概率论讨论班 (第五次)

陈昱坤

时间：2022 年 4 月 4 日

目录

1 一般定义

设 $\mathscr{B}$ 是实数域 $\mathbb{R}$ 上的 (线性)Borel 域, 在扩张实数域 $\mathbb{R}^* := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ 上我们定义扩张的 Borel 域 $\mathscr{B}^* := \sigma(\mathscr{B} \cup \{\{-\infty\}\} \cup \{\{+\infty\}\})$.

1.1 随机变量

定义 1.1. 给定概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$,$\Delta \in \mathscr{F}$, 称广义实值函数 $X : \Delta \to \mathbb{R}^*$ 为一个**广义实值随机变量**, 如果

$$\forall B \in \mathscr{B}^*[\{\omega : X(\omega) \in B\} \in \Delta \cap \mathscr{F}],$$

其中 $\Delta \cap \mathscr{F}$ 称为 Δ 在 $\mathscr{F}$ 中的**迹**.

一个**复值随机变量**是一个定义在 $\mathscr{F}$ 中的一个集合 Δ 上而在复平面中取值的函数, 其实部和虚部都是有限实值随机变量.

注 1.1. 在讨论基本性质时, 我们假设 $\Delta = \Omega$, 并且其以概率 1 取有限值. 在没有特别注明的情况下, 我们所说的随机变量都是指这种特殊的随机变量, 记为 r.v..

注 1.2. 设 $X : \Omega \to \mathbb{R}$, 它是一个随机变量当且仅当

$$\forall B \in \mathscr{B}, X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathscr{F},$$

换言之, 当且仅当可测集的原像可测. 我们也经常使用更简单的符号来表述这一结果:

$$X^{-1}(\mathscr{B}) \subset \mathscr{F}.$$

下面是集合论的一些结果:

命题 1.1. 设函数 $X : \Omega \to \mathbb{R}(\mathbb{R}^*)$,$A$ 是任意的指标集.$B_\alpha, B \subset \mathbb{R}(\mathbb{R}^*), \forall \alpha \in A$, 则有

$$X^{-1}(B^c) = (X^{-1}(B))^c$$
$$X^{-1}\left(\bigcap_{\alpha \in A} B_\alpha\right) = \bigcap_{\alpha \in A} X^{-1}(B_\alpha)$$
$$X^{-1}\left(\bigcup_{\alpha \in A} B_\alpha\right) = \bigcup_{\alpha \in A} X^{-1}(B_\alpha).$$

下面的定理给出了一个函数是否是随机变量的判据:

定理 1.1. X 是一个随机变量, 当且仅当对每个实数 x 或者对 $\mathbb{R}$ 的某个稠密子集中的所有 x, 有

$$\{\omega : X(\omega) \le x\} \in \mathscr{F}.$$

证明. 当 X 是随机变量时, 自然有

$$\{\omega : X(\omega) \le x\} \in \mathscr{F}, \forall x \in \mathbb{R}.$$

反之, 如果有上述结果成立, 即

$$X^{-1}((-\infty, x]) \in \mathscr{F}, \forall x \in \mathbb{R}.$$

我们考虑集族

$$\mathscr{C} := \{S : X^{-1}(S) \in \mathscr{F}\}.$$

那么 $\mathscr{C}$ 是一个 Borel 域, 并且包含所有形如 $(-\infty, x]$ 的集合. 又因为 $\mathscr{B}$ 是包含 $(-\infty, x], x \in \mathbb{R}$ 的最小的 Borel 域, 所以 $\mathscr{B} \subset \mathscr{C}$. 故

$$X^{-1}(\mathscr{B}) \subset X^{-1}(\mathscr{C}) \subset \mathscr{F}.$$

所以 X 是一个随机变量. □

注 1.3. 对于广义实值随机变量的情形, 只需考虑到

$$\{\omega : X(\omega) \le x\} = \{\omega : -\infty \le X(\omega) \le x\}$$

即可.

1.2 概率分布和分布函数

定义 1.2. 给定概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$, 以及随机变量 $X : \Omega \to \mathbb{R}$, 则 $\mathbb{P} \circ X^{-1}$ 定义了 $(\mathbb{R}, \mathscr{B})$ 上的一个概率测度, 称为随机变量 X 是**概率分布测度**. 定义

$$F(x) := (\mathbb{P} \circ X^{-1})((-\infty, x]) = \mathbb{P}\{X \le x\},$$

称为随机变量 X 的**分布函数**.

在上一讲中, 我们已经证明了如下结果

定理 1.2. $(\mathbb{R}, \mathscr{B})$ 上的概率测度与分布函数一一对应.

因此可以认为分布函数与概率分布测度所反映出来的随机变量 X 的信息是等价的. 我们自然地问: 这些信息能否完全刻画随机变量 X? 或者更精确地说, 给定一个概率分布测度或着一个分布函数, 是否能唯一确定概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$ 上的一个随机变量? 答案是否定的, 一个显然的反例如下:

例 1.1. 设 $\mathscr{U} = (0, 1], \mathscr{B}' := \mathscr{B} \cap \mathscr{U}, m$ 是 Lebesgue 测度, 则 $(\mathscr{U}, \mathscr{B}', m)$ 是概率空间. 在其上考虑函数 $X : [0, 1] \to \mathbb{R}, x \mapsto x, Y : [0, 1] \to \mathbb{R}, x \mapsto 1 - x$. 那么

$$F_X(y) = \mathbb{P}\{X \le y\} = y = \mathbb{P}\{Y \le y\} = F_Y(y).$$

因此 X, Y 具有相同的分布, 其共同的概率分布测度恰好为 Lebesgue 测度 m.

定义 1.3. 具有相同分布的随机变量称为是**同分布的**.

定义中所说的“同分布”显然定义了给定概率空间上随机变量全体的一个等价关系. 那么, 当我们给定了一个分布, 是否一定存在概率空间上的一个随机变量使其满足给定分布呢? 答案显然是否定的.

例 1.2. 考虑最简单的情形. 设 $\{\mathrm{pt}\}$ 是单点空间, 那么其上只有唯一一个 Borel 域 $\{\varnothing, \{\mathrm{pt}\}\}$. 在其上赋予测度 $\mathbb{P}(\varnothing) = 0, \mathbb{P}(\{\mathrm{pt}\}) = 1$, 则 $(\{\mathrm{pt}\}, \{\varnothing, \{\mathrm{pt}\}\}, \mathbb{P})$ 是一个概率空间. 其上的随机变量 $X : \{\mathrm{pt}\} \to \mathbb{R}$ 只能有分布函数

$$F(x) = \begin{cases} 1, x \geq X(\mathrm{pt}), \\ 0, x < X(\mathrm{pt}). \end{cases}$$

由此, 让我们来提一些问题:

- 有没有一个概率空间，使得任意的分布函数都能找到一个随机变量使得其满足给定的分布函数?
- 什么样的概率空间可以使得任意的分布函数都能找到一个随机变量使得其满足给定的分布函数?
- 给定一个分布函数，什么样的概率空间上不存在/存在随机变量满足这一分布函数?
- 有没有一个分布函数，如果这个概率空间上存在一个随机变量满足这一分布函数，那么对一系列的分布函数都存在相应的随机变量满足对应的分布函数?

1.3 从旧的随机变量构造新的随机变量

在本小节中我们将讨论构造随机变量的方法.

注意到一个可测空间上的实值函数是否是随机变量与该空间上赋予何种测度无关, 所以有如下定理:

定理 1.3. *设 X 是概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$ 上的 Borel 可测函数,f 是 $(\mathbb{R}, \mathscr{B})$ 上的随机变量. 那么 $f(X)$ 是一个随机变量.*

在初等概率论中, 我们经常遇到对于两个随机变量 X, Y, 考虑 $X+Y, X-Y$ 等情形. 为了能够一次性的给出这些函数是随机变量的论断, 我们考虑上述定理在**随机向量**情形的推广.

在 $\mathbb{R}^n$ 上的标准的 Borel 域 $\mathscr{B}^n$ 是由形如

$$\{(x_1, \cdots, x_n) : a_1 < x_1 \leq b_1, \cdots, a_n < x_n \leq b_n\}$$

的集合所构造成的集族所生成的 Borel 域. 或者, 等价地, 认为它是由形如

$$B_1 \times B_2 \times \cdots \times B_n := \{(x_1, \cdots, x_n) : x_i \in B_i\}, B_i \in \mathscr{B}$$

的集合所构成的集族所生成的 Borel 域. 现在考虑函数 $f : \mathbb{R}^n \to \mathbb{R}$, 如果有 $f^{-1}(\mathscr{B}) \subset \mathscr{B}^n$, 则称 f 是 **n 元 Borel 可测函数**.

现在, 让我们考虑概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$ 上的 n 个随机变量 $X_1, \cdots, X_n$, 我们形式地将其放在一起, 写为 $(X_1, \cdots, X_n)$, 称其为 **n 维随机向量**. 那么它诱导了 $(\mathbb{R}^n, \mathscr{B}^n)$ 上的一个概率测度, 其定义为

$$\nu(A) := \mathbb{P}\{(X_1, \cdots, X_n) \in A\},$$

其中,$\{(X_1, \cdots, X_n) \in A\} := \{\omega \in \Omega : (X_1(\omega), \cdots, X_n(\omega)) \in A\}$.

命题 1.2. 设 $(\Omega,\mathscr{F},\mathbb{P})$ 是一个概率空间,$f:\mathbb{R}^n\to\mathbb{R}$ 是 n 元 Borel 可测函数. 设 $(X_1,\cdots,X_n)$ 是概率空间 Ω 上的随机向量, 那么 $f(X_1,\cdots,X_n)$ 也是随机变量.

证明. 为了证明这一结果, 我们只要说明

$$(X_1,\cdots,X_n)^{-1}f^{-1}(\mathscr{B})\subset\mathscr{F}$$

即可. 因为 f 是 Borel 可测的, 所以有 $f^{-1}(\mathscr{B})\subset\mathscr{B}^n$, 因此, 我们只要证明

$$(X_1,\cdots,X_n)^{-1}(\mathscr{B}^n)\subset\mathscr{F}$$

即可. 为此, 我们只需证明 $\mathscr{B}^n$ 的生成元在 $(X_1,\cdots,X_n)^{-1}$ 的作用下都落在 $\mathscr{F}$ 之中即可.

设 $B=B_1\times B_2\times\cdots\times B_n\in\mathscr{B}^n$, 那么

$$(X_1,\cdots,X_n)^{-1}(B)=\bigcap_{k=1}^{n}X_k^{-1}(B_k)\in\mathscr{F},$$

因此对于 $\mathscr{B}^n$ 结论是对的. 又由命题1.1可知,$(X_1,\cdots,X_n)^{-1}(\mathscr{B}^n)\subset\mathscr{F}$. □

我们引入符号

$$X\wedge Y:=\max\{X,Y\},X\vee Y:=\min\{X,Y\}$$

注意到连续函数自然是 Borel 可测函数, 所以有

命题 1.3. 设 X,Y 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机变量, 那么 $X-Y,X+Y,X\cdot Y,X/Y,X\wedge Y,X\vee Y$ 都是随机变量.

下面我们考虑有一列随机变量的情形.

命题 1.4. 设 $\{X_n\}_{n=1}^{\infty}$ 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的一列随机变量, 那么

$$\inf_{n\in\mathbb{N}}X_n,\sup_{n\in\mathbb{N}}X_n,\limsup_{n\to\infty}X_n,\liminf_{n\to\infty}X_n$$

也是随机变量 (尽管未必以概率 1 取有限值).

证明. 证明第一个结论即可. 根据定理1.1, 我们只需证明对任意的 $x\in\mathbb{R}$, 有

$$\{\omega\in\Omega:\inf_{n\in\mathbb{N}}X_n\le x\}\in\mathscr{F},\forall x\in\mathbb{R}$$

即可. 根据定义, 有

$$\{\omega\in\Omega:\inf_{n\in\mathbb{N}}X_n\le x\}=\bigcup_{n\in\mathbb{R}}\{\omega\in\Omega:X_n(\omega)\le x\}\in\mathscr{F}.$$

故 $\inf_{n\in\mathbb{N}}X_n$ 可测. □

如果我们考虑

$$\Delta:=\{\omega\in\Omega:\limsup_{n\to\infty}X_n(\omega)=\liminf_{n\to\infty}X_n(\omega)\},$$

那么 $\lim\limits_{n\to\infty}X_n$ 就是 Δ 上的随机变量.

最后我们以一类随机变量来结束本节, 他们相当于实变函数论中的简单函数, 对于研究随机变量有着重要作用.

定义 1.4. 设 X 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机变量, 如果存在可数集 $B\subset\mathbb{R}$, 使得 $\mathbb{P}(X\in B)=1$, 则称 X 为**离散型随机变量**.

注 1.4. 离散型随机变量的分布函数是离散型分布函数.

注 1.5. 离散型随机变量的值域未必是离散的.

定义 1.5. 设 Ω 是一个集合,$\Delta \subset \Omega$, 定义

$$I_\Delta(x) = \begin{cases} 1, & x \in \Delta, \\ 0, & x \notin \Delta \end{cases}$$

称为 Δ 的**示性函数**.

设 Ω 上有有一个可数的分划 $\{\Lambda_j\}$, 即

$$\bigcup_j \Lambda_j = \Omega, \Lambda_i \cap \Lambda_j = \varnothing, i \neq j.$$

此时有

$$1 = I_\Omega = \sum_j I_{\Lambda_j}.$$

定义**加权划分** $\{\Lambda_j, b_j\}$, 其对应了一个随机变量

$$\varphi(\omega) = \sum_j b_j I_{\Lambda_j}(\omega), \forall \omega \in \Omega.$$

这是一个离散型随机变量. 对于任意的离散型随机变量, 假如 $\{b_j\}$ 恰为定义中的可数集 B, 那么考虑 $\Lambda_j := \{\omega : X(\omega) = b_j\}$, 则

$$X(\omega) = \sum_j b_j I_{\Lambda_j}(\omega), \text{a.s.} \omega \in \Omega.$$

如果值域 B 是有限值, 则称随机变量 X 是**简单的**. 这就是对应于实变函数论中的简单函数.

2 随机变量的期望和矩

2.1 期望 & 实变函数回顾

定义 2.1. 设 X 是概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$ 上的随机变量, 那么它的**期望**定义为

$$\mathbb{E}(X) := \int_\Omega X(\omega) \mathbb{P}(\mathrm{d}\omega).$$

有时候我们会记

$$\int_\Omega X \mathrm{d}\mathbb{P} := \int_\Omega X(\omega) \mathbb{P}(\mathrm{d}\omega).$$

对于 $\mathscr{F}$ 中的每个集合 Λ, 我们可以定义

$$\int_\Lambda X \mathrm{d}\mathbb{P} = \int_\Omega X \cdot I_\Lambda \mathrm{d}\mathbb{P}.$$

由于概率空间是有限的测度空间, 所以其具有一些一般测度空间不具有的性质. 下面是关于

定理 2.1. *(1) 设 X 是一个非负随机变量, 则存在一列递增的简单 (离散) 随机变量 $\{X_n\}$, 使得*

$$\lim_{n \to \infty} X_n(\omega) = X(\omega), \forall \omega \in \Omega,$$

且

$$\lim_{n\to\infty}\mathbb{E}(X_n)=\mathbb{E}(X).$$

(2) 可积等价于绝对可积.

(3) 积分是线性的.

(4)(对于集的可加性). 如果 Λ_n 互不相交, 则

$$\int_{\bigcup_n \Lambda_n} X\mathrm{d}\mathbb{P}=\sum_n\int_{\Lambda_n} X\mathrm{d}\mathbb{P}.$$

(5)(积分是正线性泛函). 如果在 Λ 上,$X\geq 0$ 几乎处处成立, 则

$$\int_\Lambda X\mathrm{d}\mathbb{P}\geq 0.$$

(6)(单调性). 如果在 Λ 上,$X_1\leq X\leq X_2$ 几乎处处成立, 那么有

$$\int_\Lambda X_1\mathrm{d}\mathbb{P}\leq\int_\Lambda X\mathrm{d}\mathbb{P}\leq\int_\Lambda X_2\mathrm{d}\mathbb{P}.$$

(7)(中值定理). 如果在 Λ 上,$a\leq X\leq b$ 几乎处处成立, 则

$$a\mathbb{P}(\Lambda)\leq\int_\Lambda X\mathrm{d}\mathbb{P}\leq b\mathbb{P}(\Lambda).$$

(8)(模不等式).

$$\left|\int_\Lambda X\mathrm{d}\mathbb{P}\right|\leq\int_\Lambda |X|\mathrm{d}\mathbb{P}.$$

*(9)(控制收敛定理). 如果 $\lim\limits_{n\to\infty}X_n=X$ 在 Λ 上几乎处处成立或者***依测度***成立, 并且 $|X_n|\leq Y$ 在 Λ 上几乎处处成立,$\int_\Lambda Y\mathrm{d}\mathbb{P}<\infty$, 则*

$$\lim_{n\to\infty}\int_\Lambda X_n\mathrm{d}\mathbb{P}=\int_\Lambda X\mathrm{d}\mathbb{P}=\int_\Lambda\lim_{n\to\infty}X_n\mathrm{d}\mathbb{P}.$$

(10)(有界收敛定理). 如果 $\lim\limits_{n\to\infty}X_n=X$ 在 Λ 上几乎处处成立或者依测度成立, 且存在常数 M, 使得 $|X_n|\leq M,\forall n$. 则

$$\lim_{n\to\infty}\int_\Lambda X_n\mathrm{d}\mathbb{P}=\int_\Lambda X\mathrm{d}\mathbb{P}=\int_\Lambda\lim_{n\to\infty}X_n\mathrm{d}\mathbb{P}.$$

(11)(单调收敛定理). 设 X_n 是非负随机变量 (几乎处处意义下), 如果 $X_n\uparrow X$ 几乎处处成立, 则

$$\lim_{n\to\infty}\int_\Lambda X_n\mathrm{d}\mathbb{P}=\int_\Lambda X\mathrm{d}\mathbb{P}=\int_\Lambda\lim_{n\to\infty}X_n\mathrm{d}\mathbb{P}.$$

(12)(积分与求和换序). 如果

$$\sum_n\int_\Lambda |X_n|\mathrm{d}\mathbb{P}<\infty,$$

则 $\sum_n|X_n|<\infty$, 所以 $\sum_n X_n$ 几乎处处收敛, 且有

$$\int_\Lambda\sum_n X_n\mathrm{d}\mathbb{P}=\sum_n\int_\Lambda X_n\mathrm{d}\mathbb{P}.$$

(13)(Fatou 引理). 设 X_n 是非负随机变量, 则

$$\int_\Lambda\liminf_{n\to\infty}X_n\mathrm{d}\mathbb{P}\leq\liminf_{n\to\infty}\int_\Lambda X_n\mathrm{d}\mathbb{P}.$$

下面我给出一个有启发性的结果

定理 2.2. 设 X 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机变量, 则

$$\sum_{n=1}^{\infty}\mathbb{P}\{|X|\geq n\}\leq\mathbb{E}(|X|)\leq 1+\sum_{n=1}^{\infty}\mathbb{P}\{|X|\geq n\}.$$

证明. 定义 $\Lambda_n:=\{\omega:n\leq|X|<n+1\}$, 则 $\bigcup_n\Lambda_n=\Omega$, 故

$$\mathbb{E}(|X|)=\sum_n\int_{\Lambda_n}|X|\mathrm{d}\mathbb{P}.$$

所以

$$\begin{aligned}\sum_{n=1}^{\infty}n\mathbb{P}\{n+1>|X|\geq n\}\leq\mathbb{E}(|X|)&\leq\sum_{n=0}^{\infty}(n+1)\mathbb{P}\{n+1>|x|\geq n\}\\&=\sum_{n=0}^{\infty}\mathbb{P}\{n+1>|X|\geq n\}+\sum_{n=1}^{\infty}n\mathbb{P}\{n+1>|X|\geq n\}\\&=1+\sum_{n=1}^{\infty}n\mathbb{P}\{n+1>|X|\geq n\}.\end{aligned}$$

运用 Abel 求和公式, 有

$$\begin{aligned}\sum_{n=1}^{N}n\mathbb{P}\{n+1>|X|\geq n\}&=\sum_{n=1}^{N}n(\mathbb{P}\{|X|\geq n\}-\mathbb{P}\{|X|\geq n+1\})\\&=\sum_{n=1}^{N}\mathbb{P}\{|x|\geq n\}-N\mathbb{P}\{|X|\geq N+1\}\end{aligned}$$

于是有

$$\sum_{n=1}^{N}\mathbb{P}\{|x|\geq n\}-N\mathbb{P}\{|X|\geq N+1\}=\sum_{n=1}^{N}n\mathbb{P}\{n+1>|X|\geq n\}\leq\sum_{n=1}^{N}\mathbb{P}\{|x|\geq n\}$$

如果 $\mathbb{E}(|X|)=\infty$, 那么所要证的第一个不等号自然成立; 对于第二个不等号只需注意到有上面的不等式成立即可.

如果 $\mathbb{E}(|X|)<\infty$, 那么

$$0\leq N\mathbb{P}\{|X|\geq N+1\}\leq\int_{|X|\geq N+1}|X|\mathrm{d}\mathbb{P}.$$

又因为

$$\lim_{N\to\infty}\int_{|X|\geq N+1}|X|\mathrm{d}\mathbb{P}=0,$$

所以

$$\sum_{n=1}^{\infty}n\mathbb{P}\{n+1>|X|\geq n\}=\sum_{n=1}^{\infty}\mathbb{P}\{|x|\geq n\}.$$

所以

$$\sum_{n=1}^{\infty}\mathbb{P}\{|X|\geq n\}\leq\mathbb{E}(|X|)\leq 1+\sum_{n=1}^{\infty}\mathbb{P}\{|X|\geq n\}.$$

□

推论 2.1. 特别的, 如果 X 只取整数值, 则

$$\mathbb{E}(|X|)=\sum_{n=1}^{\infty}\mathbb{P}\{|X|=n\}.$$

2.2 一个公式

本小节的目的是建立起一个运用随机变量的概率分布测度 $\mathbb{P}\circ X^{-1}$ 来重写的积分公式. 定理的表述如下:

定理 2.3. 设 X 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机变量,f 是一个 *Borel* 可测函数, 那么随机变量 $f(X)$ 的期望具有如下的计算公式:

$$\mathbb{E}(f(X))=\int_\Omega f(X)\mathrm{d}\mathbb{P}=\int_\mathbb{R} f\mathrm{d}(\mathbb{P}\circ X^{-1})=\int_{-\infty}^{+\infty} f(x)\mathrm{d}F_X(x).$$

证明. 我们运用典型方法证明. 设 $B\in\mathscr{B}$,I_B 是示性函数, 那么

$$\int_\Omega I_B(X)\mathrm{d}\mathbb{P}=\int_{X\in B}\mathrm{d}\mathbb{P}=\mathbb{P}\{X\in B\}=\mathbb{P}(X^{-1}(B))=\int_\mathbb{R} I_B\mathrm{d}(\mathbb{P}\circ X^{-1}).$$

设有 $(\mathbb{R},\mathscr{B},\mathbb{P})$ 上的带权划分 $\{\Lambda_j,n_j\}$, 则可以定义 Borel 可测函数 (同时也是离散型随机变量)

$$f(x)=\sum_j b_j I_{\Lambda_j}(x),\forall x\in\mathbb{R}.$$

此时,

$$\begin{aligned}\int_\Omega f(X)\mathrm{d}\mathbb{P}&=\int_\Omega\sum_j b_j I_{\Lambda_j}(x)\mathrm{d}\mathbb{P}=\sum_j b_j\int_\Omega I_{\Lambda_j}\mathrm{d}\mathbb{P}\\&=\sum_j\int_\mathbb{R} b_j I_{\Lambda_j}\mathrm{d}(\mathbb{P}\circ X^{-1})=\int_\mathbb{R} f\mathrm{d}(\mathbb{P}\circ X^{-1}).\end{aligned}$$

现在, 考虑任意的正 Borel 可测函数 f, 令 $\{f_m\}$ 是递增的离散型随机变量列, 且 $f_m\to f$. 根据单调收敛定理, 有

$$\int_\Omega f(X)\mathrm{d}\mathbb{P}=\lim_{m\to\infty}\int_\Omega f_m(X)\mathrm{d}\mathbb{P}=\lim_{m\to\infty}\int_\mathbb{R} f_m\mathrm{d}(\mathbb{P}\circ X^{-1})=\int_\mathbb{R} f\mathrm{d}(\mathbb{P}\circ X^{-1}).$$

最后考虑 $f=f^+-f^-$ 即可.

□

对于 n 元 Borel 函数, 也有类似的结论成立, 为了方便, 我们记

$$\mu:=\mathbb{P}\circ(X_1,\cdots,X_n)^{-1}.$$

定理 2.4. 设 $(X_1,\cdots,X_n)$ 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机向量,f 是一个 n 元 *Borel* 可测函数, 那么随机变量 $f(X)$ 的期望具有如下的计算公式:

$$\mathbb{E}(f(X))=\int_\Omega f(X)\mathrm{d}\mathbb{P}=\int_\mathbb{R} f\mathrm{d}\mu.$$

2.3 矩和常用不等式

定义 2.2. 设 X 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机变量,$a\in\mathbb{R},r>0$, 称 $\mathbb{E}(|X-a|^r)$ 为 X 的**以 a 为中心的 r 阶矩**. 如果取 $a=\mathbb{E}(X)$, 则称此时对应的矩为**中心矩**. 当 $r=2$ 时, 对应的中心矩称为**方差**, 记为 $\mathrm{Var}(X)$; 其正 (非负) 平方根称为**标准差**.

推论 2.2. 设 X 是概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 上的随机变量, 则

$$\mathbb{E}(|X-a|^r)=\int_{-\infty}^{\infty}|x-a|^r\mathrm{d}F_X(x).$$

注 2.1. 这一个计算公式表明了我们所研究的矩只与随机变量的分布函数有关. 注意到随机变量与分布函数并不是一一对应的, 由此可以想见, 即使知道了所有的矩量, 也不能确定一个随机变量.

推论 2.3. $\mathrm{Var}(X) \le \mathbb{E}(X^2)$.

证明. 根据定义, 有

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(|X-\mathbb{E}(X)|^2)\\ &= \mathbb{E}(|X|^2 - 2X\mathbb{E}(X) + |\mathbb{E}(X)|^2)\\ &= \mathbb{E}(|X|^2) - (\mathbb{E}(X))^2 \le \mathbb{E}(|X|^2).\end{aligned}$$

□

定理 2.5 (Chebyshev 不等式)**.** 设 φ 是定义在 $(0,+\infty)$ 上的严格正值增函数. 那么对于随机变量 X, 如果 $\mathbb{E}(\varphi(X)) < \infty$, 则有

$$\mathbb{P}\{|X| \ge u\} \le \frac{\mathbb{E}(\varphi(X))}{\varphi(u)}$$

证明. 直接估计即可:

$$\mathbb{E}(\varphi(X)) \ge \int_{|X|\ge u} \varphi(X)\mathrm{d}\mathbb{P} \ge \varphi(u)\mathbb{P}\{|X| \ge u\}.$$

□

这一不等式给出了截断后尾项大小的估计.

下面, 我们给出一系列 L^p 空间中的不等式, 为此我们需要先证明 Jensen 不等式.

引理 2.1 (凸函数的支撑直线)**.** 设 $\Phi(u)$ 是 $[\alpha,\beta]$ 上的下凸函数, 那么对任意的 $\gamma \in (\alpha,\beta)$,$\Phi(\gamma)$ 的左右导数存在. 并且对于介于左导数 $\Phi'_-(\gamma)$ 和 $\Phi'_+(\gamma)$ 的实数 k, 有

$$\Phi(u) - \Phi(\gamma) \ge k(u-\gamma), \forall u \in [\alpha,\beta].$$

并称直线 $\ell : u \mapsto \Phi(\gamma) + k(u-\gamma)$ 为 Φ 在点 $(\gamma,\Phi(\gamma))$ 的**支撑直线**.

证明. 根据凸函数的定义, 对于任意的 $\gamma_1 > \gamma_2$, 有

$$\frac{\Phi(\gamma_1)-\Phi(\gamma)}{\gamma_1-\gamma} > \frac{\Phi(\gamma_2)-\Phi(\gamma)}{\gamma_2-\gamma}.$$

所以函数

$$\varphi(\eta) = \frac{\Phi(\eta)-\Phi(\gamma)}{\eta-\gamma}, \eta \in [\alpha,\beta] \setminus \{\gamma\}$$

是单调递增函数, 因此, 它在 γ 处的左右极限存在, 因此 Φ 在点 γ 的左右导数存在, 并且有, 对于任意的 $\gamma_2 < \gamma$ 和 $\gamma_1 > \gamma$,

$$\frac{\Phi(\gamma_2)-\Phi(\gamma)}{\gamma_2-\gamma} \le \Phi'_-(\gamma) \le \Phi'_+(\gamma) \le \frac{\Phi(\gamma_1)-\Phi(\gamma)}{\gamma_1-\gamma}.$$

所以, 对于 $k \in \mathbb{R}$, 满足 $\Phi'_-(\gamma) \le k \le \Phi_+(\gamma)$, 有

$$\frac{\Phi(\gamma_2)-\Phi(\gamma)}{\gamma_2-\gamma} \le k \le \frac{\Phi(\gamma_1)-\Phi(\gamma)}{\gamma_1-\gamma}.$$

因此,

$$\Phi(u) - \Phi(\gamma) \ge k(u-\gamma), \forall u \in [\alpha,\beta].$$

□

定理 2.6 (带权 Jensen 不等式)**.** 设 $\Phi(u)$ 是 $[\alpha,\beta]$ 上的下凸函数,$f:[a,b]\to[\alpha,\beta]$, 非负可积函数 $p(x):[a,b]\to\mathbb{R}$ 满足

$$P=\int_a^b p(x)\mathrm{d}x>0.$$

那么有

$$\Phi(\frac{1}{P}\int_a^b f(x)p(x)\mathrm{d}x)\le\frac{1}{P}\int_a^b\Phi(f(x))p(x)\mathrm{d}x.$$

证明. 设

$$\gamma=\frac{1}{P}\int_a^b f(x)p(x)\mathrm{d}x,$$

则 $\gamma\in[\alpha,\beta]$, 当 γ 取端点值时, 不等式显然成立 ($\gamma=\beta$ 时显然取等). 假设 $\gamma\in(\alpha,\beta)$, 并令 $(\gamma,\Phi(\gamma))$ 处的支撑直线斜率为 k, 则有

$$\Phi(u)-\Phi(\gamma)\ge k(u-\gamma).$$

将 $u=f(x)$ 代入, 得

$$\Phi((f(x)))p(x)-\Phi(\gamma)p(x)\ge kp(x)(f(x)-\gamma).$$

然后在 $[a,b]$ 上积分, 有

$$\int_a^b\Phi((f(x)))p(x)\mathrm{d}x-\Phi(\frac{1}{P}\int_a^b f(x)p(x)\mathrm{d}x)\ge 0,$$

所以

$$\Phi(\frac{1}{P}\int_a^b f(x)p(x)\mathrm{d}x)\le\frac{1}{P}\int_a^b\Phi(f(x))p(x)\mathrm{d}x..$$

□

注 2.2. 注意到证明过程中并没有用到 Φ 定义在有限区间这件事, 因此对于 $\mathbb{R}$ 上的下凸函数, 不等式也成立. 同样的, 也没有依赖于在哪个空间积分这个事情, 所以对任意测度空间上, 带权 Jensen 不等式也成立.

定理 2.7 (Lyapunov 不等式)**.**

$$\mathbb{E}(|X|^r)^{1/r}\le\mathbb{E}(|X|^{r'})^{1/r'},1<r<r'<\infty.$$

证明. 我们证明

$$\mathbb{E}(|X|^r)^{r'/r}\le\mathbb{E}(|X|^{r'})=\mathbb{E}(|X|^{r\cdot(r'/r)}).$$

注意到函数 $\Phi(x)=x^{r'/r}$ 是下凸函数, 运用不带权的 Jensen 不等式 (此时需要测度空间为概率空间) 即可. □

定理 2.8 (Hölder 不等式)**.** 设 $p,q>0,1/p+1/q=1$, 则对于随机变量 X,Y, 有

$$\mathbb{E}(|XY|)\le(\mathbb{E}(|X|^p))^{1/p}(\mathbb{E}(|Y|^q))^{1/q}.$$

证明. 可以将这个不等式看成带权的 Lyapunov 不等式.

我们用带权的 Jensen 不等式证明. 如果 $|X|,|Y|$ 中有一个期望为 0, 那么不等式两边都是 0, 结论成立. 现在, 假设 $|X|,|Y|$ 的期望都大于 0, 如果 $\mathbb{E}(|X|^p)$ 与 $\mathbb{E}(|Y|^q)$ 中有一个取到无穷, 那么不等式也显然成立.

所以假设,$|X|,|Y|$ 的期望都大于 0, 且 $\mathbb{E}(|X|^p)$ 与 $\mathbb{E}(|Y|^q)$ 都是有限的. 我们以函数 $p(\omega) = |X(\omega)|^p$ 为权函数, 则

$$P = ((\mathbb{E}(|X|^p))^{1/p})^p.$$

设 $f(\omega) = |Y(\omega)||X(\omega)|^{-p/q}$,$\Phi(x) = x^q$, 所以

$$\mathbb{E}(|XY|) = \int_\Omega f(\omega)p(\omega)\mathrm{d}\mathbb{P}(\omega) \leq P \cdot P^{-1/q} \left(\int_\Omega |Y(\omega)|^q|X(\omega)|^{-p}|X(\omega)|^p \mathrm{d}\mathbb{P}(\omega)\right).$$

此即所要证的不等式. □

定理 2.9 (Minkowski 不等式)**.** 设 $1 \leq p < \infty$, 则

$$(\mathbb{E}(|X+Y|^p))^{1/p} \leq (\mathbb{E}(|X|)^p)^{1/p} + (\mathbb{E}(|Y|)^p)^{1/p}.$$

证明. 这也可以用带权的 Jensen 不等式证明, 但过程十分麻烦. 我们直接使用 Hölder 不等式证明.

$p = 1$ 时显然, 不妨设 $p > 1$.

$$\begin{aligned}
\mathbb{E}(|X+Y|^p) &= \mathbb{E}(|X+Y|^{p-1}|X+Y|) \\
&\leq \mathbb{E}(|X+Y|^{p-1}|X|) + \mathbb{E}(|X+Y|^{p-1}|Y|) \\
&\leq \mathbb{E}(|X+Y|^{(p-1)p'})^{1/p'}\mathbb{E}(|X|^p)^{1/p} + \mathbb{E}(|X+Y|^{(p-1)p'})^{1/p'}\mathbb{E}(|Y|^p)^{1/p} \\
&\leq \mathbb{E}(|X+Y|^p)^{1-1/p}\left((\mathbb{E}(|X|)^p)^{1/p} + (\mathbb{E}(|Y|)^p)^{1/p}\right)
\end{aligned}$$

所以

$$(\mathbb{E}(|X+Y|^p))^{1/p} \leq (\mathbb{E}(|X|)^p)^{1/p} + (\mathbb{E}(|Y|)^p)^{1/p}.$$

□