概率论笔记

作者：肖程哲

时间：July 5, 2022

苟日新，日日新，又日新

目录

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 $\mathscr{F}$ 上, 故需 $X^{-1}(B) \in \mathscr{F}$.

定义 2.1 (可测性)

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

♣

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

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

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

♣

由于只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 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, $\mathscr{F}_{\mathcal{X}}$ 由 $\mathcal{X}$ 的所有子集组成, 则称其为**离散随机变量** (discrete random varible). 当随机变量 $\mathcal{X} = \mathbb{R}^n$, 考虑 $\mathscr{F}_{\mathcal{X}}$ 为 $\left\{\prod_{i=1}^n (-\infty, x_i] : x_1, \ldots, x_n \in \mathbb{R}\right\}$ 生成的 Borel 代数 (最小的 σ 代数), 则称其为**连续随机变量** (continuous random varible).

♣

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

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

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

单调性 $F(x_1)\le F(x_2),\quad \forall x_1<x_2$

右连续性 $\lim_{x\to x_0^+}F(x)=F(x_0)$

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

性质 $F(x)$ 最多只有可数个间断点

命题 2.1

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

命题 2.2

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

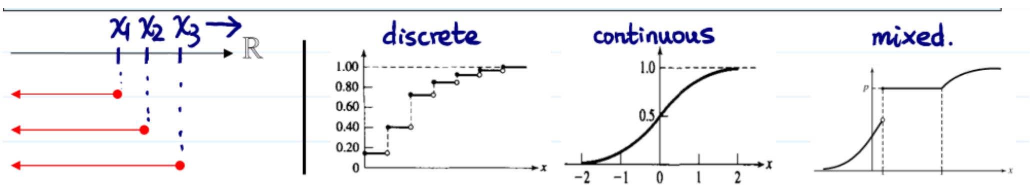
	离散	连续
一元随机变量	概率质量函数 (pmf)	概率密度函数 (pdf)
	累积分布函数 (cdf)	
	矩母函数/特征函数 (mgf/chf)	
多元随机变量	联合概率质量函数 (joint pmf)	联合概率密度函数 (joint pdf)
	联合累积分布函数 (joint cdf)	
	联合矩母函数/特征函数 (joint mgf/chf)	

定义 2.6 (累积分布函数)

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

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

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.):

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

当 X 是离散型随机变量, 设 $\mathscr{F}_{\mathcal{X}}$ 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:

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

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.):

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

当 X 是连续型随机变量, 且 $F_X : \mathbb{R}^n \to [0, 1]$ 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

记 $\sigma(X) := X^{-1}(\mathscr{F}_{\mathcal{X}}) := \{X^{-1}(B) : B \in \mathscr{F}_{\mathcal{X}}\}$. 根据前面的定义, X 可测当且仅当 $\sigma(X) \subset \mathscr{F}$. 若随机元 $X_1 : (\Omega, \mathscr{F}) \to (\mathcal{X}_1, \mathscr{F}_{\mathcal{X}_1})$ 与随机元 $X_2 : (\Omega, \mathscr{F}) \to (\mathcal{X}_2, \mathscr{F}_{\mathcal{X}_2})$ 满足 $\sigma(X_1) \perp\!\!\!\perp \sigma(X_2)$, 则称 X_1 与 X_2 **独立**, 记为 $X_1 \perp\!\!\!\perp X_2$.

注: 对任意可测映射 g_1, g_2 成立 $X_1 \perp\!\!\!\perp X_2 \implies g_1(X_1) \perp\!\!\!\perp g_2(X_2)$. 事实上, $\sigma(g(X)) \subset \sigma(X)$.

当 X_1 和 X_2 都是实值随机向量时, $X_1 \perp\!\!\!\perp X_2$ 当且仅当 (思考: $\sigma(X^{-1}(\bullet)) = X^{-1}(\sigma(\bullet))$, $\bullet \subset \mathscr{F}_{\mathcal{X}}$?)

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2), \ \forall x_1, x_2.$$

对于连续型随机向量, 刻画独立性只需要

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \ \forall x_1, x_2.$$

设随机向量 X 和 Y 有联合 (joint) 概率密度函数 $f_{X,Y}$, 可以证明 X 有概率密度函数

$$f_X(\cdot) = \int f_{X,Y}(\cdot, y),$$

称为 (X, Y) 中 X 的边际 (margin) 概率密度函数. Y **条件于** $X = x$ **的概率密度函数** $f_{Y|X}(\cdot|x)$ 满足

$$f_{X,Y}(x, \cdot) = f_{Y|X}(\cdot|x) f_X(x).$$

2.3 随机变量的函数