

多维随机变量及其分布

第 6 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

把「密度」这个工具立起来了

连续型用 f 描述, $F(x) = \int_{-\infty}^x f$; 三个分布: 均匀、指数、正态。标准化 $Z = \frac{X-\mu}{\sigma}$ 让全世界只需一张表。

也学会了求函数的分布

离散型求和, 连续型**先求 G 再求导**, 别漏雅可比因子。

★ 留下的问题

到现在为止, 我们一次只研究**一个**随机变量。可现实里要紧的往往是**几个量之间的关系**。

而这正是上一讲那个故事真正的病根——**LTCM 的模型不是不知道单个资产的分布, 是低估了它们「一起动」的程度**。「一起动」该怎么描述?

为什么需要这一讲

场景 2006 年 10 月，Netflix 拿出 **100 万美元** 公开悬赏：谁能把它的推荐算法 Cinematch 的预测误差降低 **10%**，钱就归谁。同时公开了 **1 亿条** 真实评分数据。

冲突 最自然的思路是把两件事估准：**每部电影的评分分布**（这片子口碑如何）和**每个用户的打分习惯**（这人是不是打分偏严）。这条路走得通——但它**正是 Cinematch 已经在做的**，也正是那个要被超越 10% 的基准。三年、四万多次提交，冠军队伍的功夫**几乎全花在别处**。

悬念 因为「张三会不会喜欢这部片子」这件事，不在任何一个单独的分布里。
把 X 的分布和 Y 的分布都拿到手，还差一块信息——差的是什么？

来源

Netflix Prize: 2006 年 10 月启动，奖金 100 万美元，要求把 Cinematch 的均方根误差 (RMSE, 基准 0.9525) 降低 10%; 公开数据集约 1 亿条评分。竞赛历时三年、收到 44 014 次提交，2009 年由 BellKor's Pragmatic Chaos 以 10.06% 的改进 (RMSE 0.8712) 夺冠。Bennett & Lanning, *The Netflix Prize*, KDD Cup 2007 | 访问日期 2026-08-01。

本讲学习目标

- ① 掌握描述「关系」的三个工具：**联合分布**（鸟瞰全貌）、**边缘分布**（单独看一个）、**条件分布**（给定一个看另一个）
- ② 会判断两个随机变量是否**独立**，并知道独立意味着什么、不意味着什么
- ③ 会求两个随机变量的**线性组合**（和、差）的分布

现实中变量总是关联的：收入与受教育年限、个股收益与大盘、用户年龄与点击率。本讲是从「一个变量」跨到「一群变量」的桥。

二元随机变量

定义

设 X, Y 是两个随机变量，它们的**联合分布**指：对所有实数对的集合 C ，概率 $P[(X, Y) \in C]$ 的全体。

例

掷一个骰子、抛一枚硬币。 $X =$ 点数， $Y =$ 硬币是否正面（正 1 反 0）。

(X, Y) 的取值集是 $\{(1, 1), \dots, (6, 1), (1, 0), \dots, (6, 0)\}$ 共 12 个，每个概率 $\frac{1}{12}$ 。

注意**不是**把 X 和 Y 分开描述——是把 (X, Y) 当作**一个整体**，看它落在平面上各处的概率。这个视角的转换是本讲的全部难点。

离散型的联合概率函数

定义

若 X, Y 都是离散型, 其**联合概率函数**为

$$f(x, y) = P(\{X = x\} \cap \{Y = y\})$$

满足 $\sum_{\text{所有}(x,y)} f(x, y) = 1$ 。

和第 4 讲对比

一维时 $f(x) = P(X = x)$, 画出来是一排柱子; 二维时 $f(x, y)$ 画出来是**一张表** (或平面上一片带高度的点)。

归一化条件从「一排加起来是 1」变成「**整张表加起来是 1**」。

例：一个小镇的汽车与电视机

x	y			
	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

$f(x, y)$: X 汽车数、 Y 电视机数

题

随机抽一户，求 (1) 至少两辆车**且**至少两台电视的概率； (2) 恰有一辆车的概率。

解：在表上圈区域，然后加

(1) 把 $x \geq 2$ 且 $y \geq 2$ 的六格加起来：

$$P(X \geq 2, Y \geq 2) = 0.5$$

(2) 把 $x = 1$ 那**整行**加起来：

$$P(X = 1) = 0.2$$

留意第 (2) 问：它只问 X ，却要把 Y 的所有取值**加掉**。这个动作就是「边缘分布」，二十分钟后会正式命名。

交互：联合分布 · 边缘 · 条件

换个场景想同一张表

把「汽车数 × 电视机数」换成「对甲片的评分 × 对乙片的评分」，每位用户的一对评分 (x, y) 就是表里的一个数据点。所有点的分布就是联合概率函数 $f(x, y)$ ——这是推荐算法的原始矿藏。若 $f(5, 5)$ 很高，说明喜欢一部的人多半也喜欢另一部。

现在只看两件事

- 1. 右下角那一列 $f_1(x)$ 是每行加起来得到的
- 2. 整张表加起来等于 1

点任意一行看边缘；按住 Shift 点则选列。条件分布那半边先不看，讲到再回来。



交互演示：点行/列查看边缘分布与条件分布

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

连续型的联合密度函数

定义

若存在非负函数 f , 使得对 xy 平面上的每个集合 C 都有

$$P\{(X, Y) \in C\} = \iint_C f(x, y) \, dx \, dy$$

则称 X, Y 服从连续型联合分布, f 为**联合概率密度函数**。

用归一化定常数 (二维版)

例

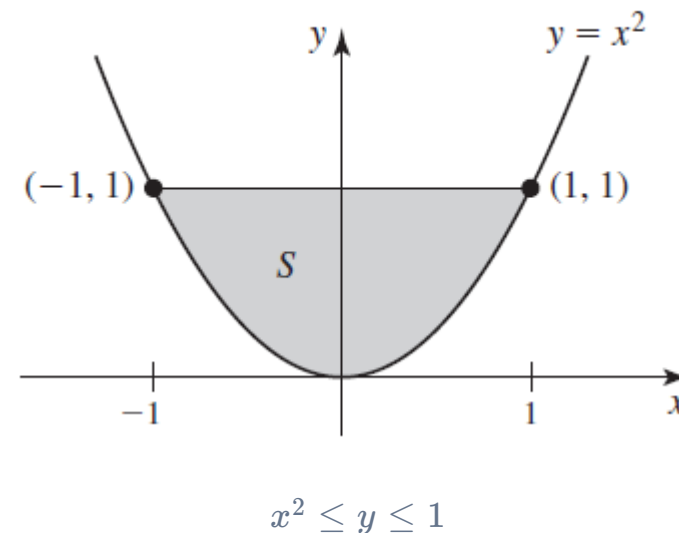
$f(x, y) = cx^2y$ ($x^2 \leq y \leq 1$), 其余为 0。求 c 。

解: 难的是定积分限, 不是积分

区域是抛物线 $y = x^2$ 与直线 $y = 1$ 围成的, x 从 -1 到 1 。先对 y 积, y 从 x^2 到 1 ; 再对 x 积:

$$\int_{-1}^1 \int_{x^2}^1 cx^2y \, dy \, dx = \frac{4}{21}c = 1$$

故 $c = \frac{21}{4}$ 。



二重积分题的全部难点在画图定限。 不画图直接写积分限, 十有八九错。这一条从本讲一直用到期末。

在区域上求概率

例（承上）

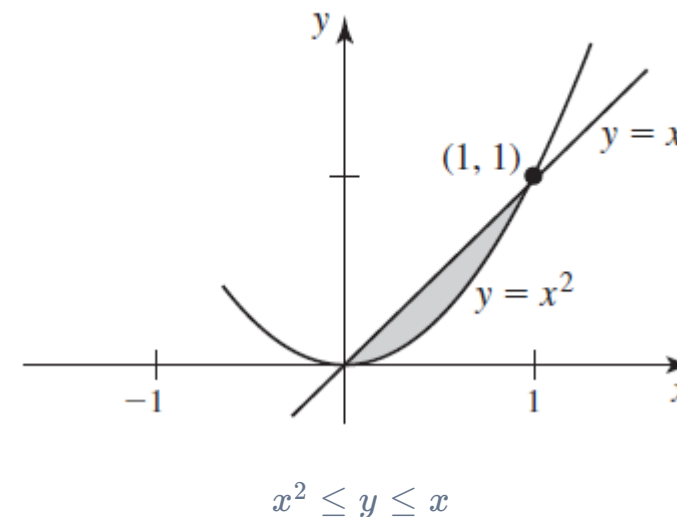
$f(x, y) = \frac{21}{4}x^2y$ ($x^2 \leq y \leq 1$)。求 $P(X \geq Y)$ 。

解

$X \geq Y$ 即 $y \leq x$ 。与原区域 $x^2 \leq y$ 求交：需 $x^2 \leq y \leq x$ ，这要求 $0 \leq x \leq 1$ 。

$$\iint_{S_0} f = \int_0^1 \int_{x^2}^x \frac{21}{4}x^2y \, dy \, dx = \frac{3}{20}$$

关键是把两个约束求交： 题目给的区域 $\cap$ 事件给的区域。漏掉任一个，积分限就错。



联合分布函数

定义

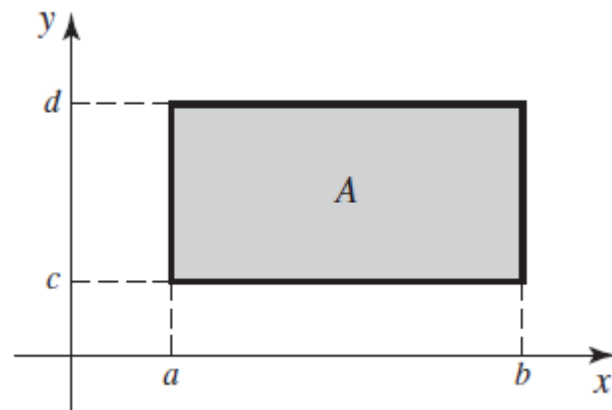
$$F(x, y) = P(X \leq x, Y \leq y)$$

X 的分布函数由 $F_1(x) = \lim_{y \rightarrow +\infty} F(x, y)$ 得到, Y 的由 $F_2(y) = \lim_{x \rightarrow +\infty} F(x, y)$ 得到。

用它算矩形区域的概率

$$\begin{aligned} &P(a < X \leq b, c < Y \leq d) \\ &= F(b, d) - F(a, d) - F(b, c) + F(a, c) \end{aligned}$$

四项，两加两减。这是第 1 讲容斥原理在二维的样子——减两次多减了一块，要加回来。



加加减减的几何含义

联合分布函数与联合密度的互换

若 X, Y 连续型且有联合密度 f , 则

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(r, s) \, dr \, ds$$

反过来**求二阶混合偏导**:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x}$$

例

X, Y 只在 $[0, 2]$ 取值, $F(x, y) = \frac{1}{16}xy(x + y)$ 。求 $F_1(x)$ 与 $f(x, y)$ 。

F_1 : 令 $y = 2$ (Y 最大只到 2, 故 $y \rightarrow \infty$ 的极限就是 $F(x, 2)$) :

$$F_1(x) = \frac{1}{8}x(x + 2), \quad 0 \leq x \leq 2$$

f : $F = \frac{1}{16}(x^2y + xy^2)$, $\partial_x F = \frac{1}{16}(2xy + y^2)$, 再对 y 求导:

$$f(x, y) = \frac{1}{8}(x + y), \quad 0 \leq x, y \leq 2$$

小结：联合分布这一段

- **离散型**用联合概率函数 $f(x, y) = P(X = x, Y = y)$, 整张表加起来是 1
- **连续型**用联合密度 $f(x, y)$, 二重积分等于 1; 概率是曲面下的体积
- **联合分布函数** $F(x, y) = P(X \leq x, Y \leq y)$ 对两类通用; 与密度互为二重积分/二阶混合偏导
- 算概率的全部难点是**画图定积分限**, 尤其要把题目区域与事件区域**求交**

边缘分布：抛开关系，单独看一个

定义

二元随机变量 $\{X, Y\}$ 中， X 或 Y 自己的分布称为**边缘分布**。

记 X 的边缘分布函数为 $F_1(x)$ 、边缘概率（密度）函数为 $f_1(x)$ ； Y 的记为 $F_2(y)$ 、 $f_2(y)$ 。

名字的来历

把联合概率写成一张表，每行加起来写在**右边缘**、每列加起来写在**下边缘**——「边缘分布」就是这么来的。

离散型的边缘分布

x	y				Total
	1	2	3	4	
1	0.1	0	0.1	0	0.2
2	0.3	0	0.1	0.2	0.6
3	0	0.2	0	0	0.2
Total	0.4	0.2	0.2	0.2	1.0

右边一列、下面一行就是边缘

公式

X 的边缘概率函数

$$f_1(x) = \sum_{\text{所有 } y} f(x, y)$$

Y 的边缘概率函数

$$f_2(y) = \sum_{\text{所有 } x} f(x, y)$$

一句话

把另一个变量「加掉」。这正是 slide 7 第 (2) 问做过的事——现在它有名字了。

边缘分布的意义：从「关系」到「流行度」

核心思想

边缘分布让我们**抛开变量间的关系**，单独审视每个变量自身的特性——比如一部电影的「大众口碑」。

$$f_1(x) = \sum_{y=1}^5 f(x, y) \quad (\text{对 } Y \text{ 的所有评分求和})$$

推荐系统里的用途

- 对联合分布求和，得到某部电影的边缘分布 $f_1(x)$
- 它描述这部片子在**所有用户**中的总体评分状况
- 这是算「平均分」、生成「Top 10 排行榜」等**非个性化推荐**的直接依据

注意「**非个性化**」这四个字。**边缘分布对所有人给出同一个答案**——这正是开场那个悬念的一半：它有用，但不够。

连续型的边缘分布

公式：求和换成积分

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) \, dy, \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) \, dx$$

边缘分布函数则由极限得到： $F_1(x) = \lim_{y \rightarrow \infty} F(x, y)$, $F_2(y) = \lim_{x \rightarrow \infty} F(x, y)$ 。

例（承前）

$f(x, y) = \frac{21}{4}x^2y$ ($x^2 \leq y \leq 1$)。求两个边缘密度。

$$f_1(x) = \int_{x^2}^1 \frac{21}{4}x^2y \, dy = \frac{21}{8}x^2(1 - x^4), \quad -1 \leq x \leq 1$$

$$f_2(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2y \, dx = \frac{7}{2}y^{5/2}, \quad 0 \leq y \leq 1$$

两次积分的限完全不同：算 f_1 时 y 从 x^2 到 1；算 f_2 时 x 从 $-\sqrt{y}$ 到 $\sqrt{y}$ 。**又是画图定限。**

★ 边缘分布决定不了联合分布

例

掷两枚硬币，每枚各掷十次。考虑两种情形：

1. X, Y 分别是**两枚**硬币的正面次数
2. X, Y **都是第一枚**硬币的正面次数

求 X, Y 的联合分布。

解：边缘相同，联合完全不同

两种情形下 X 与 Y 的**边缘分布都是** $B(10, 0.5)$ 。但

$$(1) \ p_{ij} = C_{10}^i C_{10}^j 0.5^{20}; \quad (2) \ p_{ij} = C_{10}^i 0.5^{10} \ (i = j), \ 0 \ (i \neq j)$$

这就是开场那个悬念的答案

把两个边缘都拿到手，还差的那一块，就是「它们怎么一起动」。 (2) 里知道 X 就完全知道 Y ，(1) 里知道 X 对 Y 毫无帮助——而边缘一模一样。

交互：边缘一样，联合不同

怎么用

三个联合分布轮流切换，**盯住右边那两条边缘柱子。**

要让学生看见的

切换时联合分布的表格**面目全非**（对角线 → 满格 → 反对角线），而边缘分布**纹丝不动**。

三种情形分别是：互不影响、知道一个就知道另一个、一个大另一个必定小。



交互演示：三个联合分布，边缘完全相同

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

接下来

既然边缘分布不足以决定联合分布，那自然要问：**什么时候它足够？**

换句话说：什么时候「两个变量各自的分布」就是全部信息，再没有额外的「一起动」需要交代？

随机变量的独立性

定义

对任意实数集 A, B , 若事件 $\{X \in A\}$ 与 $\{Y \in B\}$ 独立, 即

$$P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A) P(Y \in B)$$

则称随机变量 X, Y **相互独立**。

两个充要条件 (实际判定用这个)

$$X, Y \text{ 独立} \iff F(x, y) = F_1(x)F_2(y) \quad \forall x, y$$

$$X, Y \text{ 独立} \iff f(x, y) = f_1(x)f_2(y) \quad \forall x, y$$

证明思路

定义取 $A = (-\infty, x]$ 、 $B = (-\infty, y]$ 即得分布函数形式; 对连续型两边求二阶混合偏导得密度形式; 离散型取单点集直接得到。完整证明见教师笔记。

回答上一页的问题: 独立时, 边缘分布就是全部信息 ——联合分布是两个边缘的乘积, 没有额外内容。**不独立时**才需要联合分布。

应用：独立性在推荐系统里的用处

核心思想

识别变量间的**独立性**，可以帮系统避免逻辑不相关的「无效推荐」。

场景

系统发现，用户对科幻片 X 的评分与对儿童动画 Y 的评分**相互独立**：

$$f(x, y) = f_1(x)f_2(y)$$

- **避免无效推荐**：知道他是科幻迷，推儿童动画并不会提高成功率
- **构建用户画像**：「科幻偏好」与「亲子偏好」是两个相对独立的维度，应在每个维度上分别推荐

找相关性用于推荐，找独立性用于划分边界。两者同等重要——这正是「协同过滤」与「多维度画像」两条技术路线的分工。

独立性例题（一）：由独立拼出联合

例

X, Y 相互独立, 都服从 $g(x) = 2x$ ($0 \leq x \leq 1$) 。
求 $P(X + Y \leq 1)$ 。

独立 $\Rightarrow$ 联合是乘积:

$$f(x, y) = 4xy, \quad 0 \leq x, y \leq 1$$

$$P(X + Y \leq 1) = \int_0^1 \int_0^{1-x} 4xy \, dy \, dx = \frac{1}{6}$$

例

$f(x, y) = ke^{-(x+2y)}$ ($x, y \geq 0$) 。 X, Y 是否独立?

归一化给出 $k = 2$ 。两个边缘:

$$f_1(x) = e^{-x} \ (x \geq 0), \quad f_2(y) = 2e^{-2y} \ (y \geq 0)$$

$$f_1 f_2 = 2e^{-x-2y} = f, \quad \text{独立。}$$

这也是最好认的一种形态：**密度能拆成「只含 x 的因子 $\times$ 只含 y 的因子」，且定义域是矩形，就独立。**

独立性例题（二）：定义域本身会破坏独立

例

$f(x, y) = kx^2y^2$ ($x^2 + y^2 \leq 1$)。 X, Y 是否独立？

密度看起来完全可分离 ($kx^2 \cdot y^2$)，但不独立。

一个反例就够： $f(0.9, 0.9) = 0$ （因为 $0.81 + 0.81 > 1$ ，在区域外），而 $f_1(0.9) \neq 0$ 、 $f_2(0.9) \neq 0$ ，故 $f \neq f_1f_2$ 。

例

$f(x, y) = \frac{15}{4}x^2$ ($0 \leq y \leq 1 - x^2$)。求边缘并判断独立性。

$$f_1(x) = \frac{15}{4}x^2(1 - x^2), \quad f_2(y) = \frac{5}{2}(1 - y)^{3/2}$$

乘积不等于 f ，**不独立**——同样是定义域不是矩形。

★ 记住这条

定义域不是矩形 (x 的范围依赖 y)，**就一定不独立**——因为知道了 X 就限制了 Y 的取值范围，这本身就是信息。可分离只是必要条件，不是充分条件。

独立性例题（三）：离散与连续各一道

例（离散）

$f(x, y) = \frac{1}{30}(x + y)$, $x = 0, 1, 2$, $y = 0, 1, 2, 3$ 。求边缘并判断独立性。

$$f_1(x) = \sum_{y=0}^3 \frac{x+y}{30} = \frac{1}{15}(2x + 3)$$

$$f_2(y) = \sum_{x=0}^2 \frac{x+y}{30} = \frac{1}{10}(1 + y)$$

$f_1 f_2 \neq f$, **不独立**。

例（连续）

$f(x, y) = \frac{3}{2}y^2$ ($0 \leq x \leq 2$, $0 \leq y \leq 1$)。

$$f_1(x) = \int_0^1 \frac{3}{2}y^2 dy = \frac{1}{2}, \quad f_2(y) = \int_0^2 \frac{3}{2}y^2 dx = 3y^2$$

$f_1 f_2 = \frac{3}{2}y^2 = f$, **独立**。

进一步：事件 $\{X < 1\}$ 与 $\{Y \geq \frac{1}{2}\}$ 也独立——**随机变量独立 $\Rightarrow$ 由它们生成的任意事件独立**。

小结：边缘与独立性

- **边缘** = 把另一个变量加掉（或积掉）： $f_1(x) = \sum_y f$ 或 $\int f \, dy$
- **联合** $\rightarrow$ **边缘永远可以**；**边缘** $\rightarrow$ **联合不行**。差的那块就是「怎么一起动」
- **独立** $\iff f(x, y) = f_1(x)f_2(y)$ **处处成立**；只要一格不成立就否掉
- **定义域不是矩形** $\Rightarrow$ **一定不独立**，哪怕密度看起来可分离

条件分布：给定一个，另一个怎么表现

Stolen X	Brand Y					Total
	1	2	3	4	5	
0	0.129	0.298	0.161	0.280	0.108	0.976
1	0.010	0.010	0.001	0.002	0.001	0.024
Total	0.139	0.308	0.162	0.282	0.109	1.000

$X = 1$ 表示被偷， Y 是品牌

问题

由这张表能算出 X 的边缘概率，也就是**总的失窃率**。但——**给定某个品牌，它的失窃率是多少？**

用第 3 讲的条件概率

$$P(X = 1 \mid Y = y) = \frac{P(X = 1 \cap Y = y)}{P(Y = y)} = \frac{f(1, y)}{f_2(y)}$$

取出那一系列，再除以列和。

离散型的条件分布

定义

设 X, Y 的联合概率函数为 f 。对每个使 $f_2(y) > 0$ 的 y ，定义

$$g_1(x | y) = \frac{f(x, y)}{f_2(y)}$$

称为给定 $Y = y$ 时 X 的**条件概率函数**。对称地，对 $f_1(x) > 0$ 的 x

$$g_2(y | x) = \frac{f(x, y)}{f_1(x)}$$

它确实是一个概率函数

$$\sum_x g_1(x | y) = \frac{\sum_x f(x, y)}{f_2(y)} = \frac{f_2(y)}{f_2(y)} = 1$$

除以 $f_2(y)$ 这个动作，就是让那一列重新加到 1。这是第 3 讲「样本空间收缩」在二维表上的样子。

交互：回到那张表，看条件分布

这次看右下角

点一行，右下会画出 $g_2(y \mid x)$ 并显示它**加起来等于1**。

要做的对照

切到「两枚硬币（独立）」，点任意一行——**条件分布都长一样**，而且等于边缘分布。
再切到「同一枚硬币」，条件分布退化成一根柱子。

独立的含义就是：条件分布不随条件而变。



交互演示：点行/列查看条件分布的归一化

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

应用：「猜你喜欢」的数学原理

核心思想

条件分布是实现「千人千面」个性化推荐的引擎：**用已知信息预测未知偏好。**

$$g_2(y \mid x) = \frac{f(x, y)}{f_1(x)}$$

场景

- 用户给某部片子 X 打了 5 星，应该推荐什么？
- 算在 $X = 5$ 的**条件**下，他给另一部片子 Y 打各个分数的条件分布 $g_2(y \mid 5)$
- 对所有候选片子都算一遍，推荐**条件期望评分最高**的那部

这就是开场那个 100 万美元的答案

边缘分布 f_1 给出「大众口碑」，对谁都一样；**条件分布 $g_2(y \mid x)$ 才是因人而异的那部分。** 这正是「协同过滤」的核心，也是 Cinematch 与冠军算法的分野。

连续型的条件分布：一个必须绕过的困难

场景

加工分两步，第一步用时 Y ，总用时 X ，联合密度 $f(x, y) = e^{-x}$ ($0 \leq y \leq x < \infty$)。已知 Y 之后，想更新关于 X 的分布。

困难

不能直接套条件概率公式——因为 $P(Y = y) = 0$ ，分母是 0。

定义（绕过去的办法）

对每个使 $f_2(y) > 0$ 的 y ，直接**定义**

$$g_1(x | y) = \frac{f(x, y)}{f_2(y)}$$

对 $f_2(y) = 0$ 的点可任意定义，只要 g_1 仍是一个密度函数。

形式与离散型**一模一样**，但地位不同：离散型那个是推出来的，连续型这个是**定义**的——因为原来的定义在这里失效了。

连续条件分布：两道例题

例（承上）

$f(x, y) = e^{-x}$ ($0 \leq y \leq x$)。求 $g_1(x | y)$ 。

先求边缘：

$$f_2(y) = \int_y^{\infty} e^{-x} dx = e^{-y}$$
$$g_1(x | y) = \frac{e^{-x}}{e^{-y}} = e^{y-x}, \quad x \geq y$$

认得出来吗？ 这是**平移到 y 的指数分布**——已经用了 y 分钟，剩余时间仍服从原来的指数分布。第 5 讲的无记忆性，在这里又出现了。

例：由条件分布拼出联合分布

机器次品率 P 未知，设 $P \sim U[0, 1]$ 。抽查 n 件，次品数为 X 。求 (X, P) 的联合分布。

给定 $P = p$ 时 $X \sim B(n, p)$ ：

$$g_1(x | p) = C_n^x p^x (1 - p)^{n-x}$$

由乘法法则 $f(x, p) = g_1(x | p)f_2(p)$ ，而 $f_2(p) = 1$ ：

$$f(x, p) = C_n^x p^x (1 - p)^{n-x}$$

这是贝叶斯统计的雏形：把未知参数也当成随机变量。

条件分布例题（一）

例：圆盘上的均匀分布

(X, Y) 在 $S = \{(x-1)^2 + (y+2)^2 \leq 9\}$ 上均匀分布。求 $g_2(y | x)$ 与 $P(Y > 0 | X = 2)$ 。

$f = \frac{1}{9\pi}$ 。竖着切一刀， $f_1(x) = \frac{2}{9\pi} \sqrt{9 - (x-1)^2}$ ，故

$$g_2(y | x) = \frac{1}{2} [9 - (x-1)^2]^{-1/2}$$

是均匀分布——竖切得到的线段上处处等可能。

$$P(Y > 0 | X = 2) = \frac{2 - \sqrt{2}}{4} \approx 0.146$$

例

$f(x, y) = c(x + y^2)$ ($0 \leq x, y \leq 1$)。求 $g_1(x | y)$ 与 $P(X < \frac{1}{2} | Y = \frac{1}{2})$ 。

$f_2(y) = c(\frac{1}{2} + y^2)$ ，于是 c **约掉了**：

$$g_1(x | y) = \frac{x + y^2}{\frac{1}{2} + y^2}$$

$$P(X < \frac{1}{2} | Y = \frac{1}{2}) = \int_0^{1/2} \frac{x + \frac{1}{4}}{\frac{3}{4}} dx = \frac{1}{3}$$

右题里 c **自动约掉**——求条件分布常常不必先定归一化常数。省一步。

条件分布例题 (二)

例

$f(x, y) = \frac{3}{16}(4 - 2x - y)$ ($x, y > 0, 2x + y < 4$)。
求 $g_2(y | x)$ 与 $P(Y \geq 2 | X = 0.5)$ 。

$$f_1(x) = \int_0^{4-2x} f \, dy = \frac{3}{8}(x-2)^2$$

$$g_2(y | x) = \frac{4 - 2x - y}{2(x-2)^2}$$

$$P(Y \geq 2 | X = 0.5) = \int_2^3 g_2(y | 0.5) \, dy = \frac{1}{9}$$

例：泊松稀释

总产量 $Y \sim \text{Poisson}(\lambda)$ ，每件独立地以概率 p 为次品。求次品数 X 的分布。

给定 $Y = y$ 时 $X \sim B(y, p)$ 。由乘法法则再对 y 求和：

$$P(X = x) = \sum_{y=x}^{\infty} C_y^x p^x (1-p)^{y-x} \frac{\lambda^y}{y!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} e^{-\lambda p}$$

结果仍是泊松分布，参数变成 λp 。 漂亮得像变魔术，但每一步都是本讲的工具。

接下来

三个工具齐了：联合看全貌、边缘看单个、条件看联动。

但还有一类问题没解决：两个资产的收益 X 与 Y ，我关心的往往不是它们各自，而是**组合的总收益** $X + Y$ 。
它的分布怎么求？

离散型的线性组合

例

X_1 等可能取 1, 2, 3, 4, X_2 等可能取 1, 2, 3, 相互独立。求 $Y = X_1 + X_2$ 的分布。

解：穷举所有能凑出该和的组合，再相加

Y 取 2, ..., 7, 联合有 $4 \times 3 = 12$ 种等可能组合：

y	2	3	4	5	6	7
P	1/12	2/12	3/12	3/12	2/12	1/12

比如 $Y = 4$ 有 (1, 3), (2, 2), (3, 1) 三种。

一般地 $P(Y = y) = \sum_x f(x, y - x)$ ——**沿着一条斜线把联合分布加起来**。 $X_1 - X_2$ 、 $2X_1 - 3X_2$ 同理，只是斜线的方向变了。

两个可加的分布

二项分布之和

若 $X_1, \dots, X_k$ 相互独立, $X_i \sim B(n_i, p)$ (p 必须相同), 则

$$X_1 + \dots + X_k \sim B(n_1 + \dots + n_k, p)$$

为什么显然: $B(n, p)$ 就是 n 次独立伯努利试验的成功数。把 n_1 次和 n_2 次拼起来, 就是 $n_1 + n_2$ 次。

泊松分布之和

若 $X_1, \dots, X_k$ 相互独立, $X_i \sim \text{Poisson}(\lambda_i)$, 则

$$X_1 + \dots + X_k \sim \text{Poisson}(\lambda_1 + \dots + \lambda_k)$$

直观: 两个独立的泊松过程叠加, 还是泊松过程, 速率相加。「这条路每小时来 10 辆车, 那条路来 6 辆, 合起来 16 辆。」

p 相同这个条件不能省。 $B(3, 0.5) + B(3, 0.9)$ 不是二项分布 —— 它连「每次成功率相同」这个前提都不满足。

连续型的和：卷积公式

例

队列前两位顾客一起走。 X_1, X_2 独立且都服从 $f(x) = 2e^{-2x}$ ($x > 0$)。求总等待时间 $Y = X_1 + X_2$ 的分布。

老办法：先求 G 再求导 (第 5 讲)

$$G(y) = P(X_1 + X_2 \leq y) = \int_0^y \int_0^{y-x_2} 4e^{-2x_1-2x_2} dx_1 dx_2 = 1 - e^{-2y} - 2ye^{-2y}$$
$$g(y) = G'(y) = 4ye^{-2y}$$

卷积公式 (把上面这套动作一般化)

$Z = X + Y$ 的密度为

$$f_Z(z) = \int_{-\infty}^{\infty} f(z-y, y) dy$$

若 X, Y **独立**, 则

$$f_Z(z) = \int_{-\infty}^{\infty} f_1(z-y)f_2(y) dy$$

正态分布之和仍是正态分布

定理

若 $X_1, \dots, X_k$ 相互独立, $X_i \sim N(\mu_i, \sigma_i^2)$, 则

$$X_1 + \dots + X_k \sim N(\mu_1 + \dots + \mu_k, \sigma_1^2 + \dots + \sigma_k^2)$$

证明思路

对 $k = 2$ 用卷积公式: 把两个正态密度相乘后对 y 积分, 配方后可分离出一个关于 y 的完全平方 (积掉给常数) 与一个关于 z 的正态核。一般 k 由归纳得到。完整推导见教师笔记。

★ 注意是方差相加, 不是标准差相加

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} < \sigma_X + \sigma_Y$$

这就是「分散化降低风险」的数学来源——两个独立资产的组合, 波动率小于各自波动率之和。

但请回到开场: 上面这条要求独立。资产之间一旦高度相关, 方差就不再简单相加, 分散化的效果会大打折扣——第 7 讲会给出带协方差的一般公式。

例：两个均匀分布之和

例

投资者同时持股票与债券。年末股票价值 $X \sim U[1000, 4000]$ ，债券价值 $Y \sim U[800, 1200]$ ，两者独立。求组合价值 $Z = X + Y$ 的分布。

解：卷积给出一个梯形

$$f_Z(z) = \begin{cases} 8.333 \times 10^{-7}(z - 1800), & 1800 < z \leq 2200 \\ 3.333 \times 10^{-4}, & 2200 < z \leq 4800 \\ 8.333 \times 10^{-7}(5200 - z), & 4800 < z \leq 5200 \\ 0, & \text{其他} \end{cases}$$

形状值得看一眼

两个均匀分布相加，结果**不再是均匀分布**，而是一个梯形——中间平、两头斜。**再加几个，就会越来越像钟形。**

这是第 8 讲**中心极限定理**最早的一丝征兆。

不只是和： $Z = XY$ 的分布

例

(X, Y) 在矩形 $G = \{0 \leq x \leq 2, 0 \leq y \leq 1\}$ 上均匀分布。求以 X, Y 为边长的矩形面积 $S = XY$ 的密度。

解：仍是「先求 G 再求导」

$P(XY \leq s)$ 对应双曲线 $xy = s$ 下方的区域，积分后求导得

$$f(s) = \begin{cases} \frac{1}{2}(\ln 2 - \ln s), & 0 < s \leq 2 \\ 0, & \text{其他} \end{cases}$$

注意 $f(s) \rightarrow \infty$ 当 $s \rightarrow 0^+$ ——**又一次：密度可以趋于无穷**，但积分仍等于 1。第 5 讲说过三遍的那件事。

最大值与最小值的分布

定理

设 X, Y 独立, 分布函数分别为 F_X, F_Y , 则

$$F_{\max}(z) = F_X(z)F_Y(z)$$
$$F_{\min}(z) = 1 - [1 - F_X(z)][1 - F_Y(z)]$$

证明思路

$\{\max \leq z\} = \{X \leq z\} \cap \{Y \leq z\}$, 由独立性概率相乘; $\{\min > z\} = \{X > z\} \cap \{Y > z\}$, 同理相乘后取对立事件。两行就完了。完整证明见教师笔记。

为什么这么简单

因为「最大值不超过 z 」等价于「**每一个**都不超过 z 」, 「最小值超过 z 」等价于「**每一个**都超过 z 」——**都是把 max/min 翻译成「所有」, 再用独立性拆开。**

第 5 讲那道 n 个灯泡的题 (Y_1 服从 $\text{Exp}(n\lambda)$), 用的正是 $F_{\min}$ 这条。

本讲总结

① 回到开场那个问题

Netflix 的 100 万美元，买的是**联合分布里那块边缘分布装不下的信息**。

把每部电影的评分分布、每个用户的打分习惯都估准，也只是两个边缘分布——而「这个人会不会喜欢这部片子」藏在**条件分布** $g_2(y | x)$ 里。 **边缘对所有人给同一个答案，条件才因人而异**。

② 核心结论

- **三个工具**：联合 $f(x, y)$ 看全貌；边缘 $f_1 = \sum_y f$ 或 $\int f dy$ 看单个；条件 $g_2(y | x) = f / f_1$ 看联动
- **联合 $\rightarrow$ 边缘永远可以，反过来不行**——差的是「怎么一起动」
- **独立 $\iff f = f_1 f_2$ 处处成立**，等价于「条件分布不随条件而变」
- **和的分布**：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

本讲总结（续）

③ 易错提醒

- **二重积分的限要画图定**，还要把题目区域与事件区域**求交**
- **定义域不是矩形就一定不独立**，哪怕密度看起来可以拆成 x 的因子乘 y 的因子
- **方差相加，不是标准差相加**： $\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$ ，且要求独立
- 二项分布可加要求 p **相同**；泊松可加没有额外条件

④ 下一讲

本讲说「正态之和的方差相加」，但那有个前提：**独立**。不独立时该怎么算？**「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**。

下一讲：随机变量的数字特征——期望、方差，以及描述「一起动」的协方差。