終曲

2022 年 11 月 26 日

目录

1 前言 **2**

2 第一章和第二章 **2**

2.1 概率和事件 2

2.2 古典概型 3

2.2.1 抽东西 3

2.2.2 平均分组 4

2.3 条件概率和 total probability formual and Bayes formula 4

2.3.1 一些分布 4

3 随机变量和一些概念 **4**

3.1 random variable 4

3.1.1 the definition of random variables 4

3.1.2 分布函数 5

3.1.3 分类 5

3.2 重要的分布函数 5

3.2.1 正态分布 and 标准化随机变量 5

3.2.2 指数分布 6

3.3 随机变量之间的运算 6

3.4 期望和方差 6

4 随机向量 **6**

4.1 好像也没什么好讲的 6

4.2 joint distribution function and marginal distribution 6

4.3 独立性和随机向量 6

4.4 二维正态分布 6

4.5 求解形如 $g(X,Y)$ 的分布 6

4.5.1 求解 r.v. 和的分布函数 6

4.5.2 离散的情况 7

4.5.3 连续的情况 7

4.6 $\min\{X,Y\}, \max\{X,Y\}$ 的求解 8

4.7 条件概率 in random vector 9

4.7.1 第一部分 9

4.7.2 另一个部分 9

5 结束 **10**

1 前言

这里是概率论最后一个讲义. 虽然说我根本没有写过多少个讲义, 因为我觉得那实在是太过好费时间了. 但是我现在不得不悲哀地意识到, 这个方法还是有效的, 并且来说, 和我在那里无谓地耗费时间相比起来, 要好得多.

这就是前言的前言, 我此刻有些百感交集, 但还是算了, 这份心情, 说出了也不能改变任何事物.

好的让我们回到正题, 什么是概率论? The probability theory. 这首先要从某种哲学观点来看 '概率' , 这里我们就不深究了. 虽然说早至古希腊的时候就有了概率这个词, 但是 "概率为 1 的事件一定发生" 这种观点你是什么时候知道的呢? 毕竟, 以前分数也不是那么明确清晰的概念啊. 没错, "概率为 1" 实际上是一个约定, 从马后炮来看, 这个观点是自然的, 是非常简洁的, 是非常高明的. 在欧洲的中世纪, 概率的概念才有发展. 这方面我建议查询网上的资料.

总之, 在两三百年间, 概率论初步建立. 主要贡献有:

1. 概率至多为 1
2. 概率可以定义为某些事件的数量除以另一些事件的数量
3. Bernoulli 大数定理
4. de-Moivre Laplace 定理
5. Bayes 不知道干嘛了

前两条可以说是我们的第一章以及第二章的内容. 值得注意的是, 某些严格化的定义, 比如说随机变量以及样本空间等等都是之后在定义出来的. 所以说当时 non-trivial 的问题在现在看来是 trivial 的, 这种现象也不足为奇了.

在突入第一章之前, 我们不妨描述一下这个临时讲义的 outline : 我们按照课本的顺序来编排内容. 第一章第二章讲的是古典概型和贝叶斯方法. 主要涉及我们高中水平的题目. 存在部分难题. 第三章涉及到随机变量之概念, 并且有随机变量的简单分类以及分布函数等概念, 还介绍了一些分布. 第四章便是多维随机变量, 又称随机向量, 并且介绍了求解 $f(X)$ 的分布的重要方法. 以及介绍了许多重要的概念比如说协方差和二维正态分布. 第五章是大数定理和中心极限定理, 这部分是我们教科书上最垃圾的部分之一. 我们只需要掌握 chebyshev 不等式的推导和应用, 中心极限定理的概念和应用. 证明什么的基本不用管. 第六章是几个微妙的分布. 第七章便是比较重要的矩估计法和极大似然估计法.

2 第一章和第二章

2.1 概率和事件

概率测度 P 是一个函数, 我们可以写为 $\mathbb{P}$, 也可以写为 $\mathscr{P}$, 也可以单纯写为 P. 我们沿用后者.

是一个什么函数呢? 是一个将事件射到 $[0,1]$ 上的一个函数. 具体定义如下:

$$P:\mathscr{F}\to[0,1]$$

其中 $\mathscr{F}$ 是事件的集合. 并且还满足了三条公理, 称为 Kolmogorov 公理. 这里进行一个罗列

$$P(\Omega)=1 \tag{1}$$
$$\forall A\in\mathscr{F},P(A)\geq 0 \tag{2}$$
$$A_i\cap A_j=\varnothing,P\Big(\bigcup_i A_i\Big)=\sum_i P(A_i) \tag{3}$$

这三条都有中文名称, 但是没有记忆的必要. 我们这种情况之下, P 具有非常良好的性质, 在我们计算的时候我们应当使用这些性质来简化我们的计算.

▷ **Example 1.** *我们要在 n 个球中, 一次拿取 m 个球, 如果说里面有 h 个白球, 求我们至少抽取到 1 个白球的概率.*

设这个所求的事件为 A, $\bar{A}$ 是一个白球都没有的情况. 那么我们有:

$$P(A)=1-P\left(\bar{A}\right)$$

接下来是事件, 一个事件定义为 Ω 里面的一个子集, $\mathscr{F}$ 的一个成员. 主要问题是什么呢, 我们可以使用事件的交和并, 来简要表明某些事件.

▷ **Example 2.** *我们有 18 个球, 里面有 5 个白球, 和 13 个黑球, 现在抽出 4 个球, 请你表示出 "至少有 1 个白球以及至少有 1 个黑球" 这个事件.*

设 A 为有白球, B 为有黑球, 那么事件表示为

$$\Omega-\bar{A}\cup\bar{B}$$

这里还有一点没有涉及到的, 就是根据概率测度来进行一些化简, 反正这些大家都很熟了.

▷ **Example 3.**

$$\forall A,B\in\mathscr{F},P(A\cup B)=P(A)+P(B)+P(A\cap B)$$

如果说 $A\cap B=\varnothing$ 那么这条等式退化为可加性公理.

2.2 古典概型

2.2.1 抽东西

我们从一堆球里面抽取几个球, 这就是属于古典概型的范畴 [1]. 我们这样考虑, 将 Ω 视为数量有限的集合. 里面存在一些基本事件, 即 Ω 的单点集, $\{\omega\}$, 我们常把 $\{\}$ 省略.

在这些模型里, $\forall\omega_i,P(\omega_i)$ 都是相等的, 那么主要想法就是:

$$P(A)=\frac{|A|}{|\Omega|} \tag{4}$$

就是数数!

[1] 此范畴非彼范畴

▷ **Example 4.** 我们要在 n 个球中, 一次拿取 m 个球, 如果说里面有 h 个白球, 求我们至少抽取到 1 个白球的概率.
设这个所求的事件为 A, $\bar{A}$ 是一个白球都没有的情况. 那么我们有:

$$P(A) = 1 - P(\bar{A})$$

并且对于 $\bar{A}$:

$$P(\bar{A}) = \frac{C_{n-h}^{m}}{C_n^m}$$

2.2.2 平均分组

简言之就是

▷ **Example 5.** 十二东西分为 6 组, 每组两个.

$$\frac{C_{12}^2 \cdot C_{10}^2 \cdots \cdot C_2^2}{A_6^6} \tag{5}$$

下面的排列是组数, 上面的显然.

2.3 条件概率和 total probability formual and Bayes formula

这些内容实际上很简单. 公式的推导也非常简单.
B_i is a partition of Ω

$$P(A \mid B) = \frac{P(AB)}{P(B)} \tag{条件概率}$$

$$P(A) = \sum_{i=1}^{n} P(A \mid B) P(B_i) \tag{total probability}$$

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)} \tag{Bayes formula}$$

怎么回事这个公式.

但是, 我们要注意到, 虽然这些等式是显然成立的, 但是, 这样的某种分块的行为对于我们求解某些概率是非常重要的. 比如说, 很多情况下, 我们确实是知道 $P(A \mid B_i)$ 的数值, 或者说这个值比较好处理. 那么我们还是要通过 total probability formula 来得到所求的答案.

Bayes 公式就算了, 随便考考就得了的东西.

2.3.1 一些分布

Bernoulli 分布 参数为 n , 进行 n 抛硬币实验的实验.

Poisson 分布和 Poisson 逼近 参数为 λ , 形式非常像 Taylor 展开式

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{6}$$

Poisson 逼近: 当 Bernoulli 实验中 $np = \lambda$ 这个值比较小的时候, 有:

$$C_n^k (1-p)^k p \approx \frac{\lambda^k}{k!} e^{-\lambda} \tag{7}$$

3 随机变量和一些概念

3.1 random variable

3.1.1 the definition of random variables

进行一个概念的背诵. random variable 是一个从样本空间射到实数轴上的一个可测函数. 具体定义如下:

$$X : \Omega \to \mathbb{R}, \Lambda \mapsto X(\Lambda)$$

这里我们定义实数轴上 Borel 集合, 这里略去不讲, 我们说可测函数的定义是, Borel 函数的原像, 仍然是在 $\mathscr{F}$ 中的, 剩下的我就不知道了. 这里严谨的定义实际上后面基本用不到, 唯一比较有用的就是下面这条

$$P(X \in A) \equiv P\left(X^{-1}(A)\right) \tag{8}$$

我们只需要将 $X \in A$ 看作是一个事件就行了. 这个 invs 在特定场合理解起来可能有用处, 但是大部分时间都没什么用处, 因为概率论的重点其实并不是在这方面. 重点其实在随机过程. 那 tm 才有意思.

3.1.2 分布函数

分布函数我们用的很多但是本身没什么好讲, 不就是那些东西吗, 还能玩出什么花. 分布函数定义为 blahblah. 密度函数定义为 blahblah. 但是值得注意的是 $F(x) = \int_{-\infty}^{x} f(x)\ \mathrm{d}x$, 这个后面用得到.

分布函数是一个定义良好的概念. 它具有很多良好的性质:

1. 单调增
2. $\lim_{x\to+\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$
3. 右连续这个需要严谨的证明比如说 $\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$ 在什么情况下成立?
4. 间断点可数 (1. 的推论).

3.1.3 分类

书上的方法实际上只是对分布函数进行分类. 我们这里先对随机变量进行分类.

随机变量的分类 分为两类: 简单函数, 和 elementary 函数. 这里不多赘述.

Definition 1 (simple function)**.** *if we can write X as the sum of some function in the form of 1_A then we call X is simple function.*

$$X = \sum_{i=1}^{n} b_i 1_{A_i}$$

分布函数的分类 两种分类法: 1. 分为离散函数和连续函数 2. 奇异函数和绝对连续函数. 我们只考虑离散函数和绝对连续函数就行了.

Definition 2 (绝对连续)**.** 如果说存在 $f(x)$ 使得

$$F(x) = \int_{-\infty}^{x} f(x)\ \mathrm{d}x$$

那么称呼 $F(x)$ 是绝对连续的.

3.2 重要的分布函数

3.2.1 正态分布 and 标准化随机变量

定义如下, 其密度函数为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)}{2\sigma^2}} \tag{9}$$

其中 μ 是期望, 然后 σ 是标准差. 这里出现的概念我们之后很快介绍. 总之这是一个很神奇的函数, 虽然我不记得这个东西是怎么导出的了. 但是我知道这个东西还在哪里出现. 在傅里叶变换中还会出现, $\mathscr{F}[f(t)] = F(\omega)$ 我们有 $F(\omega)$ 和 $f(t)$ 的图像的形状是相同的.

总之我们一定要记牢这个东西的定义.

3.2.2 指数分布

TODO

3.3 随机变量之间的运算

3.4 期望和方差

期望的计算有很多种

$$E \mid X = \int_{-\infty}^{+\infty} x f(x)\ \mathrm{d}x$$

或这是按照定义而来:

$$E \mid X = \int_{-\infty}^{\infty} x\ \mathrm{d}F(x)$$

或者是进行简单的积分次序的交换

$$E \mid X = \int_{-\infty}^{+\infty} P(X > x)\ \mathrm{d}x$$

面对 $g(X)$ 这样的随机变量, 我们能够将上面的 x 进行一个转变

$$E \mid g(X) = \int_{-\infty}^{+\infty} g(x)\ \mathrm{d}x$$

这样我们不需要求出 $g(X)$ 的分布函数, 在某些情况之下这时非常难求的.

4 随机向量

4.1 好像也没什么好讲的

4.2 joint distribution function and marginal distribution

4.3 独立性和随机向量

4.4 二维正态分布

4.5 求解形如 $g(X,Y)$ 的分布

4.5.1 求解 r.v. 和的分布函数

给定 r.v. X,Y, 求 $X+Y$ 的分布函数 F

面对这个问题, 理论上我们可以首先引入几个例子, 然后分为离散型 r.v. 和绝对连续型的 r.v. 进行讨论. 但其实我觉得并没有这些必要, 我们直接引入即可. 但是分类讨论在这种情况之下还是必须的, 这是因为求法不同. 一边是可以直接求和的方式求解, 和本节主要内容关系不大. 绝对连续型的, 就需要从密度函数出发.

4.5.2 离散的情况

对于离散型 r.v. 设 X,Y are discrete, 仅在自然数上有定义[2], 即 $P(X=j)\geq 0$ if $j\in\mathbb{N}$. 新的分布函数是 $F(z)=P(X+Y\leq z)$

有

$$F(z)=\sum_{i=0}^{z}P(X=i)P(Y=z-i)$$

, 这种就是书上介绍的方法, 因为我们平时接触的离散 r.v. 并没有那么奇葩 (比如说 $\sum_{i=1}^{\infty}a_i=1$, $\{b_i\}$ 是有理数的穷举, 我们有离散型 r.v. $\sum_{i=1}^{\infty}a_iI_{b_i}$ [3])

▷ **Example 6.** *X , Y 相互独立, 且是参数为 λ_1,λ_2 的泊松分布, 如果说泊松分布的记号是 P 就是说 $X\sim P(\lambda_1), Y\sim P(\lambda_2)$ 问 $X+Y$ 是啥子分布?*

我们套用上面的公式

$$\begin{aligned}
F(z)&=P(X+Y\leq z)\\
&=\sum_{i=0}^{z}P(X=i)P(Y=z-i)\\
&=\sum_{i=0}^{z}\frac{\lambda_1^i}{i!}e^{-\lambda_1}\frac{\lambda_2^{z-i}}{(z-i)!}e^{-\lambda_2}\\
&=\frac{1}{z!}e^{-(\lambda_1+\lambda_2)}\sum_{i=0}^{z}\frac{z!}{i!\,(z-i)!}\lambda_1^i\lambda_2^{z-i}\\
&=\frac{1}{z!}e^{-(\lambda_1+\lambda_2)}\sum_{i=0}^{z}C_z^i\lambda_1^i\lambda_2^{z-i}\\
&=\frac{(\lambda_1+\lambda_2)^z e^{-(\lambda_1+\lambda_2)}}{z!}
\end{aligned}$$

[2] 我们遇到的大部分离散 r.v. 都是这样的

[3] $I_{b_i}(x)=1$ if $x>b_i$, else $I_{b_i}(x)=0$

$F(z) \sim P(\lambda_1 + \lambda_2)$ 这就说明, 泊松分布具有一种线性可加性. 实际上对于伯努利分布也是如此.

4.5.3 连续的情况

首先还是从定义看:

$$F(z) = P(X + Y \leq z)$$

我们可以联想到 $g(X)$ 的求法. 假设 X 是绝对连续的, $g(X)$ 的分布函数 $F = \int_G f(x)\ \mathrm{d}x$, where $G = \{x : g(x) \leq z\}$

类似的, 对于随机变量之和的分布函数

$$F(z) = \iint_G f_{X,Y}(x, y)\ \mathrm{d}x\ \mathrm{d}y$$

, 其中 $G = \{(x, y) : x + y \leq z\}$ 当然这里能够直接扩展到一般的情况, 即 $G = \{(x, y) : g(x, y) \leq z\}$, 这点我们之后在介绍.

Lemma 1. 给定了两个随机变量 X, Y , 设 $f_{X,Y}(x, y)$ 是其密度函数, 则 $Z = X + Y$ 的密度函数 f_Z 等于

$$f(z) = \int_{-\infty}^{+\infty} f(x, z - x)\ \mathrm{d}x$$

$$f(z) = \int_{-\infty}^{+\infty} f(z - x, x)\ \mathrm{d}x$$

证明.

$$\begin{aligned} F(z) &= \iint_G f(x, y)\ \mathrm{d}x\ \mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f(x, y)\ \mathrm{d}y \right) \mathrm{d}x \end{aligned}$$

令 $y = u - x$ 就有

$$F(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z} f(x, u - x)\ \mathrm{d}u \right) \mathrm{d}x$$

交换积分次序:

$$F(z) = \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f(x, u - x)\ \mathrm{d}x \right) \mathrm{d}u$$

两边求导, 就能够得出答案了. □

Theorem 1. X, Y 是独立的, 那么 $X + Y$ 的密度函数是 $f(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x)\ \mathrm{d}x$

证明. 使用上面的引理

因为是独立的, 立刻能够得到结果. □

其中 f_X, f_Y 涉及的计算称为卷积, 记为 $f_X * f_Y$

▷ **Example 7.** 设 $X, Y \sim U[-a, a]$, 求 $Z = X + Y$ 的分布函数

▷ **Example 8.** 设 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 求 $Z = X + Y$ 的分布.

而后我们指出, 引理中使用的公式实际上可以直接用来求解问题, 当你涉及两个不独立的随机变量的时候可以这么做.

4.6 $\min\{X,Y\},\max\{X,Y\}$ 的求解

$X\vee Y$ 就是两者取最大值的意思, $X\wedge Y$ 是取最小值的意思.

我们可以将 $\vee,\wedge$ 看作是两个二元函数 (而实际上他们确实是二元运算符, 所以也没什么差别), 于是这就是求解 $g(X,Y)$ 的范畴. 我们可以很轻松地解出来, 见下面的推导.

我们可以直接推导出公式:

Theorem 2. $F_{x\vee y}(z)=F_x(z)F_y(z)$

证明.

$$\begin{aligned}F_{X\vee Y}(z)&=P(X\vee Y\le z)\\&=P(X\le z\quad\&\quad Y\le z)\\&=P(X\le z)\cdot P(Y\le z)\\&=F_X(z)\cdot F_Y(z)\quad\square\end{aligned}$$

对于 $X\wedge Y$ 也是类似的.

Theorem 3. $F_{X\wedge Y}=1-(1-F_X(z))(1-F_Y(z))$

证明.

$$\begin{aligned}F_{X\wedge Y}(z)&=1-P(X\wedge Y>z)\\&=1-P(X>z\quad\&\quad Y>z)\\&=1-(1-F_X(z))(1-F_Y(z))\quad\square\end{aligned}$$

4.7 条件概率 in random vector

4.7.1 第一部分

先前我们已经知道了条件概率的一种, 即 $P(A\mid B)$, conditioning an event on an event.

现在我们介绍 conditioning a r.v. on an event, 以及相关的概率密度之计算, 分布函数之计算. 离散型的先不介绍, 因为这个东西理应介绍过了.

Definition 3. $P(X\in B\mid A)=\int_B f_{X|A}(x)\ \mathrm{d}x$, 其中 A 是一个事件. 所定义的 $f_{X|A}$ 就是我们想要的条件概率密度.

实际上我们可以将事件 A 换成是 $\mathbb{R}$ 上的一个 borel 集, 这样的概率密度记为 $f_{X|\{X\in A\}}$, 我们为了避免麻烦, 还是确保 $P(A)>0$, 但是这个条件并不是必要的, 我们之后将会处理等于 0 的情况.

Definition 4. 我们给出密度的值

$$f_{X|X\in A}=\begin{cases}\frac{f(x)}{P(X\in A)} & ,x\in A\\ 0 & ,x\notin A\end{cases}$$

我们不难验证, $f_{X|X\in A}$ 确实是一个密度函数. 这需要我们验证其在 $\mathbb{R}$ 上的积分是否等于 1.

对此我们当然能够推广到随机向量上

Definition 5.

$$f_{X,Y|(X,Y)\in A}=\begin{cases}\frac{f(x,y)}{P((X,Y)\in A)} & ,(x,y)\in A\\ 0 & ,(x,y)\notin A\end{cases}$$

这里的 A 当然是 $\mathbb{R}^2$ 上的一个 *borel* 集.

另一方面, 我们能够给出全概率公式的另一个版本

Theorem 4. A_i 是 Ω 的划分, 那么

$$f_X(x) = \sum_i^n P(A_i) f_{X|A_i}(x)$$

因为这里的划分只是有限个, 当然可以对这式子两边的密度函数进行任意的积分, 可以得到原本的那个全概率公式. [4]

4.7.2 另一个部分

接下来是另一种条件概率, i.e. conditioning a r.v. on another r.v. 我们研究的是 $f_{X|Y}(x|y)$, 意思是给定 $Y = y$ 的时候, X 的条件概率密度. 在这里我们将会去处理前面所落下的, "条件的概率是 0" 的情况.

Definition 6.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

其中 $f_Y(y)$ 不为 0

▷ **Example 9.** 对于圆盘上的一个均匀分布的随机向量 (X, Y), 我们想要计算出 $f_{X|Y}(x|y)$, 首先得是算出 $f_Y(y)$, 我们将会看到, f_Y 并不是常数, 而 $f_{X|Y}$ 将会.

5 结束

虽然很突然, 但是这份讲义就到这里了.

[4] 划分 A_i 可以是无穷个吗?