随机数学方法整理

吕铭 Lyu Ming

2016 年 1 月 5 日

目录

1 概率事件与概率空间

1. 基本事件, 样本点
2. 和事件, 交事件 $AB = A\cap B$, 对立事件 (余集 $\overline{A}$), 互斥事件 (互不相容)
3. 加法公理: 对于互不相容的事件 A_i

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \tag{1.1}$$

可以推广到至多可数个互不相交的集合

1.1 概率空间和 σ-域

概率空间三元组 $(\Omega, \mathscr{F}, P)$

1. Ω: 全体可能结果的集合
2. $\mathscr{F}$: 可观测事件的事件族, 是 Ω 幂集的子集
3. $P : F \mapsto [0,1]$ 定义概率, 要求满足加法公理, 以及必然事件概率 1

称 $\mathscr{F}$ 为 σ-域, 若:

1. $\Omega \in \mathscr{F}$
2. $A\in\mathscr{F} \to \overline{A}\in\mathscr{F}$
3. $A_i\in\mathscr{F} \to \bigcup_i A_i\in\mathscr{F}$

最小 σ-域: $\mathscr{F}_1 = \{\emptyset, \Omega\}$, 事件 A 生成的 σ-域 $\mathscr{F}_A = \{\emptyset, A, \overline{A}, \Omega\}$

1.2 概率的公理化定义

Kolmogorov 的概率公理化定义: 对于 $\mathscr{F}$ 上的函数 P, $\forall A\in\mathscr{F}$, $P(A)$ 满足:

1. 非负性: $P(A)\geq 0$
2. 规范性: $P(\Omega) = 1$
3. 可数可加性: 两两互不相容的 $A_1, A_2, \cdots \in \mathscr{F}$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \tag{1.2}$$

则称 $P(A)$ 为事件 A 的概率 (概率测度), 三元组 $(\Omega, \mathscr{F}, P)$ 称为描述该随机试验的概率空间.

定义推论包括 $P(\emptyset) = 0$, 有限可加性等.

- 概率的下 (上) 连续性: $\mathscr{F}$ 双红的事件序列 $\{A_n\}$ 满足: $A_n \subset A_{n+1}$ (或 $A_n \supset A_{n+1}$), 则

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) (\text{或} P\left(\bigcap_{n=1}^{\infty} A_n\right)) \tag{1.3}$$

注可数可加性 $\Leftrightarrow$ 有限可加性 + 概率连续性

1.3 条件概率

在概率空间 $(\Omega, \mathscr{F}, P)$ 中 $B\in\mathscr{F}$, $P(B) > 0$, 则 $\forall A\in\mathscr{F}$, 定义事件 B 发生的条件下, 事件 A 发生的条件概率

$$P_B(A) = P(A|B) = \frac{P(A\cap B)}{P(B)} \tag{1.4}$$

- $B\in\mathscr{F}$ 且 $P(B) > 0$ 时, $(\Omega, \mathscr{F}, P_B)$ 也是概率空间

1.3.1 乘法公式

对于事件列 $A_i\in\mathscr{F}$, 满足 $P(\bigcap_i A_i) > 0$, 则

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P\left(A_i \middle| \bigcap_{j=1}^{i-1} A_j\right) \tag{1.5}$$

1.3.2 全概率公式

称一族事件 B_i 为 Ω 的一个一个划分, 如果 $B_i\cap B_j = \emptyset (i\neq j)$ 且 $\cup_i B_i = \Omega$. 如果 $P(B_i) > 0$ 则称为正划分. 对于一族正划分有公式

$$P(A) = \sum_i P(B_i)P(A|B_i) \tag{1.6}$$

1.3.3 Bayes 公式

对于一族正划分, 以及 $P(A) > 0$

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} \tag{1.7}$$

$$= \frac{P(B_i)P(A|B_i)}{\sum_j P(B_i)P(A|B_j)} \tag{1.8}$$

1.4 事件独立性

定义事件 $A, B \in \mathscr{F}$ 相互独立:

$$\begin{aligned} & P(AB) = P(A)P(B) \\ \Leftrightarrow & P(A|B) = P(A) & P(B) > 0 \\ \Leftrightarrow & P(A|\overline{B}) & P(\overline{B}) > 0 \\ \Leftrightarrow & P(A|B) = P(A|\overline{B}) & 0 < P(B) < 1 \end{aligned}$$

据此可见事件的独立性实质是 σ-域的独立性.

推广到 n 个事件: $\forall k, i_t (1 \le k \le n, 1 \le i_1 \le \cdots \le i_k \le n)$

$$P\left(\bigcap_{t=1}^{k} A_{i_t}\right) = \prod_{t=1}^{k} P(A_{i_t})$$

共包含 $2^n - n - 1$ 个等式

1.4.1 事件的相关系数

对于 $P(A), P(B) \in (0,1)$, 定义相关系数

$$r(A,B) = \frac{P(AB) - P(A)P(B)}{\sqrt{P(A)(1-P(A))P(B)(1-P(B))}} \tag{1.9}$$

- $r(A,B) = 0 \Leftrightarrow A, B$ 相互独立
- $|r(A,B)| \le 1$.
  - $r(A,B) = 1 \Leftrightarrow P(A) = P(AB) = P(B)$ 即 A, B 差别仅为零概率事件
  - $r(A,B) = -1 \Leftrightarrow P(A) = P(A\overline{B}) = P(\overline{B})$ 即 $A, \overline{B}$ 差别仅为零概率事件
- $r(A,B) > 0\ (< 0)$, A, B 正 (负) 相关
- 等价于联合两点分布的随机变量相关系数

2 随机变量

2.1 分布函数

1. 分布函数

$$F(x) = P(X \le x) \tag{2.1}$$

   - 单调不减函数
   - $F(-\infty) = 0$, $F(+\infty) = 1$
   - 右连续

2. 二维联合分布

$$F(x,y) = P(X \le x, Y \le y) \tag{2.2}$$

   - x 或 y 的单调不减函数
   - 任意 $x_1 \le x_2$, $y_1 \le y_2$

$$F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1) \ge 0$$

   - $F(-\infty, y) = F(x, -\infty) = 0$,
     $F(+\infty, +\infty) = 1$
   - 关于 x 或 y 右连续

3. 边缘分布

$$F_X(x) = F(x, +\infty) \tag{2.3}$$

4. 独立性

$$F(x,y) = F_X(x)F_Y(y) \tag{2.4}$$

2.2 随机变量的数字特征

1. 数学期望 EX

$$EX = \int_{-\infty}^{\infty} x \,\mathrm{d}F_X(x) \tag{2.5}$$

   - 关于分布函数

$$\int_{-\infty}^{EX} F(x)\,\mathrm{d}x = \int_{EX}^{\infty} [1 - F(x)]\,\mathrm{d}x \tag{2.6}$$

   推论 (积分均收敛的情况下)

$$\begin{aligned} EX = & \int_0^{\infty} P(X > x)\,\mathrm{d}x \\ & - \int_{-\infty}^{0} P(X \le x)\,\mathrm{d}x \end{aligned} \tag{2.7}$$

   - 实用统计学定律

$$Eg(x) = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}F_X(x) \tag{2.8}$$

   推广到多元

$$Eg(x,y) = \iint g(x,y)\,\mathrm{d}F(x,y) \tag{2.9}$$

2. p-分位数: x 满足

$$P(X \le x) \ge p; \quad P(X \ge x) \ge 1 - p \tag{2.10}$$

   特别的 $p = 1/2$ 时称为中位数

3. k 阶 (原点) 矩 $E(X^k)$;
   k 阶 (中心) 矩: $E((X-EX)^k)$

4. 推广联合 $k+l$ 阶混合矩 $E(X^kY^l)$. (中心矩同理)

   - 高阶矩存在, 则低阶矩必存在

5. 方差 $DX = E(X-EX)^2$

2.2.1 数学期望和方差的性质

1. $|EX| \le E(|X|)$

2. $E(\sum c_iX_i) = \sum c_iEX_i$

3. $D(\sum c_iX_i) = \sum c_i^2DX_i$

4. 相互独立, 则 $E(\prod X_i) = \prod EX_i$

5. Cauchy-Schwartz 不等式:
   若 $E(X^2), E(Y^2) < +\infty$

$$(E(XY))^2 \le E(X^2)E(Y^2) \tag{2.11}$$

6. $\min_c\{E(X-C)^2\} = E(X-EX)^2 = DX$

7. DX^2 存在, 则 $DX = EX^2 - (EX)^2$

8. $DX = 0 \Leftrightarrow P(X = EX) = 1$

2.2.2 协方差和相关性

1. 定义协方差

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[(X-EX)(Y-EY)] \\ &= E(XY) - EXEY\end{aligned} \tag{2.12}$$

   来自于: $D(X+Y) = DX + DY + 2\,\mathrm{Cov}(X,Y)$

   - $\mathrm{Cov}(X,Y)$ 是对称双线性函数
   - 对于多元

$$\begin{aligned}D\left(\sum_i c_iX_i\right) &= \sum_i c_iDX_i \\ &+2\sum_{ij} c_ic_j\,\mathrm{Cov}(X_i,X_j)\end{aligned} \tag{2.13}$$

2. 随机变量的 (线性) 相关系数

$$r_{X,Y} \equiv \frac{\mathrm{Cov}\,X,Y}{\sqrt{DX}\sqrt{DY}} \tag{2.14}$$

   - X,Y (线性) 不相关: $r_{X,Y} = 0$.
     - 独立一定不相关, 反之不亦然
   - 定理 (最佳线性预测)

$$\begin{aligned}&E[X-(\hat{a}Y+\hat{b})]^2 \\ =&\min_{a,b} E[X-(aY+b)]^2 \\ =&DX(1-r_{X,Y}^2)\end{aligned} \tag{2.15}$$

   其中

$$\hat{a} = \frac{\mathrm{Cov}(X,Y)}{DY} = r_{X,Y}\sqrt{\frac{DX}{DY}} \tag{2.16}$$

   也可以理解为在最佳预测下:

$$\frac{(\hat{X}-EX)/\sqrt{DX}}{(Y-EY)/\sqrt{DY}} = r_{X,Y} \tag{2.17}$$

   - Cauchy-Schwartz 不等式:

$$|r_{X,Y}| \le 1 \tag{2.18}$$

   - $r_{X,Y} = \pm 1$ 时 X 与 Y 完全相关, 具有准确的线性关系

$$P(X = \hat{a}Y + \hat{b}|Y) = 1$$

   - $\hat{X} = \hat{a}Y + \hat{b}$ 与残差 $X - \hat{X}$ 不相关

3. 高纬 $(X_1, \cdots, X_n)$ 推广:

   - 协方差矩阵

$$\Sigma_{ij} = \mathrm{Cov}(X_i, X_j) \tag{2.19}$$

   - 相关系数矩阵

$$R_{ij} = r_{X_i,X_j} \tag{2.20}$$

   两者均为对称矩阵. Σ 对角元为方差, R 对角元 1

2.2.3 条件期望

条件期望: $E(X|Y)$

1. 全期望公式

$$E(E(X|Y)) = EX \tag{2.21}$$

2. $E(g(Y)h(X)|Y) = g(Y)E(h(X)|Y)$

3. X,Y 相互独立, 则 $E(X|Y) = EX$

4. 最佳预测

$$E[X-E(X|Y)]^2=\min_{\varphi}E[X-\varphi(Y)]^2 \tag{2.22}$$

推论

$$DX=E[X-E(X|Y)]^2+D[E(X|Y)] \tag{2.23}$$
$$E[E(X|Y)]^2\le EX^2 \tag{2.24}$$
$$r_{X,Y}^2DX\le D[E(X|Y)]\le DX \tag{2.25}$$

其中 (2.25) 式第一个不等号可以使 $\psi(Y)=\hat{a}Y+\hat{b}$ 联合 (2.15) 式得; 第二个不等号取等号当且仅当 $X=E(X|Y)$ (X 由 Y 确定)

推论: 对于有界 Borel 函数 $g(x)$: (从线性空间角度看, 定义 E(XY) 为内积)

$$E\{([X-E(X|Y)])g(Y)\}=0 \tag{2.26}$$

3 数学工具

3.1 示性函数

对于集合 A 定义:

$$I_A(x)=\begin{cases}1 & x\in A\\ 0 & x\notin A\end{cases} \tag{3.1}$$

3.2 矩母函数

随机变量 X 的矩母函数 $m_X(u)=E(\mathrm{e}^{uX})$, $u\in\mathbb{R}$. 性质

1. $m_X(u)$ 在 $u=0$ 的某一开邻域内存在, $\Rightarrow E(X^n)=m_X^{(n)}(0)$
2. X,Y 相互独立 $\Rightarrow m_{X+Y}(u)=m_X(u)m_Y(u)$
3. 当矩母函数存在时, 它可以唯一确定随机变量的分布

特别的, 对于 Poisson 分布 $X\sim P(\lambda)$, $m_X(u)=\exp[\lambda(\mathrm{e}^u-1)]$

3.3 特征函数

定义随机变量 X 的特征函数

$$\varphi_X(\theta)=E\mathrm{e}^{\mathrm{i}\theta X} \tag{3.2}$$

推广到随机向量 $\boldsymbol{X}=(X_1,\cdots X_m)^T$ 的特征函数, 记 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_m)^T\in\mathbb{R}^m$

$$\varphi(\boldsymbol{\theta})=E\mathrm{e}^{\mathrm{i}\boldsymbol{\theta}^T\boldsymbol{X}} \tag{3.3}$$

1. $|\varphi_{\boldsymbol{X}}(\boldsymbol{\theta})|\le 1$, $\varphi_X(\mathbf{0})=1$
2. $\varphi_{\boldsymbol{X}}(\boldsymbol{\theta})$ 在 $\mathbb{R}^m$ 上一致连续
3. 复共轭 $\varphi_{\boldsymbol{X}}^*(\boldsymbol{\theta})=\varphi_{\boldsymbol{X}}(-\boldsymbol{\theta})$
4. 常数 a,b, 则 $\varphi_{\boldsymbol{a}+B\boldsymbol{X}}(\boldsymbol{\theta})=\mathrm{e}^{\mathrm{i}\boldsymbol{\theta}^T\boldsymbol{a}}\varphi_{\boldsymbol{X}}(B^T\boldsymbol{\theta})$
5. 分布密度的 Fourier 变换. 逆变换 (一维)
$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\theta x}\varphi(\theta)\,\mathrm{d}\theta \tag{3.4}$$
6. 唯一性定理: $F(\boldsymbol{x})$ 由 $\varphi(\boldsymbol{\theta})$ 唯一决定, 逆转公式
$$\begin{aligned}&F(x)-F(y)=\\&\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}\theta x}-\mathrm{e}^{-\mathrm{i}\theta y}}{-\mathrm{i}\theta}\varphi(\theta)\,\mathrm{d}\theta\end{aligned} \tag{3.5}$$
7. X,Y 相互独立 $\Rightarrow \varphi_{X+Y}(\theta)=\varphi_X(\theta)\varphi_Y(\theta)$
8. $\boldsymbol{X}$ 中各分量相互独立 $\Leftrightarrow$
$$\varphi_{\boldsymbol{X}}(\boldsymbol{\theta})=\prod_i\varphi_{X_i}(\theta_i)$$
9. 矩性质: 对于 $E|X^k|<\infty$, $j\le k$
$$EX^j=\mathrm{i}^{-j}\varphi^{(j)}(0) \tag{3.6}$$
10. 混合矩, 若 $E|\prod X_i^{k_i}|<\infty$, 则:
$$\begin{aligned}&E\left(X_1^{k_1}X_2^{k_2}\cdots X_m^{k_m}\right)=\\&(-\mathrm{i})^{k_1+\cdots+k_m}\frac{\partial^{k_1+\cdots k_m}}{\partial^{k_1}\theta_1\cdots\partial^{k_m}\theta_m}\varphi(\mathbf{0})\end{aligned} \tag{3.7}$$
11. 连续性定理: 对于随机变量列 X_n 与 X, 以下两种说法等价:
    (a) 在 $F_X(x)$ 上的一切连续点上有 $F_{X_n}(x)\to F_X(x)$. 称为 X_n 依分布收敛到分布函数 F_X, 记为 $X_n\xrightarrow{D}F_X$ (或 $X_n\xrightarrow{D}X$)
    (b) $\forall\theta\in\mathbb{R}$, $\varphi_{X_n}(\theta)\to\varphi_X(\theta)$

3.3.1 一些分布的特征函数

1. Poisson 分布 $P(\lambda)$:

$$\varphi(\theta) = \exp\left[\lambda(\mathrm{e}^{\mathrm{i}\theta} - 1)\right] \tag{3.8}$$

2. 正态分布 $N(\mu, \sigma^2)$:

$$\varphi(\theta) = \exp\left[\mathrm{i}\mu\theta - \frac{1}{2}\sigma^2\theta^2\right] \tag{3.9}$$

3. m 维 Gauss 分布 $N(\boldsymbol{\mu}, \Sigma)$

$$\varphi(\boldsymbol{\theta}) = \exp\left[\mathrm{i}\boldsymbol{\theta}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\theta}^T\Sigma\boldsymbol{\theta}\right] \tag{3.10}$$

4 离散随机变量

分布列 (分布律). $p_i = P(X = x_i)$ 的表格

4.1 离散随机向量

1. 联合分布律 $p_{ij} = P(X = x_i, Y = y_j)$

2. 边缘分布律
$P(X = x_i) = \sum_j p_{ij}$, $P(Y = y_j) = \sum_i p_{ij}$

3. 条件分布律

$$P(X = x_i | Y = y_j) = \frac{p_{ij}}{\sum_k p_{kj}} \tag{4.1}$$

4.1.1 独立性

独立性定义: $\forall x_i, y_i$

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i) \tag{4.2}$$

推广到高维 $(X_1, \cdots, X_n)$, $\forall\{x_k\}, (k = 1, 2, \cdots)$

$$P(X_i = x_i, i = 1, 2, \cdots) = \prod P(X_i = x_i) \tag{4.3}$$

4.2 数学期望

1. $EX = \sum_n p_n x_n$

2. $E(X|Y = y_j) = \sum_i x_i p_{ij}$

4.3 常见分布类型

1. 二项分布 $X \sim B(n, p)$:

$$P(X = k) = C_n^k p^k q^{n-k} \tag{4.4}$$

其中 $k = 0, 1, \cdots, n$. 特别的 $n = 1$ 称为标准二值分布

- $EX = np$, $DX = npq$

2. 超几何分布

$$P(X = k) = \frac{C_M^k C_N^{n-k}}{C_{M+N}^n} \tag{4.5}$$

其中 $k = \max(0, n - N), \cdots, \min(n, M)$

- $p = M/(M + N) > 0$ 在 $M + N \to \infty$ 时近似为二项分布
- $EX = nM/(M + N)$
- $DX = npq(M + N - n)/(M + N - 1)$

3. 几何分布 $X \sim Ge(p)$

$$P(X = k) = qp^{k-1} \tag{4.6}$$

其中 $q = 1 - p$, $k = 1, 2, \cdots$

- $EX = 1/p$, $DX = q/p^2$

截止 m 的几何分布 $P(X = m) = p^m$, $P(X > m) = 0$

- $EX = (1 - q^m)/p$

4. 负二项分布 $X \sim NB(r, p)$ (或称 Pascal 分布)

$$P(X = k) = C_{k-1}^{r-1} p^r q^{k-r} \tag{4.7}$$

其中 $q = 1 - p$, $k = r, r + 1, \cdots$

- $EX = rq/p$, $DX = rq/p^2$

5. Poisson 分布 $X \sim P(\lambda)$ 或 $\mathrm{Poisson}_\lambda$

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} \tag{4.8}$$

- $EX = DX = \lambda$

详见 6.4

5 连续型随机变量

定义: 对于随机变量 X, 存在非负可积函数 $f(x)$ $(x \in \mathbb{R})$, 使 $\forall a < b$

$$P(a < X \le b) = \int_a^b f(x)\,\mathrm{d}x \tag{5.1}$$

$$\Leftrightarrow F(x) = \int_{-\infty}^x f(u)\,\mathrm{d}u \tag{5.2}$$

称 $f(x)$ 为 X 的概率密度函数

1. $f(x) \ge 0$, $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$
2. 连续型随机变量的分布函数 $F(x)$ 是连续函数, 反之不亦然
3. 在 $f(x)$ 的连续点 $F'(x) = f(x)$
4. $f(x)$ 不唯一

5.1 连续随机向量

1. 定义 (二维): 存在非负可积函数 $f(x, y)$, 使得

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v)\,\mathrm{d}u\,\mathrm{d}v \tag{5.3}$$

$$\Leftrightarrow P((x, y) \in C) = \iint_C f(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{5.4}$$

其中 $C = \{(x, y) | a < x \le b, c < y \le d\}$. 称 $f(x, y)$ 为 (X, Y) 的联合分布密度函数

- $f(x, y) \ge 0$; $\iint_{\mathbb{R}^2} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1$
- 若 $f(x, y)$ 在 (x, y) 处连续, 则

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{5.5}$$

2. 边缘分布

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y \tag{5.6}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x \tag{5.7}$$

3. 条件密度, 对于 $f_Y(y) > 0$ (要求绝对收敛)

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} \tag{5.8}$$

5.1.1 独立性

独立性定义: $\forall x, y$

$$f(x, y) = f_X(x) f_Y(y) \Leftrightarrow f_{X|Y}(x|y) = f_X(x) \tag{5.9}$$

5.1.2 随机变量的函数分布

对于随机向量 $\boldsymbol{X}$ 的联合概率密度函数 $f(\boldsymbol{y})$, 有双射 $\boldsymbol{Y} = T(\boldsymbol{X})$ $(T : D \mapsto R)$, 则 $\boldsymbol{Y}$ 有概率密度函数

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f(\boldsymbol{x}(\boldsymbol{y})) \left| \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} \right| I_R(\boldsymbol{y}) \tag{5.10}$$

其中 $\partial \boldsymbol{x} / \partial \boldsymbol{y}$ 表示 Jacob 行列式

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \cdots \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \cdots \\ \vdots & \vdots & \ddots \end{vmatrix} \tag{5.11}$$

据此导出相关公式:

$$\begin{aligned} f_{aX+bY}(z) &= \int_{-\infty}^{\infty} \frac{1}{|b|} f\left(x, \frac{z - ax}{b}\right) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} \frac{1}{|a|} f\left(\frac{z - by}{a}, y\right) \mathrm{d}y \end{aligned} \tag{5.12}$$

$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) \mathrm{d}x \tag{5.13}$$

$$f_{X/Y}(z) = \int_{-\infty}^{\infty} |y| f(zy, y)\,\mathrm{d}x \tag{5.14}$$

特别的, 对于 (5.12) 式, 当 X, Y 相互独立, 有卷积公式

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\,\mathrm{d}y \\ &\equiv f_X * f_Y = f_Y * f_X \end{aligned} \tag{5.15}$$

5.2 数学期望

1. 一维: 如果

$$\int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x < \infty \tag{5.16}$$

则称 X 的数学期望存在, 定义为

$$EX = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \tag{5.17}$$

2. 高维同理

$$Eg(X, Y) = \iint_{\mathbb{R}^2} g(x, y) f(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{5.18}$$

5.3 常见分布类型

1. 均匀分布 $X \sim U[a,b]$

$$f(x) = \frac{1}{b-a} I_{[a,b]}(x) \tag{5.19}$$

- $EX = (a+b)/2$; $DX = (b-a)^2/12$

二维 $(X,Y) \sim U(D)$

$$f(x,y) = I_D(x,y)\frac{1}{|D|} \tag{5.20}$$

2. 指数分布 $X \sim E(\lambda)$ (或 Exp_λ)

$$f(x) = \lambda \mathrm{e}^{-\lambda x} I_{[0,+\infty)}(x) \tag{5.21}$$

- $EX = 1/\lambda$, $DX = 1/\lambda^2$
- $EX^n = \Gamma(n+1)/\lambda^n = n!/\lambda^n$

关于指数分布一下命题等价:

(a) $X \sim E(\lambda)$

(b) 无记忆性: $\forall s, t \geq 0$,

$$P(X > s+t | X > s) = P(X > t)$$

(c) 恒定增长率:

$$\lim_{h\to 0} \frac{1}{h} P(x < X \leq x+h | X > x) = \lambda$$

3. 正态分布 $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{5.22}$$

特别的 $N(0,1)$ 称为标准正态分布

- 标准正态分布的分布函数 $F(x) = \Phi(x)$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-u^2/2}\,\mathrm{d}u \tag{5.23}$$

特殊值: $\phi(-x) = 1 - \Phi(x)$, $\Phi(0) = 1/2$

- $EX = \mu$, $DX = \sigma^2$
- $X \sim N(0,1)$ 的 n 阶矩

$$EX^n = \begin{cases} (2k-1)!! & n = 2k \\ 0 & n = 2k-1 \end{cases} \tag{5.24}$$

- 可加性: 对于相互独立的 $X_i \sim N(\mu_i, \sigma_i^2)$,

$$X = \sum_i a_i X_i + b \sim N(\sum_i a_i\mu_i + b, \sum_i a_i^2\sigma_i^2) \tag{5.25}$$

4. 二维正态分布 $(X,Y) \sim N(\mu_1, \mu_2, \rho, \sigma_1^2, \sigma_2^2)$

$$f(x,y) = \frac{\exp\left[-Q(x,y)/2\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho}} \tag{5.26}$$

其中 $|\rho| < 1$, $\sigma_i > 0$; $Q(x,y)$ 是二次型

$$Q(x,y) = \frac{1}{1-\rho^2}\left[x^{*2} - 2\rho x^* y^* + y^{*2}\right] \tag{5.27}$$

标准化变量: $x^* = (x-\mu_1)/\sigma_1$, $y^* = (y-\mu_2)/\sigma_2$.

特别的 $N(0,0,\rho,1,1)$ 为二元 ρ-标准正态分布

- 边缘分布 $X_1 \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$
- 相关系数 $r_{X,Y} = \rho$
- X, Y 相互独立 $\Leftrightarrow \rho = 0$
- $X|Y = y \sim N(\mu', \sigma'^2)$
- $\mu' = E(X|Y) = \mu_1 + \sigma_1\rho(Y-\mu_2)/\sigma_2$
  - 非线性预测恰为线性预测, 线性回归的依据
- $\sigma'^2 = D(X|Y) = (1-\rho^2)\sigma_1^2$

5. m 维 Gauss 分布 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$
$\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{\mu}$, 其中 $\boldsymbol{Z}$ 的各分量独立 $Z_i \sim N(0,1)$ $(i = 1, \cdots, n)$ 构成的 m 维随机向量, $A_{m\times n}$ 为常数. 特别的, 一维 Gauss 分布是正态分布或常数.

- 分布密度: 对于行列式 $|\Sigma| > 0$ 的正态分布

$$\begin{aligned} f(\boldsymbol{x}) = &\left(\frac{1}{2\pi}\right)^{m/2} \frac{1}{\sqrt{|\Sigma|}} \times \\ &\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right] \end{aligned} \tag{5.28}$$

- 特别的, $m = 2$ 时, 二维正态分布

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{5.29}$$

- 期望 $E\boldsymbol{X} = \boldsymbol{\mu}$, 协方差矩阵 $\Sigma = AA^T$
- 特别的, $\mathrm{Rank}[A] = m$ 满秩时称为 m 维正态分布. 等价于 Σ 行列式非零/正定
- 以下等价于定义:
  - (正态分布时) 密度函数为 (5.28) 式
  - 满足 (3.10) 式特征函数定义

- $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma) \Leftrightarrow \forall \boldsymbol{a} \in \mathbb{R}^m$, $\boldsymbol{a}^T \boldsymbol{X}$ 服从一维 Gauss 分布 (正态或者常数)

- $B\boldsymbol{X} + \boldsymbol{b} \sim N(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^T)$
- $\boldsymbol{X}$ 各分量相互独立 $\Leftrightarrow \mathrm{Cov}(X_i, X_j) = 0$ 即 Σ 对角. 推广:
  - $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ 相互独立 $\Leftrightarrow \Sigma$ 分块对角
  - $A\boldsymbol{X}$ 与 $B\boldsymbol{X}$ 独立 $\Leftrightarrow A\Sigma B^T = 0$

6 随机过程

1. 随机过程: 概率空间 $(\Omega, \mathscr{F}, P)$ 中一族依赖参数 $t \in T$ 的随机变量 $X = \{X_t | t \in T\}$

2. 指标集 T

3. (样本) 轨道: 对于固定的基本事件 ω, 定义在 T 上的实值函数 $X_t(\omega)$,

6.1 有限维分布族

对于一个随机过程, 定义

$$F_{t_1, \cdots, t_k}(x_1, \cdots, x_k) = P(X_{t_i} \le x_i, i = 1, \cdots, k)$$

为过程的有限维分布族, 满足

1. 对称性: $\{(t_i, x_i)\}$ 的排序不影响取值

2. 相容性: $F_{\boldsymbol{t}, t_i}(\boldsymbol{x}, +\infty) = F_{\boldsymbol{t}}(\boldsymbol{x})$

- Kolmogorov 定理: 满足对称性和相容性的函数族总是某个随机过程使的有限维分布族

6.2 独立增量过程

1. 独立增量过程: 随机过程 $X = \{X_t\}$ 中, 任意 s 个互不相交的区间上的增量 $X_{m_i} - X_{n_i}$ $(i = 1, \cdots, s)$ 都相互独立

2. 时齐的独立增量过程: $X_{m+n} - X_m$ $(n > 0)$ 对于一切 m 同分布

6.3 随机徘徊

离散事件, 离散状态的随机过程.

1. (1-维) 简单随机徘徊: 概率空间 $(\Omega, \mathscr{F}, P)$ 上的独立同分布随机变量序列 $\{Z_n\}$ 满足

$$Z_n \sim \begin{pmatrix} -1 & 1 \\ q & p \end{pmatrix} \quad n = 1, 2, \cdots$$

其中 $q = 1 - p$. 令

$$X_n = X_0 + \sum_{i=1}^{n} Z_i$$

称 $X = \{X_n, n \ge 1\}$ 为 (1-维) 简单随机徘徊. 特别的 $p = q = 1/2$ 时候称为对称的简单随机徘徊

2. 分布性质: 令 $X_0 = x$

   - $EX_n = x + n(p - q)$
   - $DX_n = 4npq$
   - $\mathrm{Cov}(X_n, X_m) = 4pq \min(m, n)$

3. 轨道性质:
记 T_y^x 表示 x 出发首次到达 y 的时刻, $\phi(x) = P(T_d^x < T_c^x)$, 对于 $c + 1 \le x \le d - 1$

$$\phi(x) = p\phi(x+1) + q\phi(x-1)$$

考虑 $\phi(c) = 0$, $\phi(d) = 1$, 递推得到 ($p = q$ 渐近)

$$\phi(x) = \frac{1 - (q/p)^{x-c}}{1 - (q/p)^{d-c}}$$

   - $P(T_d^x < T_c^x) + P(T_c^x < T_d^x) = 1$, 即从 $[c, d]$ 出发, 一定到达边界
   - 若 $x > c$, 则

$$P(T_c^x \le \infty) = \begin{cases} (q/p)^{x-c} & p > 1/2 \\ 1 & p \le 1/2 \end{cases}$$

6.4 Poisson 过程与 Poisson 分布

连续时间, 离散状态的随机过程.
模型条件:

- 独立增量性, 时齐性

- 普通性:

$$P(\omega : N_{t+h}(\omega) - N_t(\omega) = n)$$
$$= \begin{cases} \lambda h + o(h) & n = 1 \\ o(h) & n \ge 2 \\ 1 - \lambda h + o(h) & n = 0 \end{cases}$$

6.4.1 Poisson 分布的性质

$X \sim P(\lambda)$ 或 Poisson_λ

$$P(X=k) = \frac{(\lambda)^k}{k!}\mathrm{e}^{-\lambda} \tag{6.1}$$

1. 可加性:

$$X_i \sim P(\lambda_i) \Rightarrow S_n = \sum_i^n X_i \sim P\left(\sum_i^n \lambda_i\right) \tag{6.2}$$

2. 随机选择下的不变性: 对于 $C \sim P(\lambda)$, 对于每个个体以保留概率 p 筛选得到 Y 个个体, 则 $Y \sim P(\lambda p)$

$$\sum_{n=k}^{\infty} C_n^k p^k q^{n-k} \frac{\lambda^n \mathrm{e}^{-\lambda}}{n!} = \frac{(\lambda p)^k}{k!}\mathrm{e}^{-\lambda p} \tag{6.3}$$

3. Poisson 近似定理: 对于二项分布 $B(n,p)$, $np=\lambda$

$$\lim_{n\to\infty} B(n, \lambda/n) = P(\lambda) \tag{6.4}$$

即:

$$\lim_{n\to\infty} C_n^k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \tag{6.5}$$

4. $EX = DX = \lambda$

5. 与二项式分布的关系: 对于独立的 $X_i \sim P(\lambda_i)$

$$\begin{aligned} & P(X_2=k|X_1+X_2=n) \\ = \ & C_n^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k} \end{aligned} \tag{6.6}$$

6. 与正态分布的关系, 对于 $X_\lambda \sim P(\lambda)$

$$\lim_{\lambda\to\infty} \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{D} N(0,1) \tag{6.7}$$

使用特征函数可得.

7. 复合 Poisson 分布: $N \sim P(\lambda)$, 独立同分布的 Y_i 以及

$$X = \sum_{i=1}^N Y_i$$

- $EX = EN\, EY_1 = \lambda EY_1$
- $DX = \lambda EY_1^2$

6.4.2 Poisson 过程

1. 强度 λ 的 Poisson 过程 $N = \{N_t, t \geq 0\}$:

   (a) $N_0 = 0$

   (b) 独立增量过程

   (c) $N_{s+t} - N_s \sim P(\lambda t)$

   强度 λ 的 Poisson 过程等价于初值为 0 的普通性时齐独立增量过程.

2. $EX = DX = \lambda t$, $\mathrm{Cov}(N_s, N_t) = \lambda \min(s,t)$

3. 在 $N_t = n$ 条件下, 对于 $u<t$: $N_u \sim B(n, u/t)$

6.5 Yule 过程

连续时间, 离散状态的随机过程. Yule 过程 N_t, 已知 $N_0 = n_0$, 记 $p_n(t) = P(N_t = n)$. 模型得到

$$p_n(t+h) = p_n(t)(1-n\lambda h) + p_{n-1}(t)(n-1)\lambda h + o(h)$$

可解得

$$p_n(t) = C_{n-1}^{n-n_0} \mathrm{e}^{-\lambda n_0 t}\left(1-\mathrm{e}^{-\lambda t}\right)^{n-n_0} \tag{6.8}$$

6.6 Gauss 过程

连续时间, 连续状态的随机过程.

定义: 对于 $X = \{X_t | t \in \mathbb{R}\}$, 任意有限维 $\boldsymbol{X} = (X_{t_1}, \cdots X_{t_n})^T \sim N(\boldsymbol{\mu_t}, \Sigma_{\boldsymbol{t}})$. 其中 $\boldsymbol{\mu_t}$, $\Sigma_{\boldsymbol{t}}$ 依赖时间 $\boldsymbol{t} = (t_1, \cdots, t_n)$

6.7 Brown 运动

连续时间, 连续状态的随机过程.

定义随机过程 $B = \{B_t | t \geq 0\}$: 为 Brown 运动

1. B 是独立增量过程

2. $\forall s \geq 0, t > 0$, 增量 $B_{s+t} - B_s \sim N(0, Dt)$

3. (可导出) 对于固定的 ω, $B_t(\omega)$ 关于时间连续

Brown 运动是 Gauss 过程.

特别的, 当 $D=1$ 时称为标准 Brown 运动 (默认)

6.7.1 Einstein 模型

1. 独立增量过程
2. 时齐, 且 $\sigma(t) \equiv E(B_{t+h} - B_h)^2$ 存在且连续
   - 于是有 $\sigma(t+s) = \sigma(t) + \sigma(s) \Rightarrow \sigma(t) = Dt$
3. 空间对称性 $EB_t = 0$

据此有结论, 一维 $B_t \sim N(0, Dt)$. 可从特征函数 $\varphi_t(\theta) = E\mathrm{e}^{\mathrm{i}\theta B_t}$ 入手得到它的偏微分方程

$$\frac{\partial \varphi_t}{\partial t} = -\frac{1}{2} D\theta^2 \varphi_t \tag{6.9}$$

6.7.2 Brown 运动的不变性

1. 定义等价于: B 是 Gauss 过程且 $EB_t = 0$, $E(B_s B_t) = \min(s, t)$
2. 平移不变性: $\{B_{t+a} - B_a | t \geq 0\}$ 是 Brown 运动
3. 尺度不变形: $\{B_{ct}/\sqrt{c} | t \geq 0\}$ 是 Brown 运动
4. $\{tB_{1/t} | t \geq 0\}$ 是 Brown 运动

6.7.3 Brown 运动的轨道

1. 轨道 $B_t(\omega)$ 关于 t 处处连续, 处处不可导

$$\begin{aligned} &P\left(\omega : \left|\frac{B_{t+\Delta t}(\omega) - B_t(\omega)}{\Delta t}\right| \leq C\right) \\ =&\Phi(C\sqrt{\Delta t}) - \Phi(-C\sqrt{\Delta t}) \xrightarrow{\Delta t \to 0} 0 \end{aligned} \tag{6.10}$$

2. 镜面对称性: $P(B_t \geq 0) = P(B_t < 0) = 1/2$
3. 定义首达时

$$T_a = \inf\{t > 0 | B_t = 1\} \tag{6.11}$$

   其分布

$$\begin{aligned} P(T_a \leq t) &= \sqrt{\frac{2}{\pi}} \int_{|a|/\sqrt{t}}^{\infty} \mathrm{e}^{-x^2/2} \,\mathrm{d}x \\ &= 2\left[1 - \Phi\left(\frac{|a|}{\sqrt{t}}\right)\right] \end{aligned} \tag{6.12}$$

   $a > 0$ 时候, 可以从 $P(B_t \geq a) = P(B_t \geq a | T_a \leq a)P(T_a \leq t) = P(T_a \leq t)/2$ 得到

4. $P(T_a < \infty) = 1$: 几乎所有轨道都能在有限时间内经过指定点
5. $ET_a = +\infty$: 经过指定点的期望时间是无穷
6. 最远距离的分布

$$\begin{aligned} X &= \max_{0 \leq s \leq t} B_s \\ f(x) &= \frac{2}{\sqrt{2\pi t}} \mathrm{e}^{-x^2/2t} \quad (x \geq 0) \end{aligned} \tag{6.13}$$

7 极限定理

1. 依分布收敛: 在 $F_X(x)$ 上的一切连续点上

$$X_n \xrightarrow{D} X \Leftrightarrow \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{7.1}$$

   - Polya 定理: $X_n \xrightarrow{D} X \Rightarrow F_{X_n}(x)$ 一致收敛到 $F_X(x)$

2. 依概率收敛: $\forall \varepsilon > 0$

$$X_n \xrightarrow{P} X \Leftrightarrow \lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0 \tag{7.2}$$

   - $g(x)$ 连续, $\xi_n \xrightarrow{P} \xi$, 则 $g(\xi_n) \xrightarrow{g} (\xi)$ 可以推广到随机向量 $\boldsymbol{\xi}$
   - $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$
   - $X_n \xrightarrow{P} a \Leftrightarrow X_n \xrightarrow{D} a$ (a 常数)

3. 依概率 1 收敛:

$$X_n \xrightarrow{\text{a.s.}} X \Leftrightarrow P\left(\lim_{n\to\infty} X_n = X\right) = 1 \tag{7.3}$$

7.1 大数定律

$\{X_n\}$ 满足 (弱) 大数定律, 即

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i \xrightarrow{P} 0 \tag{7.4}$$

1. Chebyshev 大数定律: $\{X_n\}$ 两两不相关, 且方差一致有界 $(\forall i, DX_i \leq C)$, 则 $\{X_n\}$ 满足大数定律
   (a) 引理 (Chebyshev 不等式): 对于有限方差 X, $\forall \varepsilon > 0$

$$P\left(|X - EX| \geq \varepsilon\right) \leq \frac{DX}{\varepsilon^2} \tag{7.5}$$

   (使用范围内 $g \leq (x - EX)^2/\varepsilon^2$ 进行缩放)

   (b) 均值 $\overline{X} = \sum_i X_i/n$ 分布的估计:

$$D\overline{X} = \frac{\sigma^2}{n}; \quad P(|\overline{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n} \tag{7.6}$$

2. Khintchine 大数定律: $\{X_n\}$ 独立同分布, $EX_i = \mu$, 则 $\{X_n\}$ 满足大数定律.

   - Chebyshev 大数定律是它的推广. 可以直接通过特征函数得到

3. Bernoulli 大数定律: 以概率 p 独立重复 n 次的 Bernoulli 概型中, μ_n 表示发生次数, $\mu_n/n \xrightarrow{P} p$

   - Bernoulli 概型: $P(X=1)=p$, $P(X=0)=q=1-p$ 的二值分布, 进行独立重复实验
   - 满足强大数定律 $\mu_n/n \xrightarrow{\text{a.s.}} p$

7.2 中心极限定理

$\{X_n\}$ 满足中心极限定理, 即

$$S_n^* = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n EX_i}{\sqrt{D\left(\sum_{i=1}^n X_i\right)}} \xrightarrow{D} N(0,1) \tag{7.7}$$

或 $\sum_i X_i$ 近似满足正态分布

1. Levy-Lindeberg 中心极限定理: $\{X_n\}$ 独立同分布, 且具有有限的数学期望 $(EX_i = \mu)$ 和方差 $(DX_i = \sigma^2 \neq 0)$, 则 $\{X_n\}$ 满足中心极限定理
   (使用特征函数证明)

2. 推论 De Moivre-Laplace 定理: 概率 p 的 Bernoulli 实验中, 发生次数 μ_n $(\sim B(n,p))$

$$\frac{\mu_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1) \tag{7.8}$$

   (与 (6.4) 式区分)

3. Liapunov 定理: 独立随机变量 $\{X_i\}$, $EX_i = \mu_i$, $DX_i = \sigma_i^2$, 记 $B_n^2 = \sum \sigma_i^2$, 若满足 Liapunov 条件, 即 $\exists \delta > 0$

$$\lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{i=1}^n E|X_i - \mu_i|^{2+\delta} = 0 \tag{7.9}$$

   则由

$$S_n = \sum_{i=1}^n X_i - N\left(\sum_{i=1}^n \mu_i, B_n^2\right) \xrightarrow{D} 0 \tag{7.10}$$

   - (7.9) 主要保证: $\forall \tau > 0$

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} \frac{|X_i - \mu_i|}{B_i} > \tau\right) = 0 \tag{7.11}$$

     即各项对于总和的极限分布不产生显著影响