



# 概率论笔记

作者：肖程哲

时间：July 7, 2022



苟日新，日日新，又日新

# 目录

<b>第 1 章 概率基础</b>	<b>1</b>
1.1 概率空间 . . . . .	1
1.2 古典概型与几何概率 . . . . .	3
1.2.1 古典概型 . . . . .	3
1.3 条件概率 . . . . .	3
<b>第 2 章 随机变量</b>	<b>5</b>
2.1 随机变量的分布 . . . . .	5
2.2 多元随机变量 . . . . .	7
2.2.1 边际分布 . . . . .	8
2.2.2 独立 . . . . .	8
2.2.3 条件分布 . . . . .	9
2.3 随机变量的函数 . . . . .	10
<b>第 3 章 常见分布</b>	<b>8</b>
3.1 离散分布 . . . . .	8
3.2 连续分布 . . . . .	8
3.3 正态分布及其导出分布 . . . . .	8
<b>第 4 章 随机变量的数值特征</b>	<b>9</b>
4.1 期望与方差 . . . . .	9
4.2 矩母函数与特征函数 . . . . .	9
4.3 熵与信息 . . . . .	9
<b>第 5 章 概率极限</b>	<b>10</b>
5.1 收敛 . . . . .	10
5.2 大数定理 . . . . .	10
5.3 中央极限定理 . . . . .	10
<b>第 A 章 基本数学工具</b>	<b>11</b>
A.1 排列与组合 . . . . .	11

## 第2章 随机变量

### 内容提要

- 离散与连续随机变量
- 一元与多元
- cdf, pmf, pdf
- 条件分布
- 独立随机变量
- 随机变量函数的分布
- 次序随机变量

在概率论中, 主要关心  $X$  取值于数值集合  $\mathcal{X}$  中某个子集  $B$  的可能性, 即希望得到  $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$ . 概率论不关心具体的样本点  $\omega \in \Omega$ , 将其记为  $\{X \in B\} = X^{-1}(B)$ . 由于  $\mathbb{P}$  定义在  $\mathcal{F}$  上, 故需  $X^{-1}(B) \in \mathcal{F}$ .

#### 定义 2.1 (可测性)

设所有值得关心的  $B \subset \mathcal{X}$  组成  $\mathcal{F}_{\mathcal{X}}$ , 且  $\forall B \in \mathcal{F}_{\mathcal{X}}$  都满足  $\{X \in B\} \in \mathcal{F}$ , 则称  $X$  为  $\mathcal{F}/\mathcal{F}_{\mathcal{X}}$  可测的. 当  $\mathcal{F}_{\mathcal{X}}$  不引起混淆时, 简记为关于  $\mathcal{F}$  可测, 写作  $X \in \mathcal{F}$ .



由于原像保持交、并、补等集合运算, 且  $\mathcal{F}$  是  $\sigma$  代数, 可将  $\mathcal{F}_{\mathcal{X}}$  扩张为合适的最小的  $\sigma$  代数, 即  $\sigma(\mathcal{F}_{\mathcal{X}})$ , 因此可测映射的定义不妨只考虑  $\mathcal{F}_{\mathcal{X}}$  是  $\sigma$  代数的情况.

#### 定义 2.2 (随机变量)

为了表示因随机性而变动的量, 称可测映射(measurable mapping)

$$X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathcal{X}, \mathcal{F}_{\mathcal{X}}), \quad \omega \in \Omega \mapsto X(\omega) \in \mathcal{X}$$

为随机元 (random element), 也称随机变量 (random variable). 其中  $\mathcal{F}_{\mathcal{X}}$



由于只考虑  $\mathcal{F}_{\mathcal{X}}$  是  $\sigma$  代数的情况, 可将随机变量看作将原概率空间映射到新概率空间的方式. 新样本空间由 Borel 点集 构成, 对应的概率测度等于原像的.

**注** 使用随机变量  $X$  时, 有两个可能的含义:

- $X$  的 (随机) 取值
- $X$  的分布

#### 定义 2.3 (离散与连续随机变量)

当  $\mathcal{X}$  是(至多可数的)离散点集,  $\mathcal{F}_{\mathcal{X}}$  由  $\mathcal{X}$  的所有子集组成, 则称其为离散随机变量 (discrete random variable).

当随机变量  $\mathcal{X} = \mathbb{R}^n$ , 考虑  $\mathcal{F}_{\mathcal{X}}$  为  $\{\prod_{i=1}^n (-\infty, x_i] : x_1, \dots, x_n \in \mathbb{R}\}$  生成的 Borel 代数 (最小的  $\sigma$  代数), 则称其为连续随机变量 (continuous random variable).



## 2.1 随机变量的分布

#### 定义 2.4

称随机元  $X$  诱导的概率测度

$$\mathbb{P}\{X \in \bullet\}, \bullet \in \mathcal{F}_{\mathcal{X}}$$

为  $X$  的概率分布 (distribution/law)



**注** 对于随机变量, 他的取值是随机的, 但他的分布是固定的

**定义 2.5 (单变量分布函数)**

一个函数  $F : \mathbb{R} \rightarrow [0, 1]$  称为一个单变量分布函数, 当其满足以下性质时:

**单调性**  $F(x_1) \leq F(x_2), \forall x_1 < x_2$

**右连续性**  $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$

**有界性**  $\lim_{n \rightarrow -\infty} F(x) = 0, \lim_{n \rightarrow \infty} F(x) = 1$



**性质**  $F(x)$  最多只有可数个间断点

**命题 2.1**

对每个分布  $Q : \mathcal{B}_1 \rightarrow [0, 1]$  都存在唯一一个分布函数  $F_Q : \mathbb{R} \rightarrow [0, 1]$  使得  $F_Q(x) = Q[(-\infty, x)], \forall x \in \mathbb{R}$  成立。

**命题 2.2**

对每个分布函数  $F : \mathbb{R} \rightarrow [0, 1]$  都存在唯一一个分布  $Q_F : \mathcal{B}_1 \rightarrow [0, 1]$  使得  $Q_F[(-\infty, x)] = F(x), \forall x \in \mathbb{R}$  成立。

**定理 2.1**

分布函数可以唯一决定概率分布, 即:

$$Q_{F_Q} = Q, \quad F_{Q_F} = F$$

把随机变量  $X$  服从分布函数  $F(x)$  简记作  $X \sim F(x)$



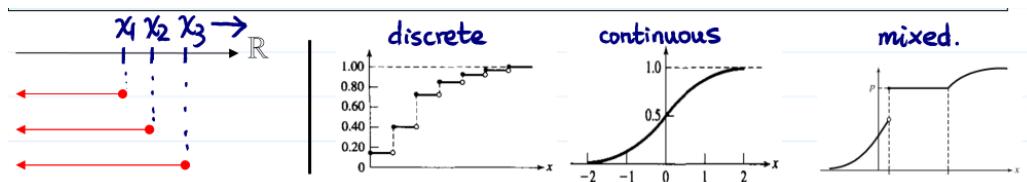
	离散	连续
一元随机变量	概率质量函数 (pmf)	概率密度函数 (pdf)
	累积分布函数 (cdf)	
	矩母函数/特征函数 (mgf/chf)	
多元随机变量	联合概率质量函数 (joint pmf)	联合概率密度函数 (joint pdf)
	联合累积分布函数 (joint cdf)	
	联合矩母函数/特征函数 (joint mgf/chf)	

**定义 2.6 (累积分布函数)**

此时  $X = (X_1, \dots, X_n)^\top$  的分布由(累积)分布函数 (cumulative distribution function, c.d.f.)

$$F_X(x_1, \dots, x_n) = \mathbb{P}\{X_1 \leq x_1, \dots, X_n \leq x_n\}, \quad x_1, \dots, x_n \in \mathbb{R}.$$

唯一刻画. 把随机变量  $X$  服从分布函数  $F(x)$  简记作  $X \sim F(x)$

**定义 2.7 (概率质量函数)**

当且仅当函数  $p(x)$  满足下述条件时, 被称为概率质量函数 (probability mass function, p.m.f.):

- $p(x) \geq 0$
- $\sum_{x \in \mathcal{X}} p(x) = 1$



当  $X$  是离散型随机变量, 设  $\mathcal{F}_X$  由  $\mathcal{X}$  的所有子集组成, 此时  $X$  的分布由

$$p_X(x) = \mathbb{P}\{X = x\} = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}), \quad x \in \mathcal{X}$$

唯一刻画. 其与分布函数间的关系为:

- $F(x) = \sum_{t \leq x} P(X = t) = \sum_{t \leq x} p(t)$
- $p(x) = P(X = x) = F(x) - F(x-)$

#### 定义 2.8 (概率密度函数)

当且仅当函数  $f(x)$  满足下述条件时, 被称为概率密度函数 (probability density function, p.d.f.):

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$



当  $X$  是连续型随机变量, 且  $F_X : \mathbb{R}^n \rightarrow [0, 1]$  可微 (或者更一般地, 绝对连续), 此时  $X$  的分布由

$$f_X := \frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}$$

唯一刻画. 其与分布函数间的关系为:

- $F(x) = \int_{-\infty}^x f(t) dt$
- $f(x) = \frac{d}{dx} F(x)$

**注** 即使对于  $f(x) > 0$  的  $x$ ,  $P(X = x) = x \int_x^x f(t) dt = 0$ , 即连续型随机变量在实轴上任意一点的概率测度为零. 概率密度函数  $f(x)$  代表的是在此位置上单位长度的概率, 可能是一个很大的值.

## 2.2 多元随机变量

#### 定义 2.9 (随机向量)

若随机变量  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  定义在同一概率空间  $(\Omega, \mathcal{F}, \mathbb{P})$  上, 则称

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

构成一个  $n$  维随机向量, 亦称  $n$  维随机变量.



#### 命题 2.3

若  $B_n$  为  $\mathbb{R}^n$  上任一博雷尔点集, 有

$$\{X(\omega) \in B_n\} \in \mathcal{F}$$



#### 定义 2.10

称  $n$  元函数

$$F(x_1, x_2, \dots, X_n) = \mathbb{P}\{X_1(\omega) < x_1, X_2(\omega) < x_2, \dots, X_n(\omega) < x_n\}$$

为随机向量  $X(\omega)$  的联合分布函数 (joint cdf).



当  $n = 2$  时, 有

$$\mathbb{P}((a_1, b_1) \leq X < (a_2, b_2)) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \quad (2.1)$$

**性质** 多元分布函数的一些性质:

1. 单调性: 关于每个变元是单调不减函数;

2.

$$F(x_1, x_2, \dots, -\infty, \dots, X_n) = 0$$

$$F(+\infty, +\infty, \dots, +\infty) = 1$$

3. 关于每个变元右连续.

4. 在二元场合, 还应该有: 对任意  $a_1 < b_1, a_2 < b_2$ , 都有

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0$$

为保证2.1式中的概率的非负性, 性质4是必须的, 而且由性质4可以推出单调性, 但存在着反例说明, 由单调性并不能保证性质4的成立(见习题12). 这是多元场合与一元场合的不同之处.

### 2.2.1 边际分布

#### 定义 2.11

对于多维随机变量  $X$ , 只考虑其中一个分量的分布时, 称其为  $X$  的 **边际分布或边缘分布**. 对于分量  $X_i$ , 其**边缘分布函数** (marginal cdf) 为:

$$F_{X_i}(x_i) = \mathbb{P}\{X_i \leq x_i\} = F(\infty, \dots, x_i, \dots, \infty)$$

### 2.2.2 独立

#### 定义 2.12 (独立随机变量)

若随机变量  $X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  联合分布函数可分解成各分量边缘分布函数的乘积, 即:

$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n), \quad \forall x_1, x_2, \dots, x_n \in \mathbb{R}$$

则称随机变量  $X$  各分量相互独立

**注** 对于一般的多元随机变量, 其各分量边缘分布不足以描述联合分布的情况. 但若其各分量独立则可以.

#### 定理 2.2

对于连续情况:

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n) \\ \Leftrightarrow f(x_1, x_2, \dots, x_n) &= f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) \end{aligned}$$

对于离散情况:

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n) \\ \Leftrightarrow p(x_1, x_2, \dots, x_n) &= p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n) \end{aligned}$$

#### 定理 2.3

若随机变量  $X, Y$  独立, 则其变换  $Z = g(X), W = h(Y)$  也独立.

泛化情况: 若随机向量  $\{X\}_n$  各分类独立, 则其变换  $\{Y\}_n = g(\{X\}_n)$  各分类也独立.

### 2.2.3 条件分布

#### 定义 2.13

对一切使  $P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{+\infty} p_{ij} > 0$  的  $y_j$ , 称

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \dots \quad (2.2)$$

为给定  $Y = y_j$  条件下  $X$  的条件分布列. 若  $p_X(x) = 0$ , 则定义其为 0.



设二维连续随机变量  $(X, Y)$  的联合密度函数为  $p(x, y)$ , 边际密度函数为  $p_X(x), p_Y(y)$ .

在离散随机变量场合, 其条件分布函数为  $P(X \leq x | Y = y)$ . 但是, 因为连续随机变量取某个值的概率为零, 即  $P(Y = y) = 0$ , 所以无法用条件概率直接计算  $P(X \leq x | Y = y)$ , 一个很自然的想法是: 将  $P(X \leq x | Y = y)$  看成是  $h \rightarrow 0$  时  $P(X \leq x | y \leq Y \leq y + h)$  的极限, 即

$$\begin{aligned} P(X \leq x | Y = y) &= \lim_{h \rightarrow 0} P(X \leq x | y \leq Y \leq y + h) \\ &= \lim_{h \rightarrow 0} \frac{P(X \leq x, y \leq Y \leq y + h)}{P(y \leq Y \leq y + h)} \\ &= \lim_{h \rightarrow 0} \frac{\int_{-\infty}^x \int_y^{y+h} p(u, v) dv du}{\int_y^{y+h} p_Y(v) dv} \\ &= \lim_{h \rightarrow 0} \frac{\int_{-\infty}^x \left\{ \frac{1}{h} \int_y^{y+h} p(u, v) dv \right\} du}{\frac{1}{h} \int_y^{y+h} p_Y(v) dv}. \end{aligned}$$

当  $p_Y(y), p(x, y)$  在  $y$  处连续时, 由积分中值定理可得

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \int_y^{y+h} p_Y(v) dv &= p_Y(y), \\ \lim_{h \rightarrow 0} \frac{1}{h} \int_y^{y+h} p(u, v) dv &= p(u, y). \end{aligned}$$

所以

$$P(X \leq x | Y = y) = \int_{-\infty}^x \frac{p(u, y)}{p_Y(y)} du.$$

至此, 我们可以定义连续随机变量的条件分布如下.

#### 定义 2.14

对一切使  $p_Y(y) > 0$  的  $y$ , 给定  $Y = y$  条件下  $X$  的条件分布函数和条件密度函数分别为

$$F(x|y) = \int_{-\infty}^x \frac{p(u, y)}{p_Y(y)} du, \quad (2.3)$$

$$p(x|y) = \frac{p(x, y)}{p_Y(y)}. \quad (2.4)$$

同理对一切使  $p_Y(y) > 0$  的  $x$ , 给定  $X = x$  条件下  $Y$  的条件分布函数和条件密度函数分别为

$$F(y|x) = \int_{-\infty}^y \frac{p(x, v)}{p_X(x)} dv \quad (2.5)$$

$$p(y|x) = \frac{p(x, y)}{p_X(x)}. \quad (2.6)$$



**注** 对于每一个固定的  $x$ ,  $p_{Y|X}(y|x)$  是一个关于  $y$  的概率质量函数;  $f_{Y|X}(y|x)$  是一个关于  $y$  的概率密度函数与概率三定理的对应:

乘法法则  $p_{XY}(x, y) = p_{Y|X}(y|x)p_X(x)$ ,  $f_{XY}(x, y) = f_{Y|X}(y|x)f_X(x)$

全概率公式  $p_Y(y) = \sum_x p_{Y|X}(y|x)p_X(x)$ ,  $f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx$

$$\text{Bayes 原理} \quad p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_x p_{Y|X}(y|x)p_X(x)}, \quad f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx}$$

## 2.3 随机变量的函数