第零章、预备知识

§0.1 概率空间

- 随机试验= 概率空间$(\Omega, \mathscr{F}, P)$.
- 样本空间/全集/所有试验结果: Ω.
- 样本/点/试验结果: ω,
- 事件/子集: $A, B, \cdots$. 关于ω 的性质/要求.
- σ 代数: $\mathscr{F}$. 集合系, 满足三条性质.
- 集合系$\mathscr{E}$ 生成的σ 代数: $\sigma(\mathscr{E}) := \bigcap_{\mathscr{F}\text{是}\sigma\text{代数}, \mathscr{F}\supseteq\mathscr{E}} \mathscr{F}$.
- 一般地, Ω 可换为非空集$\mathbb{X}$, 称$(\mathbb{X}, \mathscr{F})$ 为可测空间.
- 例. 离散型. $\mathbb{X}$ 可数; **默认** $\mathscr{F} = 2^{\mathbb{X}} := \{A : A \subseteq \mathbb{X}\}$.
- 例. 连续型. $\mathbb{X} = \mathbb{R}$, **默认** $\mathscr{F} = \mathscr{B} := \sigma(\text{开集}) = \sigma(\text{区间})$.

- 概率: P. $\mathscr{F}$ 上的函数, 满足三条性质:

$$\text{非负、规范、可列可加性.}$$

- 一般地, 设$(\mathbb{X},\mathscr{F})$ 为可测空间. 若$\mu:\mathscr{F}\to\mathbb{R}_+\cup\{\infty\}$ 满足

$$\mu(\varnothing)=0\text{、可列可加性,}$$

则称μ 为$\mathscr{F}$ (或$\mathbb{X}$) 上的测度.

- 概率是测度, 也称概率测度.
- 若$0<\mu(\mathbb{X})<\infty$, 则有限测度μ 可归一化为概率测度$\hat{\mu}$.

$$\hat{\mu}(A):=\frac{\mu(A)}{\mu(\mathbb{X})},\quad \forall A\in\mathscr{F}.$$

- 例. 离散型. $\mathbb{X}$ 可数.

概率μ 对应$\mathbb{X}$ 上的(概率)分布列 $\{\mu_i : i\in\mathbb{X}\}$. (测度类似.)

$$\mu_i=\mu(\{i\}),\quad \forall i\in\mathbb{X}.$$

- 例. 连续型. $(\mathbb{R}, \mathscr{B})$. 称$\mathbb{R}$ 上的概率测度为分布.
- 分布函数 $F: \mathbb{R} \to \mathbb{R}$, 满足三条性质:

  单调上升、规范($F(\infty) = 1$ & $F(-\infty) = 0$)、右连续.
- 注: F 是分布函数; $\mathscr{E} = \{(-\infty, x] : x \in \mathbb{R}\}$ 对交运算封闭, 即

  $$A, B \in \mathscr{E} \Rightarrow AB \in \mathscr{E};$$

  且$\sigma(\mathscr{E}) = \mathscr{B}$. 故在$(\mathbb{R}, \mathscr{B})$ 上存在唯一的分布μ 满足

  $$F(x) = \mu\big((-\infty, x]\big), \quad x \in \mathbb{R}.$$
- 分布函数$F \longleftrightarrow$ 分布μ.
- 定理*: 若$\mathscr{E}$ 对交运算封闭且$\mu|_{\mathscr{E}} = \hat{\mu}|_{\mathscr{E}}$, 则$\mu|_{\sigma(\mathscr{E})} = \hat{\mu}|_{\sigma(\mathscr{E})}$.

§0.2 从随机变量到随机过程

- 随机变量$X:\Omega\to\mathbb{R}$, $\omega\mapsto X(\omega)$, 满足可测性要求:

$$\{X\leqslant x\}:=\{\omega: X(\omega)\leqslant x\}\in\mathscr{F},\quad \forall x.$$

- 离散型随机变量. X 取值范围为可数集 $\mathbb{X}\subseteq\mathbb{R}$.
- 重点: X 的分布/分布列.

$$\{P(X=x):\ x\in\mathbb{X}\}.$$

- 注: 定义随机变量X 时只需要$(\Omega,\mathscr{F})$, 不需要概率P.
  但一般情况是X 出自某概率模型, 因此有默认的P.
- 注:可测性要求:

$$\{X=x\}\in\mathscr{F},\quad \forall x\in\mathbb{X}.$$

- 注: 仅需关注P 在$\sigma(X):=\sigma\left(\{\{X=x\}:x\in\mathbb{X}\}\right)$ 上的限制.
- 注: $\sigma(X)$ 是使得X 可测的最小的σ 代数.

概率论课程中的离散型随机变量.

- 两点分布:

$$P(X=1)=p, \quad P(X=0)=1-p.$$

- 二项分布$B(n,p)$:

$$P(X=i)=C_n^i p^i (1-p)^{n-i}, \quad i=0,1,\cdots,n.$$

- 泊松分布:

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

- 几何分布:

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

推广离散型随机变量的定义.

- 随机试验/概率模型: $(\Omega, \mathscr{F}, P)$.
- 设S 为非空的可数集. 若$X : \Omega \to S$ 满足可测性要求:

$$\{X = i\} = \{\omega : X(\omega) = i\} \in \mathscr{F}, \quad \forall i \in S,$$

则称X 为离散型随机变量.

- 称S 为X 的(取)值空间/状态(State)空间/位置空间.
- 值/状态/位置: $i, j, k \cdots \in S$.
- 不妨设$S = \{1, \cdots, N\}$, $\{0, 1, 2, \cdots\}$, $\cdots$
- $X \sim \mu$, X 服从分布(列) μ, X 的分布μ_X 为μ:

$$P(X = i) = \mu_i, \quad \forall i \in S.$$

- $X \stackrel{d}{=} Y$, 同分布: $\mu_X = \mu_Y$, 同分布列.

- 状态.

  例. 抛硬币, $S=\{\mathrm{H},\mathrm{T}\}$.

  例. 投骰子, $S=\{$红,橙,黄,绿,蓝,紫$\}$.

  ① ②

- 位置.

  例. $S=\{1,\cdots,5\}$.

  $X=$ (静态的)粒子的位置.

  ③ ④

- 例. 设X 服从泊松分布:

  ⑤

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

理解为: 将粒子按以上分布列置于$\mathbb{Z}_+$ 中的随机位置X.

离散型随机向量.

- 随机试验/概率模型: $(\Omega, \mathscr{F}, P)$. 设S 为非空的可数集.
- 随机向量$\vec{X} = (X_1, \cdots, X_n)$. 其中, $X_m : \Omega \to S$, $\forall m$.
- $\vec{X}$ 取值于S^n 的离散型随机变量.

$$S^n := \{\vec{x} = (x_1, \cdots, x_n) : x_1, \cdots, x_n \in S\},$$
$$\vec{x} = \vec{X}(\omega), \quad (x_1, \cdots, x_n) = (X_1(\omega), \cdots, X_n(\omega)).$$

- 称$\vec{X}$ 的(在S^n 上的)分布为n 维联合分布. (边缘/条件分布.)
- 注: 仅需关注P 在$\sigma(\vec{X})$ 上的限制,

$$\begin{aligned}\sigma(\vec{X}) =& \sigma(X_1, \cdots, X_n) := \sigma\left(\{\{\vec{X} = \vec{x}\} : \vec{x} \in S^n\}\right) \\ =& \sigma\left(\{\{X_m = i\} : 1 \leqslant m \leqslant n, i \in S\}\right).\end{aligned}$$

- 注: $\sigma(\vec{X})$ 为使得所有X_m 均可测的最小的σ 代数.
- 注: 类似地, 可设$S_1, \cdots, S_n$ 均为可数集, X_m 取值于S_m.

离散型随机变量序列.

- 无穷维向量$\vec{X} = (X_1, X_2, \cdots)$. 其中, $X_m : \Omega \to S$, $\forall m$.
- 注: 仅需交代全部有限维联合分布.
- 例. 事件的独立性→ 随机变量的独立性; 独立同分布, i.i.d..
- 例. Bernoulli 试验. 譬如,

$$P(X_1 = H, X_2 = H, X_3 = T) = p^2(1-p), \quad \cdots$$

- 例. $X_1, X_2, \cdots$ 独立且都服从参数为λ 的泊松分布, 指:

$$P(X_1 = i_1, \cdots, X_n = i_n) = e^{-\lambda n} \prod_{k=1}^{n} \frac{\lambda^{i_k}}{i_k!}, \quad \forall n, \forall i_1, \cdots, i_n.$$

对无穷维向量$\vec{X} = (X_1, X_2, \cdots)$ 仅需交代全部有限维联合分布.

为什么?

- 仅需关注P 在 “$\sigma(\vec{X})$” 上的限制, 其中

$$\sigma(\vec{X}) := \sigma\left(\{\{X_m = i\} : m \geqslant 1, i \in S\}\right).$$

  为使得所有X_i 均可测的最小的σ 代数.

- $\mathscr{E}$ 满足交运算封闭, 且$\sigma(\mathscr{E}) = \sigma(\vec{X})$. 其中,

$$\mathscr{E} = \bigcup_{n=1}^{\infty} \sigma(X_1, \cdots, X_n) = \bigcup_{n=1}^{\infty} \sigma\left(\{\{X_m = i\} : 1 \leqslant m \leqslant n, i \in S\}\right).$$

- 由定理*, 若$\mu|_{\mathscr{E}} = \hat{\mu}|_{\mathscr{E}}$, 则$\mu|_{\sigma(\mathscr{E})} = \hat{\mu}|_{\sigma(\mathscr{E})}$.
- $\mu|_{\mathscr{E}}$ 即为全部有限维联合分布.

离散时间参数、离散状态空间的随机过程.

- 即，离散型随机变量序列$\vec{X} = (X_0, X_1, X_2, \cdots)$.
- 将$\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$ 理解为离散时间. 时间参数记为n.
  初始时刻$n = 0$,
  下一时刻$n = 1$,
  再下一时刻$n = 2, \cdots$
- X_n = 动态的粒子在时刻n 的位置;
  演变的系统在时刻n 的状态.
  例. $X_0 = 1, X_1 = 4, X_2 = 4, X_3 = 3, \cdots$
- $\vec{X}$ 记录下粒子的运动过程/轨道;
  或系统的演变过程.

① ② ③ ④ ⑤

- 轨道: 以时间n 为自变量、取值于位置空间S 的函数;

  无穷维向量; 无穷长的S 字符串.

- 轨道空间:

$$S^{\mathbb{Z}_+} = \{\vec{x} = (x_0, x_1, x_2, \cdots) : x_n \in S, \forall n \geqslant 0\}.$$

- $\vec{X} : \Omega \to S^{\mathbb{Z}_+}$:

$$\vec{x} = \vec{X}(\omega), \quad (x_0, x_1, x_2, \cdots) = (X_0(\omega), X_1(\omega), X_2(\omega), \cdots)$$

- 将$\vec{X}$ 改记为$\{X_n : n \geqslant 0\}$, 一族随机变量.
- 进一步: 往谈论$\vec{X}$ 是取值于$S^{\mathbb{Z}_+}$ 的随机轨道. 可测性要求?

推广随机变量的定义—随机元.

- 随机变量$X:\Omega\to\mathbb{R}$, $\omega\mapsto X(\omega)$, 满足可测性要求:

$$\{X\leqslant x\}:=\{\omega:X(\omega)\leqslant x\}\in\mathscr{F},\quad \forall x.$$

- 注: $\mathscr{B}=\sigma(\mathscr{E})$, 其中$\mathscr{E}=\{(-\infty,x]:x\in\mathbb{R}\}$.

$$\begin{array}{ccc}\Omega & \longrightarrow & \mathbb{R}\\ \{X\leqslant x\} & \longleftarrow & (-\infty,x]\end{array}$$

- $X^{-1}D:=\{X\in D\}$, $X^{-1}\mathscr{E}:=\{X^{-1}D:D\in\mathscr{E}\}$.

  命题: 下图可交换.

$$\begin{array}{ccc}X^{-1}\mathscr{E} & \longleftarrow & \mathscr{E}\\ \downarrow & & \downarrow\\ \sigma(X^{-1}\mathscr{E}) & \longleftarrow & \sigma(\mathscr{E})\end{array}$$

- 称$\sigma(X):=\sigma\big(\{X\leqslant x\}:x\in\mathbb{R}\big)$ 为X 生成的σ 代数. 它是使X 可测的最小的σ 代数. 可测性要求的本质是$\sigma(X)\subseteq\mathscr{F}$.

- 随机试验/概率模型: $(\Omega, \mathscr{F}, P)$.
- 设$(\mathbb{X}, \mathscr{S})$ 为可测空间. 若$X : \Omega \to \mathbb{X}$ 满足可测性要求:

$$\{X \in D\} \in \mathscr{F}, \quad \forall D \in \mathscr{S},$$

则称X 为取值$\mathbb{X}$ 的随机元.

- 注: 称$\mathbb{X}$ 上的概率测度为分布.
- $X \sim \mu$, X 服从分布μ, X 的分布μ_X 为μ:

$$P(X \in D) = \mu(D), \quad \forall D \in \mathscr{S}.$$

例. 离散时间参数、离散状态空间的随机过程.

- 设$(\Omega, \mathscr{F}, P)$ 为概率模型, S 为可数集,
  $\vec{X} = (X_0, X_1, X_2, \cdots)$, 其中, $X_m : \Omega \to S$.
- 令

$$\mathbb{X} = S^{\mathbb{Z}_+} = \{\vec{x} = (x_0, x_1, x_2, \cdots) : x_n \in S, \forall n \geqslant 0\},$$

$$\mathscr{S} = \sigma\left(\left\{\{\vec{x} : x_n = i\} : n \in \mathbb{Z}_+, i \in S\right\}\right).$$

- 则, $\vec{X}$ 可理解为取值$\mathbb{X}$ 的随机元, 即随机轨道.
- 注: 仅需交代全部有限维联合分布列.

$$P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n), \quad n \geqslant 0, i_0, i_1, \cdots, i_n \in S.$$

连续时间参数、离散状态空间的随机过程.

- 设$(\Omega, \mathscr{F}, P)$ 为概率模型, S 为可数集.
- 取连续时间参数 $T = \mathbb{R}_+ = [0, \infty)$.
- **一族随机变量**: $\mathbf{X} = \{X_t, t \geqslant 0\}$, 其中$X_t : \Omega \to S$, $\forall t \geqslant 0$.
- 使得所有X_t 均可测的最小σ 代数:

$$\sigma(\mathbf{X}) := \sigma\left(\{\{X_t = i\} : t \geqslant 0, i \in S\}\right).$$

- $\mathbf{X}$ 取值于轨道空间

$$\mathbb{X} = S^T := \{\mathbf{x} = \{x_t, t \geqslant 0\} : x_t \in S, \forall t\}.$$

- 取$\mathscr{S} = \sigma\left(\{\{\mathbf{x} : x_t = i\} : t \geqslant 0, i \in S\}\right)$, 则$\mathbf{X}$ 可理解为取值$\mathbb{X}$ 的随机元, 即随机轨道.
- 仍然仅需交代全部有限维联合分布.

$$\mathscr{E} = \bigcup_{n=1}^{\infty} \bigcup_{t_1, \cdots, t_n \geqslant 0} \sigma\left(X_{t_1}, \cdots, X_{t_n}\right).$$

连续时间参数、连续状态空间的随机过程.

- 设$(\Omega, \mathscr{F}, P)$ 为概率模型, 取连续时间参数$T = \mathbb{R}_+ = [0, \infty)$.
- 一族取值于$\mathbb{R}$ 的随机变量: $\mathbf{X} = \{X_t : t \geqslant 0\}$.
- $(X_{t_1}, \cdots, X_{t_n})$ 均为连续型随机向量, $\forall n \geqslant 1$, $t_1 < \cdots < t_n$.

  注: 允许$X_0 \equiv x_0$.
- 使得所有X_t 均可测的最小σ 代数:

$$\sigma(\mathbf{X}) := \sigma\left(\{\{X_t \leqslant x\} : t \geqslant 0, x \in \mathbb{R}\}\right).$$

- $\mathbf{X}$ 取值于轨道空间

$$\mathbb{X} = \mathbb{R}^T := \{\mathbf{x} = \{x_t, t \geqslant 0\} : x_t \in \mathbb{R}, \forall t\}.$$

- 取$\mathscr{S} = \sigma\left(\{\{\mathbf{x} : x_t \leqslant x\} : t \geqslant 0, x \in \mathbb{R}\}\right)$, 则$\mathbf{X}$ 可理解为取值$\mathbb{X}$ 的随机元, 即随机轨道.

- 仍然仅需交代全部有限维联合分布.

$$\mathscr{E} = \bigcup_{n=1}^{\infty} \bigcup_{t_1,\cdots,t_n \geqslant 0} \sigma\left(X_{t_1}, \cdots, X_{t_n}\right).$$

- 等价地, 交代全部有限维联合密度.
- 注: 类似地, 以上离散时间参数或连续时间参数均可取到负半轴, 均可取区间段.

  例. $T = \mathbb{Z}$, $\{0, 1, \cdots, N\}$, $\{N_1, \cdots, N_2\}$;

  或, $T = \mathbb{R}$, $[0, T]$, $[T_1, T_2]$.

独立性.

- 事件的独立性.
- 随机变量的独立性: $\forall A_1, \cdots, A_n$,

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

- 等价条件, 离散型、连续型.
- 命题0.2.7. 设X, Y 为离散型随机变量, 分别取值于S_1, S_2. 若

$$P(X = i, Y = j) = \mu_i P(Y = j), \quad \forall i \in S_1, j \in S_2,$$

则X 与Y 相互独立, 且$X \sim \mu$.

- 证: 上式两边对j 求和, 可得$X \sim \mu$. 再将其再代入上式, 便知X 与Y 相互独立. □

设以下随机变量都取值于可数集S, 或者都取值于$\mathbb{R}$.

- 设$\mathbf{X} = \{X_\alpha : \alpha \in I\}$ 与$\mathbf{Y} = \{Y_\beta : \beta \in J\}$ 是两族随机变量.
- 若任意$(X_{\alpha_1}, \cdots, X_{\alpha_n})$ 与任意$(Y_{\beta_1}, \cdots, Y_{\beta_m})$ 相互独立, 则称$\mathbf{X}$ 与$\mathbf{Y}$ 相互独立.
- 原因: 令

$$\mathscr{E}_{\mathbf{X}} := \bigcup_{n=1}^{\infty} \bigcup_{\alpha_1,\cdots,\alpha_n \in I} \sigma\left(X_{\alpha_1}, \cdots, X_{\alpha_n}\right),$$

$$\mathscr{E}_{\mathbf{Y}} := \bigcup_{m=1}^{\infty} \bigcup_{\beta_1,\cdots,\beta_m \in I} \sigma\left(Y_{\beta_1}, \cdots, Y_{\beta_m}\right).$$

  则它们都满足交运算封闭.
- 命题: 因此, $\mathscr{E}_{\mathbf{X}}$ 与$\mathscr{E}_{\mathbf{Y}}$ 独立蕴含着$\sigma(\mathscr{E}_{\mathbf{X}})$ 与$\sigma(\mathscr{E}_{\mathbf{Y}})$ 独立.
- 类似地可定义$\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(n)}$ 独立,
  $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots$ 独立, 或独立同分布.

§0.3 期望与收敛性

- 设随机变量X 取实数值.
- 离散型: $EX := \sum_{i\in S} iP(X=i)$.
- 连续型: $EX := \int_{-\infty}^{\infty} xp_X(x)\mathrm{d}x$.
- 统一定义: $EX = \int_0^\infty P(X>x)\mathrm{d}x - \int_0^\infty P(X<-x)\mathrm{d}x$.
- 若某事件的概率为1, 则称它几乎必然(a.s.)发生.
- 注: 更一般地, 验证该事件 $\supseteq A$ 且$P(A)=1$, 即可.
- 几乎必然收敛, $X_n \xrightarrow{\text{a.s.}} X$: $P(\lim_{n\to\infty} X_n = X) = 1$.
- 注: 验证$P(A)=1$ 且$\forall\omega\in A$, $X_n(\omega)\to X(\omega)$ 即可.
- 依概率收敛, $X_n \xrightarrow{P} X$: $P(|X_n - X|>\varepsilon)\to 0$, $\forall\varepsilon>0$.

- 关键的事件: $A_n = \{|X_n - X| > \varepsilon\}$.
  $X_n \xrightarrow{P} X$ iff $P(A_n) \to 0$;
  $X_n \xrightarrow{\text{a.s.}} X$ iff $P\left(\bigcup_{n\geqslant N} A_n\right) \to 0$.
- Borel-Cantelli引理: 若$\sum_{n=1}^{\infty} P(A_n) < \infty$, 则$P\left(A_n \text{ i.o.}\right) = 0$,

$$\{A_n \text{ i.o.}\} = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n.$$

- 推论: 若$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty, \forall \varepsilon > 0$, 则$X_n \xrightarrow{\text{a.s.}} X$.
- 推论: 设$\eta_1, \eta_2, \cdots$ i.i.d. 且期望存在. 则$\eta_n/n \xrightarrow{\text{a.s.}} 0$.
- 强大数定律/SLLN: 设$X_1, X_2, \cdots$ i.i.d.且期望有意义.
  则$(X_1 + \cdots + X_n)/n \xrightarrow{\text{a.s.}} EX_1$.

- 有界收敛定理, BCT:

  设$X_n \xrightarrow{P} X$. 若$\exists M$ 使得$|X_n| \leqslant M, \forall n \geqslant 1$, 则$EX_n \to EX$.

- 单调收敛定理, MCT, Levi定理:

  设X_n 非负且单调上升到X, 则$EX_n \to EX$.

- Lebesgue 控制收敛定理, DCT:

  设$X_n \xrightarrow{\text{a.s.}} X$. 若$|X_n| \leqslant Y, \forall n$ 且$E|Y| < \infty$, 则$EX_n \to EX$.

- Fubini 定理(离散时间参数版本):

  设或者X_n 都非负, 或者$\sum_{n=1}^{\infty} E|X_n| < \infty$,

  则$E\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} EX_n$.

- Fubini 定理(连续时间参数版本):

  设或者X_t 都非负, 或者$\int_0^{\infty} E|X_t|\mathrm{d}t < \infty$,

  则$E\int_0^{\infty} X_t\mathrm{d}t = \int_0^{\infty} EX_t\mathrm{d}t$.

§0.4 条件概率、条件分布与条件期望

- 条件概率: 设$P(A)>0$, 称

$$P(B|A)=P_A(B)=\frac{P(AB)}{P(A)}$$

为在A 发生的条件下, 事件B 的(条件)概率.

- 注: 只要见到“在A 发生的条件下”,“已知A”,“假设A”之类, 则其后谈及的“概率”都指采用P_A 进行计算.
- 例(离散型). 当A 成立时, X 的分布列为$\{P_A(X=i), i\in S\}$.
- 例(离散型). “若A (发生), 则X 与Y 独立”指:

$$P_A(X=i,Y=j)=P_A(X=i)P_A(Y=j),\quad \forall i,j.$$

而不是$P(X=i,Y=j)=P(X=i)P(Y=j),\ \forall i,j.$

- 条件分布: 条件分布列、条件密度.
- 条件期望$E(X|Y)$:
  1. 固定j, 用在$\{Y=j\}$ 的条件下, X 的条件分布列/条件密度求X 的期望, 得到$\varphi(j) := E(X|Y=j)$.
  2. 令$E(X|Y) := \varphi(Y)$.
- 注: 条件期望$E(X|Y)$ 是随机变量.
- 重期望公式: $EX = EE(X|Y)$.