\infty} r(x)f(x) \mathrm{d}x$$

## ★ 这条省掉了一整步

第 5 讲求  $Y = r(X)$  的分布要「先求  $G$  再求导」, 很麻烦。但如果只要期望, 用  $X$  的密度直接积就行, 不必求  $Y$  的分布。

## 例

$f(x) = 2x$  ( $0 < x < 1$ ), 求  $E(X^{1/2})$ 。

$$\int_0^1 x^{1/2} \cdot 2x \mathrm{d}x = \frac{4}{5}$$

## 例

$f(x) = 3x^2$  ( $0 < x < 1$ ), 求  $E(1/X)$ 。

$$\int_0^1 \frac{1}{x} \cdot 3x^2 \mathrm{d}x = \frac{3}{2}$$

## 二元函数的期望

例

$f(x, y) = 1$  ( $0 \leq x, y \leq 1$ ) , 求  $E(X^2 + Y^2)$ 。

$$\int_0^1 \int_0^1 (x^2 + y^2) dx dy = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

例

$f(x, y) = 12y^2$  ( $0 \leq y \leq x \leq 1$ ) , 求  $E(XY)$ 。

$$\int_0^1 \int_0^x xy \cdot 12y^2 dy dx = \int_0^1 3x^5 dx = \frac{1}{2}$$

积分限又是上一讲那套:  $y$  从 0 到  $x$ 。

一般地  $E[r(X, Y)] = \iint r(x, y) f(x, y) dx dy$  —— 同样**不必先求**  $r(X, Y)$  **的分布**。

## 期望的三条性质（本讲最有用的一页）

- **线性（对常数）**：  $E(aX + c) = aE(X) + c$ ；特别地  $E(c) = c$ 、 $E(aX) = aE(X)$
- **可加（★ 不需要独立）**：  $E(X + Y) = E(X) + E(Y)$
- **可乘（需要独立）**：  $X, Y$  独立  $\Rightarrow E(XY) = E(X)E(Y)$

### 证明思路

可加性：  $E(X + Y) = \iint (x + y)f(x, y) = \iint xf + \iint yf$ ，前者对  $y$  积得边缘  $f_1$ ，即  $E(X)$ ，后者同理——**全程没用到独立**。  
可乘性：独立时  $f = f_1f_2$ ，二重积分拆成两个一重积分之积。完整证明见教师笔记。

### ★ 第二条与第三条的差别是本讲最重要的区分

$E(X + Y) = E(X) + E(Y)$  **永远成立**，哪怕  $X, Y$  强烈相关。

$E(XY) = E(X)E(Y)$  **只在独立时成立**——不独立时两边的差，正是本讲后半的主角**协方差**。

## 用性质重做两道题

### 二项分布的期望，第二种解法

$X \sim B(n, p)$  可写成  $n$  个独立伯努利之和  $X = \sum_{i=1}^n X_i$ ，每个  $E(X_i) = p$ 。由**可加性**

$$EX = \sum_{i=1}^n E(X_i) = np$$

**一行。**对比 slide 8 那半页组合恒等式。

### 配对问题

$n$  封信随机装进  $n$  个信封， $X$  为装对的封数。求  $EX$ 。

令  $X_i = 1$  表示第  $i$  封装对，则  $P(X_i = 1) = \frac{1}{n}$ ， $E(X_i) = \frac{1}{n}$ ：

$$EX = \sum_{i=1}^n E(X_i) = n \cdot \frac{1}{n} = 1$$

**无论  $n$  多大，平均都恰好装对 1 封。** 而这些  $X_i$  并不独立——可加性不需要独立，这里就用上了。

## 用独立性拆开期望

例

$X_1, X_2, X_3$  独立,  $E(X_i) = 0$ 、 $E(X_i^2) = 1$ 。求  $E[X_1^2(X_2 - 4X_3)^2]$ 。

$X_1^2$  与  $(X_2 - 4X_3)^2$  独立, 故可乘:

$$= E(X_1^2) \cdot E[(X_2 - 4X_3)^2]$$

$$= 1 \cdot E(X_2^2 - 8X_2X_3 + 16X_3^2) = 1 - 0 + 16 = 17$$

中间项  $E(X_2X_3) = E(X_2)E(X_3) = 0$ 。

例：两级滤水

第一次滤去  $X_1$  比例杂质, 第二次滤去  $X_2$ 。  $X_1, X_2$  独立同分布,  $f(x) = 4x^3$  ( $0 < x < 1$ )。求两次后剩余比例  $Y$  的期望。

$Y = (1 - X_1)(1 - X_2)$ , 由独立性

$$E(Y) = [1 - E(X_1)]^2 = \left(1 - \frac{4}{5}\right)^2 = 0.04$$

$$(E(X_1) = \int_0^1 x \cdot 4x^3 = \frac{4}{5}。)$$

**两级过滤把杂质降到 4%**——比单级的 20% 好得多。

## 应用：投资组合的预期收益

### 场景

组合总收益率  $R_p = \sum_i w_i R_i$  ( $w_i$  是各资产权重)。

### 由期望的线性性

$$E[R_p] = \sum_i w_i E[R_i]$$

**组合的预期收益就是各资产预期收益的加权平均** ——不需要知道资产之间怎么联动。

### 记住这句话，本讲后半要打脸

**收益可以这样简单相加，风险不行。** 这是现代投资组合理论的出发点，但只解决了一半问题：「我能期望赚多少」。另一半——「我要冒多大风险」——要等到方差与协方差。

## 条件期望

### 定义

给定  $X = x$  时  $Y$  的**条件期望**，就是条件分布的期望：

$$E[Y \mid x] = \sum_{\text{所有 } y} y g_2(y \mid x) \quad \text{或} \quad E[Y \mid x] = \int_{-\infty}^{\infty} y g_2(y \mid x) \mathrm{d}y$$

### 注意它是 $x$ 的函数

$E[Y \mid x]$  不是一个数，是**一条曲线**——每给一个  $x$ ，就有一个对应的平均值。上一讲的「猜你喜欢」要的就是它：给定用户对甲片打 5 分，他对乙片的平均打分。

# 回归的本质：拟合条件期望

## 核心问题

已知房屋面积  $X$ ，如何**最优地**预测价格  $Y$ ？

## 理论上的最佳答案

在所有可能的预测值中，使**均方误差**  $E[(Y - \text{预测值})^2 | X = x]$  最小的那个，**恰好就是条件期望**  $E[Y | X = x]$ 。

## 证明思路

设预测值为  $c$ ，记  $m = E[Y | x]$ 。则  $E[(Y - c)^2 | x] = E[(Y - m + m - c)^2 | x] = E[(Y - m)^2 | x] + (m - c)^2$ （交叉项  $2(m - c)E[Y - m | x] = 0$ ）。右边第一项与  $c$  无关，第二项在  $c = m$  时取最小值 0。完整证明见教师笔记。

## 于是所有回归模型的目标都是同一个

$E[Y | X = x]$  是  $x$  的函数，称为**回归函数**。**线性回归**假设它是直线  $\beta_0 + \beta_1 x$ ，**多项式回归**假设它是多项式，**神经网络**用更复杂的非线性函数去逼近它。

**分歧只在「用什么函数族去逼近」，目标完全相同。**

## 接下来

期望回答了「中心在哪」。但——

学生甲成绩稳定在 90 分，学生乙在 85–95 之间波动。**两人的期望都是 90**，可谁的成绩更让人放心？

**还需要一个数来描述「散得有多开」。**

# 方差

## 定义

称  $X - EX$  为**离差**。离差**平方**的期望称为**方差**：

$$D(X) = E[(X - EX)^2]$$

离散型  $\sum_x (x - EX)^2 f(x)$ ，连续型  $\int (x - EX)^2 f(x) dx$ 。其平方根  $\sqrt{D(X)}$  称为**标准差**。

## 为什么要平方

直接取  $E(X - EX)$  恒等于 0（正负抵消），没有信息。平方消掉符号；**而且平方比取绝对值好——它可导，能求最优**，上一页那个「均方误差最小」正是靠这一点。

代价是**量纲变了**（元  $\rightarrow$  元<sup>2</sup>），所以又定义标准差把量纲开回来。

## 例：两只股票，同样的期望

例

股价  $A \sim U[25, 35]$ ,  $B \sim U[15, 45]$ 。求各自的期望与方差。

解

$E(A) = E(B) = 30$ ——期望完全相同。

$$D(A) = \int_{25}^{35} (a - 30)^2 \cdot \frac{1}{10} da = \frac{25}{3} \approx 8.33$$

$$D(B) = \int_{15}^{45} (b - 30)^2 \cdot \frac{1}{30} db = 75$$

标准差分别是 2.89 与 8.66，相差三倍。

一般地， $U[a, b]$  的方差是  $\frac{(b-a)^2}{12}$ ——区间宽度翻三倍，方差翻九倍。期望看不出的差别，方差一眼看出。

## 方差的三条性质

- **线性变换**:  $D(aX + c) = a^2 D(X)$ ; 特别地  $D(c) = 0$ 、 $D(aX) = a^2 D(X)$
- **可加 (★ 需要独立)**:  $X, Y$  独立  $\Rightarrow D(X + Y) = D(X) + D(Y)$
- **化简公式 (最常用)**:  $D(X) = E(X^2) - (EX)^2$

### 证明思路

化简公式:  $E[(X - EX)^2] = E[X^2 - 2X EX + (EX)^2] = E(X^2) - 2(EX)^2 + (EX)^2$ 。

线性变换:  $D(aX + c) = E[(aX + c - aEX - c)^2] = a^2 E[(X - EX)^2]$ ——**常数  $c$  被消掉了**, 平移不改变离散程度。完整证明见教师笔记。

### ★ 与期望对照着记

$E(X + Y) = E(X) + E(Y)$  **不需要独立**;  $D(X + Y) = D(X) + D(Y)$  **需要独立**。

而且  $D$  里的系数是  $a^2$  **不是**  $a$ ——所以标准差才是  $|a|$  倍。

## 几何分布与二项分布的方差

几何分布：错位相减，再来一次

$$(1-p)EX^2 = \sum_{k \geq 1} (k-1)^2 (1-p)^{k-1} p$$

相减,  $k^2 - (k-1)^2 = 2k - 1$ :

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

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

二项分布：两种解法

硬算用  $E[X(X-1)]$  的技巧:

$$E(X^2) = E[X(X-1)] + EX = n(n-1)p^2 + np$$

$$D(X) = n^2 p^2 - np^2 + np - n^2 p^2 = np(1-p)$$

或者一行:  $X = \sum X_i$  独立, 每个  $D(X_i) = p(1-p)$ , 由可加性

$$D(X) = np(1-p)$$

注意右边这次**用到了独立**——而 slide 21 求期望时没用到。 **同一个分解，方差比期望多一个前提。**

## 另外三个分布的方差

泊松 ( $\lambda$ )

$$E[X(X-1)] = \lambda^2 \sum_{x \geq 2} \frac{e^{-\lambda} \lambda^{x-2}}{(x-2)!} = \lambda^2$$

$$D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

**期望与方差都等于  $\lambda$** ——第 4 讲说过的那个「最好认的特征」，现在证出来了。

指数 ( $\lambda$ )

$$D(X) = \frac{1}{\lambda^2}$$

标准差 =  $1/\lambda$  = 期望。**指数分布的标准差等于它的期望**，也是个好认的特征。

正态 ( $\mu, \sigma^2$ )

$$D(X) = \sigma^2$$

记号里那个  $\sigma^2$  **本来就是方差**——第 5 讲写记号时埋的，现在兑现。

## 应用：机器学习的损失函数

### AI 训练的终极目标

模型训练就是最小化某个「损失」的**期望**。回归任务最常用的是均方误差：

$$\min_{\text{模型参数}} E[(Y_{\text{真实}} - Y_{\text{预测}})^2]$$

### 把它和前面两页连起来

- 这个式子的形状**就是方差的形状**：离差平方的期望
- 而 slide 27 已经证过：使它最小的预测值是**条件期望**  $E[Y | X]$
- 所以**整个机器学习的训练过程，是在用数据逼近一条条件期望曲线**

换个损失函数就换了目标：用绝对误差  $E|Y - \hat{Y}|$ ，最优预测变成**条件中位数**而非均值。 **损失函数的选择，等于在选你要预测分布的哪个特征。**

# 小结：期望与方差对照表

| 分布                    | $EX$        | $D(X)$         |
|-----------------------|-------------|----------------|
| 伯努利 $(p)$             | $p$         | $p(1 - p)$     |
| 二项 $B(n, p)$          | $np$        | $np(1 - p)$    |
| 几何 $(p)$              | $1/p$       | $(1 - p)/p^2$  |
| 泊松 $(\lambda)$        | $\lambda$   | $\lambda$      |
| 均匀 $U[a, b]$          | $(a + b)/2$ | $(b - a)^2/12$ |
| 指数 $(\lambda)$        | $1/\lambda$ | $1/\lambda^2$  |
| 正态 $N(\mu, \sigma^2)$ | $\mu$       | $\sigma^2$     |

这张表期末一定要能默写。第 8 讲起全部推断都从它出发。

# 协方差：把「一起动」压缩成一个数

## 动机

两个不独立的随机变量，依赖到什么程度？ 数学成绩与语文成绩不独立，但也不相等——怎么描述？

## 定义

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

展开后得到常用的计算式：

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

## ★ 它正是 slide 19 那个「差」

当时说： $E(XY) = E(X)E(Y)$  只在独立时成立。 **不独立时两边的差，就是协方差。**

符号的含义： $X$  比平均大的时候  $Y$  也偏大  $\rightarrow$  两个离差同号  $\rightarrow$  乘积为正  $\rightarrow \text{Cov} > 0$ 。

# 协方差：一道例题与六条性质

## 例

$f(x, y) = 2xy + 0.5$  ( $0 \leq x, y \leq 1$ )。求  $\text{Cov}(X, Y)$ 。

$$E(X) = E(Y) = \int_0^1 \int_0^1 x(2xy + 0.5) dy dx = \frac{7}{12}$$

$$E(XY) = \int_0^1 \int_0^1 xy(2xy + 0.5) dy dx = \frac{25}{72}$$

$$\text{Cov} = \frac{25}{72} - \left(\frac{7}{12}\right)^2 = \frac{1}{144}$$

**正的，但很小**——两者略微正相关。

## 六条性质

- $\text{Cov}(X, X) = D(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $E(XY) = E(X)E(Y) + \text{Cov}(X, Y)$
- $D(X + Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$
- $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$
- $\text{Cov}(X, aY + b) = a \text{Cov}(X, Y)$

**第一条说方差是协方差的特例**（自己和自己的协方差）；**第四条是上一讲那个问题的答案**——不独立时方差不再简单相加，差的正好是  $2\text{Cov}$ 。独立时  $\text{Cov} = 0$ ，退回旧公式。

## 应用：风险分散这顿「免费午餐」

### 双资产组合的风险

$$D(R_p) = w_A^2 D(R_A) + w_B^2 D(R_B) + 2w_A w_B \text{Cov}(R_A, R_B)$$

### 对比 slide 24

收益是**线性**相加的：  $E[R_p] = w_A E[R_A] + w_B E[R_B]$ ，与相关性无关。

风险**不是**——最后那一项带着协方差。

**若  $\text{Cov} < 0$ ，总风险小于各部分风险之和**：同样的期望收益，更低的波动。这就是「不要把鸡蛋放在一个篮子里」的数学原理，也是金融学里常说的唯一的免费午餐。

### AI 里的同一件事：PCA

高维特征之间的关系用**协方差矩阵**  $\Sigma$  描述。主成分分析通过分析  $\Sigma$ ，找到一组**协方差为 0** 的新坐标轴，实现降维与特征解耦。「**让协方差为零**」是两个领域共同的目标。

# 交互：拖动相关系数，看免费午餐从哪来

## 怎么用

两个资产的期望与波动可调，主角是那根**相关系数**滑块。

## 当堂要做的三个实验

- 1. 拖  $\rho$ ，盯住**纵轴**——期望收益**纹丝不动**
- 2.  $\rho$  拖到 **+1**：曲线塌成直线，分散化的好处彻底消失
- 3.  $\rho$  拖到 **-1**：曲线碰到纵轴，存在一个权重让波动**恰好为 0**



交互演示：拖动相关系数，观察风险—收益曲线如何弯折

[https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/portfolio\\_frontier.html](https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/portfolio_frontier.html)

# 相关系数：把协方差标准化

协方差有个毛病：量纲

$\text{Cov}(2X, Y) = 2\text{Cov}(X, Y)$ ——**换个单位，数就变了**。用元还是用万元，算出来差一万倍，没法横向比较。

## 定义相关系数

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

可以证明  $|\rho| \leq 1$ 。特别地：

- $|\rho| = 1 \iff X, Y$  **线性**相关（一个是另一个的线性函数）
- $\rho = 0$  称为  $X, Y$  **不相关**
- **独立  $\Rightarrow$  不相关，反之不成立**（下下页给反例）

## $|\rho| \leq 1$ 的证明思路

对任意实数  $t$ ,  $E[(X - EX)t + (Y - EY)]^2 \geq 0$ 。展开是关于  $t$  的二次式  $D(X)t^2 + 2\text{Cov}t + D(Y) \geq 0$ ，故判别式  $4\text{Cov}^2 - 4D(X)D(Y) \leq 0$ ，即  $|\rho| \leq 1$ 。等号成立当且仅当那个平方恒为 0，即  $Y$  是  $X$  的线性函数。完整证明见教师笔记。

## 应用：CAPM 的 Beta 与特征去冗余

### 金融：资本资产定价模型的 $\beta$

$\beta$  衡量个股相对整个市场的风险：

$$\beta_i = \frac{\text{Cov}(R_i, R_m)}{D(R_m)} = \rho_{i,m} \frac{\sigma_i}{\sigma_m}$$

$\rho_{i,m}$  越高、 $\beta$  越大，股票「进攻性」越强；反之「防御性」强。 $\beta$  是构建投资策略的关键指标，而它本质上就是一个标准化的协方差。

### AI：特征工程中的去冗余

建模前算所有输入特征的**相关系数矩阵**是标准操作。

若两个特征的  $|\rho|$  接近 1，说明信息高度重复。为防止过拟合、提高效率，通常**从中移除一个**。

两个领域用的是同一个数：**金融用它度量「跟大盘走多紧」，AI 用它度量「这两列数据是不是在说同一件事」。**

## ★ 独立与不相关不是一回事

### 反例

$X$  等可能取  $-1, 0, 1$ , 令  $Y = X^2$ 。

**显然不独立**——知道  $X$  就完全知道  $Y$ 。

但算协方差:  $E(X) = 0$ ,

$$\text{Cov}(X, Y) = E(X^3) - E(X)E(X^2) = 0 - 0 = 0$$

**不相关。**

### 为什么会这样

$\rho$  只度量**线性**关系。 $Y = X^2$  是一个完美的非线性关系, 而它的线性成分恰好为零 (抛物线关于  $y$  轴对称, 正负抵消)。

$\rho = 0$  **只说明「没有线性关系」, 不说明「没有关系」。**

### 另一头的极端

若  $Y = aX + b$ , 则  $\text{Cov} = aD(X)$ 、 $D(Y) = a^2 D(X)$ ,

$$\rho = \frac{aD(X)}{|a|D(X)} = \text{sgn}(a)$$

**完全线性时**  $\rho = \pm 1$ , 符号由斜率决定。

# 二元正态分布

## 定义

$X, Y$  服从二元正态分布，若联合密度为

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x-\mu_1}{\sigma_1} \right) \left( \frac{y-\mu_2}{\sigma_2} \right) + \left( \frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\}$$

## 三条性质

- 边缘分布分别是  $X \sim N(\mu_1, \sigma_1^2)$ 、 $Y \sim N(\mu_2, \sigma_2^2)$
- 式子里那个  $\rho$  就是相关系数  $\rho(X, Y)$
- $\rho = 0$  是  $X, Y$  独立的充分必要条件

## ★ 第三条是个特例，别推广

上一页刚说「不相关不一定独立」，这里却说「等价」——**因为这里多了一个前提：联合分布是二元正态。**

在正态的世界里， $\rho$  装下了两个变量之间的全部关系；**出了这个世界就不行。** 这也解释了为什么金融模型偏爱正态——它让「相关性」这一个数就够用了。

## $n$ 维正态分布（选讲）

### 密度函数

记  $\mathbf{X}, \boldsymbol{\mu}$  为  $n$  维列向量,  $\mathbf{C}$  为**协方差矩阵** ( $c_{ij} = \text{Cov}(X_i, X_j)$ ) , 则

$$f(x_1, \dots, x_n) = (2\pi)^{-n/2} |\mathbf{C}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right\}$$

### 四条性质

- 每个分量都是正态; 反之若各分量正态**且相互独立**, 则整体是  $n$  维正态
- 任意**线性组合**是一维正态, 反之亦然
- 线性函数  $(Y_1, \dots, Y_k)$  仍是多维正态
- **相互独立**  $\iff$  **两两不相关**

注意第一条的「**且相互独立**」不能去掉: 各分量都正态, 不足以保证整体是多维正态。第四条是二维那条性质的推广, 同样只在正态世界成立。

## 两个不等式（选讲）

### 琴生不等式

若  $g$  是凹函数，则

$$E[g(X)] \leq g(E[X])$$

等号成立当且仅当  $g$  是线性函数（或  $X$  退化为常数）。

它正是开场那个悖论的解答方向。丹尼尔·伯努利 1738 年的想法是：人关心的不是钱  $X$ ，是效用  $g(X)$ ，而  $g$  是凹的（钱越多，多一块钱的价值越小）。于是即便  $E[X] = \infty$ ， $E[g(X)]$  仍可以是有限的。

### 柯西—施瓦茨不等式

$$[E(XY)]^2 \leq E(X^2) E(Y^2)$$

把  $X, Y$  换成离差  $X - EX$ 、 $Y - EY$ ，立刻得到

$$\text{Cov}(X, Y)^2 \leq D(X)D(Y)$$

这就是  $|\rho| \leq 1$ 。

## 本讲总结

### ① 回到开场那个赌局

它的期望**不是无穷大，而是不存在**——级数不绝对收敛，定义里那四个字就是为它准备的。

更深一层：即便期望存在，**一个数也不足以指导决策**。丹尼尔·伯努利 1738 年的答案是：人最大化的不是钱的期望，是**效用**的期望，而效用函数是凹的——由琴生不等式， $E[g(X)] \leq g(E[X])$ ，无穷的期望换成有限的效用。

**这就是期望效用理论的起点，也是整个现代经济学的基石之一。**

# 本讲总结 (续)

## ② 核心结论

- **期望**:  $E(X + Y) = E(X) + E(Y)$  不需要独立;  $E(XY) = E(X)E(Y)$  需要独立
- **方差**:  $D(X) = E(X^2) - (EX)^2$ ;  $D(aX + c) = a^2 D(X)$ ;  $D(X + Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$
- **协方差**就是  $E(XY) - E(X)E(Y)$ ——独立时两边的那个差
- **相关系数**  $\rho = \text{Cov}/(\sigma_X \sigma_Y)$ ,  $|\rho| \leq 1$ , 只度量线性关系
- **回归的本质是拟合条件期望**  $E[Y | X = x]$

## ③ 易错提醒

- 期望可加**不要独立**, 方差可加**要独立**——最高频的混淆
- $D(aX) = a^2 D(X)$ , 系数是**平方**
- **独立**  $\Rightarrow$  **不相关**, 但**不相关**  $\nRightarrow$  **独立** ( $Y = X^2$ ); 只有在正态下两者才等价

## ④ 下一讲

本讲算出  $E$  与  $D$ , 但它们描述的还是一个随机变量。如果把很多个独立同分布的随机变量**平均**起来会怎样?

绪论里布丰投针的频率会稳定到概率上、第 6 讲两个均匀分布相加变成梯形再加会变钟形——**这两件事其实是同一件事, 下一讲给它们名字: 大数定律与中心极限定理。**