概率论

孙天阳

中国科学技术大学数学科学学院

tysun@mail.ustc.edu.cn

2024 年 11 月 17 日

目录

Chapter 1

概率空间与独立性

1 条件概率

2 条件概率习题

Chapter 2

随机变量与分布函数

1 随机变量

定义 1.1. 设 $(\Omega, \mathcal{F}, P)$ 是一个概率空间, 称一个可测函数 $X\colon \Omega \to \mathbb{R}$ 为一个随机变量.

考虑随机变量有如下好处: 首先它将实验结果数值化, 使我们能够用数学工具来分析随机现象. 其次, 相较于弄清楚 Ω 中有哪些可测集, 每个可测集的测度是多少, 随机变量可以帮助我们聚焦在自己关心的事情上. 比如考虑一个只取值 0 和 1 的随机变量 X, 其实也就是说我们只关心 $X = 0$ 发生的概率与 $X = 1$ 发生的概率, 而不在乎其背后的概率空间 Ω 具体长什么样. 用严格的数学语言来说,

定义 1.2. 给定可测映射 $f\colon (X_1, \Sigma_1) \to (X_2, \Sigma_2)$ 和测度 $\mu : \Sigma_1 \to [0, +\infty]$. 定义推出测度 $f_*\mu$ 为

$$f_*\mu(B) = \mu(f^{-1}(B)), \quad \forall B \in \Sigma_2.$$

把这套语言用在概率空间和随机变量上, 也就是说我们可以通过随机变量 X 将概率空间上的测度 P 推出到 $\mathbb{R}$ 上, 然后研究推出测度 f_*P. 接下来的很多理论上的讨论只需从 f_*P 出发就可以了.

2 分布函数

命题 2.1. 设 $F(x)$ 是分布函数，则

(1) $F(x)$ 单调递增.

(2) $\lim\limits_{x\to-\infty} F(x)=0, \lim\limits_{x\to+\infty} F(x)=1.$

(3) $F(x)$ 右连续.

证明. □

注记.

(1) 可以证明，满足上述三条的函数必为某概率空间上某随机变量的分布函数

(2) $F(x)$ 单调递增的充分条件是 F 几乎处处可导且 $F'>0$.

3 随机向量

定义 3.1.

(1) 设 $X_1, \cdots, X_n$ 为 $(\Omega, \mathcal{F}, P)$ 上的随机变量，称 $\vec{X} = (X_1, \cdots, X_n)$ 为 n 维随机向量，称 n 元函数 $F(x_1, \cdots, x_n) = P(X_1 \leqslant x_1, \cdots, X_n \leqslant x_n)$ 为 $\vec{X}$ 的联合分布函数.

(2) 若 $\vec{X}$ 只取 $\mathbb{R}^n$ 中的可列个点,则称 $\vec{X}$ 是离散型随机向量,称 n 元函数 $f(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n)$ 为联合分布列.

(3) 若存在非负可积函数 $f(x_1, \cdots, x_n)$ 使得

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \cdots, u_n) \mathrm{d}u_1 \cdots \mathrm{d}u_n,$$

则称 $\vec{X}$ 为连续型随机向量，称 f 为联合密度函数.

注记. 任意两个随机变量都有联合分布函数，但即使是两个连续型随机变量，组合在一起也不一定是连续型随机向量.

4 随机向量习题

2.5.6 函数

$$F(x,y)=1-\mathrm{e}^{-xy},\quad 0\leqslant x,y<\infty$$

是某对随机变量的联合分布函数吗？

证明. □

Chapter 3

离散型随机变量

- 当你讨论两个随机变量的互动，比如 $X+Y$，XY，你是必须要知道 X 和 Y 在概率空间 Ω 上的行为的，或者至少你得知道 X 与 Y 是独立的.

1 分布列与独立性

例 1.1. 二项分布

几何分布称随机变量 X 服从以 $p\in(0,1)$ 为参数的几何分布，是指 X 的分布列满足

$$f(k)=P(X=k)=q^{k-1}p,\quad k=1,2,\cdots$$

称随机变量 X 服从以 $\lambda>0$ 为参数的 $Poisson$ 分布，是指 X 的分布列满足

$$f(k)=P(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad k=0,1,2,\cdots$$

记作 $X\sim P(\lambda)$. 若 $X\sim P(\lambda)$，容易算得 $E[X]=\lambda$.

定义 1.1. 称离散型随机变量 X 和 Y 独立，若 $\forall x,y\in\mathbb{R}$，有

$$P(X=x,Y=y)=P(X=x)P(Y=y).$$

称 $X_1,\cdots,X_n$ 相互独立，若任意 $x_i\in\mathbb{R}$,

$$P(X_1=x_1,\cdots,X_n=x_n)=P(X_1=x_1)\cdots P(X_n=x_n).$$

更一般地，称可列个随机变量 $\{X_n\}$ 相互独立，是指其中任意有限多个相互独立.

2 例子

3 期望

- 期望实际上只跟 X 推出的 $\mathbb{R}$ 上的测度有关，X 的期望是 $f(x) = x$ 这个函数在测度 X_*P 下的积分.

- $g(X)$ 是期望是 $f(x) = g(x)$ 这个函数在测度 X_*P 下的积分.

定理 3.1. 期望算子 E 满足

(1) 非负性. 当 $X \geqslant 0$ 时，$E[X] \geqslant 0$.

(2) 归一性. $E[1] = 1$.

(3) 线性性. $E[aX + bY] = aE[X] + bE[Y]$.

命题 3.1. 设 X 与 Y 期望均存在. 若 X 与 Y 独立，则

$$E[XY] = E[X]E[Y].$$

4 协方差

5 依赖性

定义 5.1. 联合分布与联合分布列

引理 5.1. 离散随机变量 X 和 Y 是独立的当且仅当

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall x, y \in \mathbb{R}.$$

更一般地, X 和 Y 是独立的当且仅当 $f_{X,Y}$ 能够被分解为分别只依赖于 x 与 y 的函数 $g(x)$ 与 $h(y)$ 的乘积.

6 依赖性习题

3.6.8 设 X 和 Y 有联合质量分布函数

$$f(j,k)=\frac{c(j+k)a^{j+k}}{j!k!},\quad j,k\geqslant 0,$$

其中 a 是一个常数，求 $c, P(X=j), P(X+Y=r)$ 和 $E[X]$.

证明.

$$\begin{aligned}
\sum_{j,k\geqslant 0}\frac{(j+k)a^{j+k}}{j!k!} &= \sum_{n=0}^{\infty}\sum_{j=0}^{n}\frac{na^n}{j!(n-j)!}\\
&= \sum_{n=0}^{\infty}\frac{na^n}{n!}\sum_{j=0}^{n}\frac{n!}{j!(n-j)!}\\
&= \sum_{n=0}^{\infty}\frac{n(2a)^n}{n!}\\
&= 2a\sum_{n=0}^{\infty}\frac{(2a)^n}{n!}\\
&= 2a\mathrm{e}^{2a}\\
P(X=j) &= \frac{1}{2a\mathrm{e}^{a}}\frac{a^j}{j!}(a+j)\\
E[X] &= \sum_{j=1}^{\infty}\frac{1}{2a\mathrm{e}^{a}}\frac{a^j}{j!}(a+j)j\\
&= \frac{1}{2a\mathrm{e}^{a}}\left[a^2\sum_{j=1}^{+\infty}\frac{a^{j-1}}{(j-1)!}+a\sum_{j=1}^{+\infty}\frac{a^{j-1}}{(j-1)!}j\right]\\
x\mathrm{e}^{x} &= \sum_{n=0}^{\infty}\frac{x^{n+1}}{n!}\\
\mathrm{e}^{x}+x\mathrm{e}^{x} &= \sum_{n=0}^{\infty}\frac{x^n}{n!}(n+1)\\
E[X] &= a+\frac{1}{2}
\end{aligned}$$

□

7 条件期望

定义 7.1. 设 (X,Y) 是离散型随机向量，若 $P(X=x)>0$，则称

$$f_{Y|X}(y|x)=P(Y=y\big|X=x)$$

是给定 $X=x$ 下 Y 的条件分布列，

$$F_{Y|X}(y|x)=P(Y\leqslant y\big|X=x)$$

是给定 $X=x$ 下 Y 的条件分布，

$$\psi(x)=E[Y\big|X=x]=\sum_{y}yf_{Y|X}(y\big|x)$$

是给定 $X=x$ 下 Y 的条件期望，

8 条件期望习题

3.11.4

证明. 按定义计算 $f_{Y|Z}$ 和 $f_{Z|Y}$，帮助你熟悉概念的基础得不能再基础的题. □

3.7.5 设机器的寿命（单位：天）是以 f 为质量函数的随机变量 T. 假设机器已经工作了 t 天，问机器的剩余寿命的期望是：

(1) $f(x) = (N+1)^{-1}, x \in \{0, 1, \cdots, N\}$

(2) $f(x) = 2^{-x} x = 1, 2, \cdots$

证明. 本题是对事件求条件期望，所以仅仅是权重重新分配的加权平均. □

3.7.6 设 $X_1, X_2, \cdots$ 独立同分布，期望是 μ，设 N 是取非负整数值的随机变量，与 X_i 独立. 令 $S = X_1 + X_2 + \cdots + X_N$. 证明 $E(S\big|N) = \mu N$.

证明. 我想这背后应该有一些东西，但目前按照朴素的想法来理解还是很容易的. □

9 随机游走

9.1 n 步游走最远距离

设 $S_0 = 0$，记 $M_n = \max\{S_0, S_1, \cdots, S_n\}$.

- 不是指绝对值的最大，而是指通常序关系 $\cdots < -1 < 0 < 1 < \cdots$ 下的最大.
- 但绝对值最大也是值得关心的问题.

直接考虑 $M_n = r$ 的概率并不容易，往往是先考虑 $M_n \geqslant r$ 的概率，还要用全概率公式讨论 S_n 的位置. 确定下 S_n 的位置的好处在于左右步数的分配被确定，从而概率的计算问题转化为轨道的数数问题.

剩下的计算是自然的：

$$P(M_n \geqslant r, S_n = b) = \begin{cases} P(S_n = b), & b \geqslant r \\ \left(\dfrac{q}{p}\right)^{r-b} P(S_n = 2r - b) & b < r \end{cases}$$

- $M_n \geqslant r$ 被转化为经过直线 $y = r$ 至少一次.
- 当 $S_n = b$ 落在直线 $y = r$ 下方时要想数过程中经过 $y = r$ 至少一次的轨道当然是把 $S_n = b$ 对称过去. 同时虽然轨道数是相同的，但因为 $2r - b$ 相较 b 靠右 $2r - 2b$ 个格，粒子就得向右多跳 $r - b$ 步，向左少跳 $r - b$ 步，从而会多一个 $\left(\dfrac{q}{b}\right)^{r-b}$ 的系数.
- 上式对于 $r \geqslant 0$ 都是对的.

对称随机游走的好处是，即使左右步数的分配不同，即终点的位置不同，不同轨道的权重仍然是相同的，反映在上面的式子中，就是系数 $\left(\dfrac{q}{p}\right)^{r-b}$ 没掉了，从而我们可以直接相加，有

$$P(M_n \geqslant r) = P(S_n \geqslant r) + P(S_n \geqslant r + 1).$$

10 随机游走习题

3.9.1 设 T_k 是粒子从 $S_0 = k$ 出发，到最终被 0 和 N 处的吸收壁吸收的概率，其中 $0 \leqslant k \leqslant N$. 如 $T_0 = T_N = 0$. 证明

(1) $P(T_k < \infty) = 1$ 对任意 $0 \leqslant k \leqslant N$ 成立.

(2) $E[T_k^m] < \infty$ 对任意 $0 \leqslant k \leqslant N$ 和 $m \geqslant 1$ 成立.

设 S_n 为简单对称随机游走，$S_0 = 0$，记 $\tau_k = \inf\{n \geqslant 1 : S_n = k\}$，

Chapter 4

生成函数及其应用

1 生成函数

1.1 数列的生成函数

1.2 非负整值随机变量的生成函数

1.3 典型分布的特征函数

1.4 与数字特征的关系

Chapter 5

连续型随机变量

1　概率密度函数

定义 1.1. 称随机变量 X 是连续型随机变量，如果存在可积函数 $f:\mathbb{R}\to[0,+\infty)$，使得

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

称 $f(x)$ 为概率密度函数.

2 例子

2.1 正态分布

正态分布的概率密度函数我是记不住的，哪怕能背下来时间久了也会忘，但是我只需要知道

-

定义 2.1. 设 $X_1, \cdots, X_n$ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，若对任意 $x_1, \cdots, x_n \in \mathbb{R}$ 成立

$$P(X_1 \leqslant x_1, \cdots, X_n \leqslant x_n) = P(X_1 \leqslant x_1) \cdots P(X_n \leqslant x_n),$$

则称 $X_1, \cdots, X_n$ 是相互独立的.

2.2 Γ 分布

- $\Gamma(t) = \displaystyle\int_0^{+\infty} x^{t-1}\mathrm{e}^{-x}\mathrm{d}x$
- $\displaystyle\int_0^{+\infty} \frac{1}{\Gamma(t)} x^{t-1}\mathrm{e}^{-x}\mathrm{d}x = 1$
- $\displaystyle\int_0^{+\infty} \frac{1}{\Gamma(t)} (\lambda x)^{t-1}\mathrm{e}^{-\lambda x}\mathrm{d}(\lambda x) = \int_0^{+\infty} \frac{1}{\Gamma(t)} \lambda^t x^{t-1}\mathrm{e}^{-\lambda x}\mathrm{d}x = 1$

3 一般理论

定义 3.1. 设 X 与 Y 是独立的连续型随机变量，称 $Z := X+Y$ 为 X 与 Y 的卷积.

命题 3.1. 设 (X,Y) 有密度 $f(x,y)$，则 $Z=X+Y$ 有密度

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

证明.

$$\begin{aligned}
P(Z\leqslant z)&=P(X+Y\leqslant z)\\
&=\iint_{x+y\leqslant z} f(x,y)\mathrm{d}x\mathrm{d}y\\
&=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f(x,y)\mathrm{d}y\mathrm{d}x\\
&=\int_{-\infty}^{+\infty}\int_{-\infty}^{z} f(x,y-x)\mathrm{d}y\mathrm{d}x\\
&=\int_{-\infty}^{z}\int_{-\infty}^{+\infty} f(x,y-x)\mathrm{d}y\mathrm{d}x\\
f_Z(z)&=\int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x
\end{aligned}$$

□

4 条件期望

5 多元正态分布

定义 5.1. 称 $\vec{X}=(X_1,\cdots,X_n)$ 服从 n 元正态分布，如果 $\vec{X}$ 有概率密度函数

$$f(\vec{x})=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T\right\},$$

其中 $\vec{\mu}\in\mathbb{R}^n$，Σ 为正定矩阵. 记 $\vec{X}\sim N(\vec{\mu},\Sigma)$.

验证 $f(\vec{x})$ 确为概率密度函数：令 $\vec{X}=\vec{\mu}+\vec{Y}O$，其中 O 是使得 $O\Sigma O^T=\Lambda=diag(\lambda_1,\cdots,\lambda_n)$ 的正交矩阵. 由线性代数的知识，这样的 O 是存在的. 则

$$\begin{aligned}\int_{\mathbb{R}^n}f(\vec{x})\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n&=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\int_{\mathbb{R}^n}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T\right\}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n\\&=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\int_{\mathbb{R}^n}\exp\left\{-\frac{1}{2}\vec{y}\Lambda^{-1}\vec{y}^T\right\}\mathrm{d}y^1\wedge\cdots\wedge\mathrm{d}y^n\\&=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\prod_{i=1}^n\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}\frac{y_i^2}{\lambda_i}\right\}\mathrm{d}y^i\\&=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\prod_{i=1}^n\sqrt{2\pi\lambda_i}=1.\end{aligned}$$

命题 5.1. 设 $\vec{X}\sim N(\vec{\mu},\Sigma)$，则

(1) $E[\vec{X}]=\vec{\mu}$.

(2) $Cov(X_i,X_j)=\Sigma_{ij}$.

证明.

(1)

$$\begin{aligned}E[X_i]&=\int_{\mathbb{R}^n}x_if(\vec{x})\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n\\&=\int_{\mathbb{R}^n}(\mu_i+y_jo_{ji})\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left\{-\frac{1}{2}\vec{y}\Lambda^{-1}\vec{y}^T\right\}\mathrm{d}y^1\wedge\cdots\wedge\mathrm{d}y^n\\&=\mu_j\end{aligned}$$

(2)

□

定理 5.1. 设 $\vec{X}\sim N(\vec{\mu},\Sigma)$，$D$ 是 n 阶非奇异方阵. 令 $\vec{Y}=\vec{X}D$，则 $\vec{Y}\sim N(\vec{\mu}D,D^T\Sigma D)$.

证明.

$$\begin{aligned}f_X(\vec{x})\mathrm{d}\vec{x}&=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T\right\}\mathrm{d}\vec{x}\\&=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left\{-\frac{1}{2}(\vec{y}D^{-1}-\vec{\mu})\Sigma^{-1}(\vec{y}D^{-1}-\vec{\mu})^T\right\}\mathrm{d}\vec{y}D^{-1}\\&=\frac{1}{\sqrt{(2\pi)^n|D^T\Sigma D|}}\exp\left\{-\frac{1}{2}(\vec{y}-\vec{\mu}D)(D^T\Sigma D)^{-1}(\vec{y}-\vec{\mu}D)^T\right\}\mathrm{d}\vec{y}\end{aligned}$$

$$= f_Y(\vec{y})\mathrm{d}\vec{y}$$

□

定理 5.2. 设 $\vec{X} \sim N(\vec{\mu}, \Sigma)$，$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$，则 $\vec{X}^{(i)} \sim N(\vec{\mu}^{(i)}, \Sigma_{ii})$.

定理 5.3. 设 $\vec{X} \sim N(\vec{\mu}, \Sigma)$，则 $X_1, \cdots, X_n$ 相互独立 $\Longleftrightarrow$ Σ 为对角阵.

Chapter 6

特征函数与极限定理

1 再谈期望

- 一方面，随机变量 X 的期望 $E[X]$ 就是可测函数 X 关于测度 P 的积分 $\int_\Omega X \mathrm{d}P$.
- 另一方面，给定随机变量 X，设其分布函数为 F. X 将测度 P 推出到 $\mathbb{R}$ 上得到测度 μ_F. 称 Borel 可测函数 $g : \mathbb{R} \to \mathbb{R}$ 关于 μ_F 的积分为 Lebesgue-Stieltjes 积分，记作 $\int_{\mathbb{R}} g \mathrm{d}F$.

定理 1.1.

$$E[g(X)] = \int_{\mathbb{R}} g \mathrm{d}F.$$

2 特征函数

- 特征函数可将求各阶矩的积分运算化成微分运算.
- X 的函数的特征函数比 X 的函数的概率密度函数容易求.

定义 2.1. 复随机变量

- 复随机变量与实二维随机向量的本质区别在于复随机变量可以做乘法.
- 复随机变量的独立性的定义
- 复随机变量的期望的定义
- Z_1 与 Z_2 独立 $\Longrightarrow E[Z_1Z_2]=E[Z_1]E[Z_2]$.

定义 2.2. $\phi_X(t):=E[\mathrm{e}^{\mathrm{i}tX}]$

定理 2.1.

(1) $\phi(0)=1$

(2) ϕ 在 $(-\infty,+\infty)$ 上一致连续.

(3) ϕ 非负定. $\forall t_1,\cdots,t_n\in\mathbb{R}$, $\forall z_1,\cdots,z_n\in\mathbb{C}$，成立

$$\sum_{j,k=1}^{n} z_j\bar{z}_k\phi(t_j-t_k)\geqslant 0.$$

证明.

(2)

$$|\phi(t+h)-\phi(t)|=\left|\int \mathrm{e}^{\mathrm{i}tx}(\mathrm{e}^{\mathrm{i}hx}-1)\mathrm{d}F\right|\leqslant\int|\mathrm{e}^{\mathrm{i}hx-1}|\mathrm{d}F\xrightarrow{DCT}0,\quad h\to 0.$$

□

定理 2.2. 当 X 与 Y 独立时，

$$\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t).$$

证明.

$$\phi_{X+Y}(t)=E[\mathrm{e}^{\mathrm{i}t(X+Y)}]=E[\mathrm{e}^{\mathrm{i}tX}\mathrm{e}^{\mathrm{i}tY}]=E[\mathrm{e}^{\mathrm{i}tX}]E[\mathrm{e}^{\mathrm{i}tY}]=\phi_X(t)\phi_Y(t).$$

□

当 X 与 Y 独立时，X 与 Y 的加法对应于概率密度函数的卷积，而卷积的傅里叶变换对应于傅里叶变换的乘法.
5.7.1 给出两个不独立的随机变量 X,Y 满足 $\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t),\forall\ t\in\mathbb{R}$.

证明. □

-

定理 2.3. 设 (X, Σ, μ) 是一个测度空间，G 是 $\mathbb{C}$ 中的开集，$f: X \times G \to \mathbb{C}$. 若

(1) 对任意取定的 $x \in X$，$f(x, z)$ 是关于 z 的全纯函数.

(2) 对任意取定的 $z \in G$，$F(z) := \int_X f(x, z) \mathrm{d}\mu$ 存在.

(3) 存在 X 上的非负可积函数 g，使得

$$\left|\frac{\partial f}{\partial z}(x, z)\right| \leqslant g(x), \quad \forall x \in X, z \in G.$$

那么 F 全纯且 $F'(z) = \int_X \frac{\partial f}{\partial z}(x, z) \mathrm{d}\mu$.

证明. □

3 极限定理

定理 3.1. 设 $\{X_n\}$ 是独立同分布的随机变量序列，且 $\mu = E[X_1]$ 存在，令 $S_n = \sum_{k=1}^{n} X_k$，则

$$\frac{1}{n}S_n \xrightarrow{D} \mu.$$

证明. 由 Levy-Cramer 连续性定理，只需证明 $\phi_n(t) = E[\mathrm{e}^{\mathrm{i}t\frac{S_n}{n}}]$ 逐点收敛到 $\mathrm{e}^{\mathrm{i}\mu t}$.

记 $\phi(t) = E[\mathrm{e}^{\mathrm{i}tX_1}]$，则 $\phi_n(t) = (\phi(\frac{t}{n}))^n$.

又 $\phi(\frac{t}{n}) = 1 + \frac{\mathrm{i}\mu t}{n} + o(\frac{t}{n})$，所以

$$\lim_{n\to+\infty} \phi_n(t) = \left(1 + \frac{\mathrm{i}\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n = \mathrm{e}^{\mathrm{i}\mu t}.$$

□

Chapter 7

几种收敛

1 四种收敛

定义 1.1. 设 X, X_n 为 (Ω, F, P) 上的随机变量，

(1)

(2)

(3)

设 X, X_n 为随机变量，不必定义在同一测度空间，

(4) 对 $F_X(x) = P(X \leqslant x)$ 的连续点处，有 $F_{X_n}(x) \to F_X(x)$. 记作 $X_n \xrightarrow{D} X$.

命题 1.1. $X_n \xrightarrow{P} X \Longrightarrow X_n \xrightarrow{D} X$.

证明. □

定理 1.1.

(1) 若 $X_n \xrightarrow{D} c$，则 $X_n \xrightarrow{P} c$，其中 $c \in \mathbb{R}$ 是常数.

(2)

(3)

证明.

(1) $P(|X_n - c| > \varepsilon) = P(X_n > c + \varepsilon) + P(X_n < c - \varepsilon) = 1 - P(X_n \leqslant c + \varepsilon) + P(X_n < c - \varepsilon)$

(2)

(3)

□

2 结论拾零

2.1 Markov 不等式

2.2 Borel-Cantelli 引理

例 2.1. 设 $\{X_n\}$ 相互独立，服从参数为 1 的指数分布，试证

$$P(\limsup_{n\to+\infty}\frac{X_n}{\log n}=1)=1.$$

证明. $P(\frac{X_n}{\log n}\geqslant 1+\varepsilon)=\frac{1}{n^{\alpha+1}}$ □

例 2.2. 设 $\{X_n\}$ 相互独立，服从标准正态分布 $N(0,1)$，试证

$$P(\limsup_{n\to+\infty}\frac{|X_n|}{\sqrt{\log n}}=\sqrt{2})=1.$$

证明.

$$P(\limsup_{n\to+\infty}\frac{|X_n|}{\sqrt{\log n}}\geqslant\sqrt{2}(1+\varepsilon))=0,\quad \varepsilon>0$$

$$P(\limsup_{n\to+\infty}\frac{|X_n|}{\sqrt{\log n}}\geqslant\sqrt{2}(1+\varepsilon))=1,\quad \varepsilon\leqslant 0$$

$$P(\limsup_{n\to+\infty}\frac{|X_n|}{\sqrt{\log n}}=\sqrt{2})=1$$

□

3 强大数律

定理 3.1. 设 $\{X_k\}$ 独立同分布，$E[X_1^2] < +\infty$，$E[X_1] = \mu$，则

(1) $\frac{1}{n}\sum_{n=1}^{\infty} X_i \xrightarrow{2} \mu$

(2) $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$

太长不看版：样本均值收敛于期望值.

4 习题

7.11.7 Show that $X_n \xrightarrow{a.e.} X$ whenever $\sum_n E(|X_n - X|^r) < \infty$.

Chapter 8

外篇

1 信息熵

2 Lindeberg 替换术

Chapter 9

我的一些观察

1

如果 $f(x)$ 是概率密度函数，那么 $f_\lambda(x) := \lambda f(\lambda x)$ 也是概率密度函数，对于任意的 $\lambda > 0$.

- 期望
$$E[X_\lambda] = \int x f_\lambda(x)\mathrm{d}x = \frac{1}{\lambda}\int \lambda x f(\lambda x)\mathrm{d}\lambda x = \frac{1}{\lambda}E[X]$$

2

2.1

想证明随机变量 A 是随机变量 B，首先你要知道 B 的分布/概率密度函数，然后直接验证 A 的分布/概率密度函数就是 A 的分布/概率密度函数.

例 2.1. 设 $X_1, X_2, \cdots, X_n$ 是独立的标准正态分布，证明：

(1) X_1^2 服从 $\chi^2(1)$.

(2)

(3)

证明. □

3 独立性

定义 3.1. 设 $X_1, \cdots, X_n$ 为 (Ω, F, P) 上的随机变量，称它们是独立的，如果对任意 $x_1, \cdots, x_n \in \mathbb{R}$ 有

$$P(X_1 \leqslant x_1, \cdots, X_n \leqslant x_n) = P(X_1 \leqslant x_1) \cdots P(X_n \leqslant x_n).$$

命题 3.1. $X_1, \cdots, X_n$ 相互独立，如果对任意 $B_1, \cdots, B_n \in B(\mathbb{R})$，有

$$P(X_1 \in B_1, \cdots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n).$$

4 用到 Markov 不等式的习题

7.11.10 Show that $X_n \xrightarrow{P} 0$ iff

$$\lim_{n\to+\infty} E\left(\frac{|X_n|}{1+|X_n|}\right) = 0$$

证明. 注意到 $g(u) = \dfrac{u}{1+u}$ 是 $[0,\infty)$ 上的单调递增函数. 因此，对任意 $\varepsilon > 0$，由 Markov 不等式，

$$P(|X_n| > \varepsilon) = P\left(\frac{|X_n|}{1+|X_n|} > \frac{\varepsilon}{1+\varepsilon}\right) \leqslant \frac{1+\varepsilon}{\varepsilon} \cdot E\left(\frac{|X_n|}{1+|X_n|}\right).$$

□

5

4.1.3

6 最大值与最小值

4.5.4