数学期望与条件数学期望

蒋辉

南京航空航天大学数学系

September 16, 2019

- 概率空间 $(\Omega, \mathscr{F}, P)$:
  - 样本空间 Ω: 随机试验所有可能的结果组成的集合.
  - 事件域 $\mathscr{F}$: Ω 中某些子集组成的集合, 满足
    (1). $\Omega \in \mathscr{F}$;
    (2). 若 $A \in \mathscr{F}$, 则 $A^c \in \mathscr{F}$;
    (3). 若 $A_n \in \mathscr{F}$, $n = 1, 2, \cdots$, 则 $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$.
  - 概率 $P : \mathscr{F} \to \mathbb{R}^+$ 满足
    (1). $\forall A \in \mathscr{F}, 0 \le P(A) \le 1$;
    (2). $P(\Omega) = 1$;
    (3). 对于 $A_n \in \mathscr{F}$, $n = 1, 2, \cdots$, 有 $P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

■ 称 $X:\Omega\to\mathbb{R}$ 为随机变量, 若 $\forall x\in\mathbb{R}$, 有

$$\left\{X\le x\right\}\in\mathscr{F}.$$

更进一步, 称

$$F(x)=P\Big(X\le x\Big)$$

为 X 的分布函数.

■ 分布函数 $F(x)$ 满足:

- $F(x)$ 为非降函数;
- $F(-\infty)=0,\ F(+\infty)=1$;
- $F(x)$ 为右连续.

■ 注: 随机向量 $\Big(X_1,\cdots,X_n\Big)$ 的分布函数为:

$$F(x_1,\cdots,x_n)=P\Big(X_1\le x_1,\cdots,X_n\le x_n\Big),\quad x_1,\cdots,x_n\in\mathbb{R}.$$

独立性

- 独立事件族: $\left(\Omega, \mathscr{F}, P\right)$ 为概率空间, 称 $\mathscr{G} \subset \mathscr{F}$ 为独立事件族, 若对任意的 $A_1, \cdots, A_n \in \mathscr{G}$, 有 $P\left(\bigcup_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$.
- 独立随机变量族: $\mathbb{T}$ 为指标集, 称 $\left\{X_t, t \in \mathbb{T}\right\}$ 为独立随机变量族, 若对任意的 $n \in \mathbb{N}$, 以及 $t_1, \cdots, t_n \in \mathbb{T}$, 有

$$P\left(X_{t_1} \le x_1, \cdots, X_{t_n} \le x_n\right) = \prod_{i=1}^n P\left(X_{t_i} \le x_i\right), \quad x_1, \cdots, x_n \in \mathbb{R}.$$

■ 定义积分 $\int_a^b f(x)dF(x)$:

若 $f(x), F(x)$ 为 $[a,b]$ 上的两个函数, 在 $[a,b]$ 上插入分点:

$$a = x_0 < x_1 < \ldots < x_n = b.$$

在每一个小区间 $[x_{i-1}, x_i]$ 中任选一点 ξ_i. 作和式:

$$\sigma_n = \sum_{i=1}^{n} f(\xi_i)(F(x_i - F(x_{i-1}))).$$

若当 $\lambda_n := \max_{1\le i\le n}\{x_i - x_{i-1}\} \to 0$ 时, $\sigma_n \to \sigma$, 则称 $f(x)$ 在 $[a,b]$ 上关于 $F(x)$ R-S可积, 且记

$$\sigma = \int_a^b f(x)dF(x).$$

Theorem

若 $f(x)$ 为 $[a,b]$ 上的连续函数，$F(x)$ 为 $[a,b]$ 上的单调函数，则 $f(x)$ 关于 $F(x)$ 在 $[a,b]$ 上可积.

特别地，当 $F(x)=x$ 时，R-S积分为黎曼积分.

- **基本性质：**

(1). $\int_a^b(\alpha f(x)+\beta g(x))\,dF(x)=\alpha\int_a^b f(x)dF(x)+\beta\int_a^b g(x)dF(x)$

(2). $\int_a^b f(x)d\,(\alpha F(x)+\beta G(x))=\alpha\int_a^b f(x)dF(x)+\beta\int_a^b f(x)dG(x)$

(3). $\int_a^b f(x)dF(x)=\int_a^c f(x)dF(x)+\int_c^b f(x)dF(x)$

(4). 若 $F(x)$ 为 $[a,b]$ 上的增函数，则

$$\left|\int_a^b f(x)dF(x)\right|\leq\int_a^b|f(x)|dF(x).$$

- **三个重要的例子:**

(1). 若 $f(x)$ 为 $[a,b]$ 中的点 c 处连续, 令

$$F_c(x) = \begin{cases} 1, & x \geq c; \\ 0, & x < c. \end{cases}$$

则 $\int_a^b f(x)dF_c(x) = f(c)$.

(2). 若 $F(x)$ 为阶梯函数(跳跃函数), 且在 x_i, $i = 1,2,...,n$ 处的振幅为 p_i, 则

$$\int_a^b f(x)dF(x) = \sum_{i=1}^n f(x_i)p_i.$$

(3). 若 $F'(x) = p(x)$, 则

$$\int_a^b f(x)dF(x) = \int_a^b f(x)p(x)dx.$$

■ 定义:

- 设 X 的分布函数为 $F_X(x)$, 且 $\int_{-\infty}^{\infty}|x|dF_X(x)<\infty$, 则称

$$EX=\int_{-\infty}^{\infty}xdF_X(x)$$

为 X 的数学期望.

- 若 X 为离散型随机变量, 即

$$P(X=x_i)=p_i,\quad i=1,2,\ldots$$

则 $EX=\sum_{i=1}^{\infty}x_ip_i$.

- 若 X 为连续型随机变量且密度函数为 $p(x)$, 则

$$EX=\int_{-\infty}^{\infty}xp(x)dx.$$

■ 随机变量函数的期望:

- 设 X 的分布函数为 $F_X(x)$, $g(x)$ 为一元Borel函数, 且 $Y=g(X)$, 则

$$EY=\int_{\mathbb{R}} g(x)dF(x).$$

- 若 $(X_1,...X_n)$ 的分布函数为 $F(x_1,...,x_n)$, $g(x_1,...,x_n)$ 为 n 元Borel函数, 且 $Y=g(X_1,...,X_n)$, 则

$$EY=\int_{\mathbb{R}^n} g(x_1,...,x_n)dF(x_1,...,x_n).$$

■ 期望的性质:

- 若 EX 存在, 则

$$|EX| \leq E|X|;$$

- Hölder 不等式: 对于任意的对偶数 p, q $(p > 0, q > 0, \frac{1}{p} + \frac{1}{q} = 1)$, 有

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

  特别地, $E|XY| \leq \sqrt{EX^2 EY^2}$.

- 对于 $k > 0$, 若 $E|X|^k < \infty$, 则 $E|X|^r < \infty$, 其中 $0 \leq r \leq k$.

■ 定义: 给定概率空间 $(\Omega, \mathfrak{F}, P)$, 令 $B \in \mathfrak{F}$ 且 $P(B) > 0$. $F(x|B)$ 为 X 关于 B 的条件分布函数, 则 X 关于 B 的条件期望为:

$$E(X|B) = \int_{\mathbb{R}} x dF(x|B).$$

■ 注:

- 若 X 为取值 $x_i, i = 1, 2, \ldots$ 的离散型随机变量, 则

$$E(X|B) = \sum_{i=1}^{\infty} x_i P(X = x_i | B).$$

- 若 X 为连续型随机变量, 则 $E(X|B) = \int_{-\infty}^{\infty} x p(x|B) dx$.

- 若 X, B 相互独立 ($\forall x$, $\{X \le x\}$ 与 B 独立), 则 $E(X|B) = EX$.
- 令 $Q(\cdot) = P(\cdot|B)$, 则 Q 为 $(\Omega, \mathfrak{F})$ 上的概率, 且

$$E(X|B) = E_Q(X).$$

■ 例: 设 $X \sim \varepsilon(\lambda)$, $B = \{X > t_0\}, t_0 > 0$. 求 $E(X|B)$.

■ 一些重要公式:

- 全期望公式: 令 $B_i, i=1,2,\ldots$ 为样本空间 Ω 的划分 $(\Omega=\sum_{i=1}^{\infty} B_i)$, 且 $P(B_i)>0, i=1,2,\ldots$. 则

$$EX=\sum_{i=1}^{\infty} P(B_i)E(X|B_i).$$

例子: 设r.v.s $\{X_n, n\geq 1\}$ i.i.d., r.v. ξ 取正整数值且与 $\{X_n, n\geq 1\}$ 相互独立, 求

$$E(X_1+\ldots+X_{\xi}), \quad D(X_1+\ldots+X_{\xi}).$$

- 条件全期望公式: 令 $B_i, i=1,2,\ldots$ 同上, 且且 $P(AB_i)>0, i=1,2,\ldots$. 则

$$E(X|A)=\sum_{i=1}^{\infty} P(B_i|A)E(X|AB_i).$$

注: 令 $Q(\cdot)=P(\cdot|B)$.

- 令 X 为r.v., $A, B \in \mathfrak{F}$, $P(B) > 0$. 则

$$E(I_A) = P(A), \quad E(I_A|B) = P(A|B), \quad E(XI_B) = P(B)E(X|B),$$

其中

$$I_A(\omega) = \begin{cases} 1, & \omega \in A; \\ 0, & \omega \notin A. \end{cases}$$

■ 定义: 令 $g(\cdot)$ 为 *Borel*-函数. 在 $Y = y$ 的条件下, $g(X)$ 的条件数学期望为

$$\begin{aligned} E\left(g(X)|y\right) &= E\left(g(X)|Y = y\right) \\ &= \int_{\mathbb{R}} g(x)dF_{X|Y}(x|y). \end{aligned}$$

- 若 (X, Y) 为离散型 $r.v.$, 且 $P(Y = y) > 0$, 则

$$E(g(X)|y) = \sum_i g(x_i)P(X = x_i|Y = y).$$

- 若 (X, Y) 为连续型 $r.v.$, 则

$$E(g(X)|y) = \int_{\mathbb{R}} g(x)f_{X|Y}(x|y)dx.$$

- **注**: 令 $g(\cdot)$ 为 $Borel$-函数, 则
  - $E(g(X)|y)$ 为关于 y 的函数;
  - $E(g(X)|Y) := E(g(X)|y)_{y=Y}$ 称为 $g(X)$ 关于 Y 的条件数学期望;
  - $E(g(X)|Y)$ 是 Y 的函数.

例:

- 令 $X \sim U(0,1)$, $Y \sim U(X,1)$, 求 $E(Y|X)$.
- 令 $(X,Y) \sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 求

$$E(X|Y), \quad E(Y|X), \quad E(X^2|Y), \quad E(Y^2|X).$$

提示:

$$X|Y=y \sim N\left(\mu_1+\frac{\rho\sigma_1}{\sigma_2}(y-mu_2),\sigma_1^2(1-\rho^2)\right)$$

与

$$Y|X=x \sim N\left(\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-mu_1),\sigma_2^2(1-\rho^2)\right).$$

- 条件数学期望的重要性质
  - 若 $X \geq 0$, 则 $E(X|Y) \geq 0$;
  - $\forall a, b \in \mathbb{R}$,
  $$E(aX + bY|Z) = aE(X|Z) + bE(Y|Z);$$
  - 若 X, Y 相互独立, 则 $E(X|Y) = EX$;
  - $E(Xg(Y)|Y) = g(Y)E(X|Y)$, $E(g(Y)|Y) = g(Y)$;
  - 重期望公式
  $$E(E(X|Y)) = EX;$$
  - $E(E(X|Y)g(Y)) = E(Xg(Y))$;
- 例: 令 $X \sim U(0,1)$, $Y \sim U(X,1)$, 求 EY.

- 条件数学期望的本质–投影

$$E(X-g(Y))^2 = E(X-E(X|Y))^2 + E(E(X|Y)-g(Y))^2.$$

- 条件数学期望的重要性质(续)
  - $E\big(E(X|Z)\big|Y,Z\big) = E(X|Z);$
  - $E\big(E(X|Y,Z)\big|Z\big) = E(X|Z).$

例:

- 设 $r.v.X, Y$ 具有二阶矩, 若 $E(X|Y) = EX$, 证明 $Cov(X, Y) = 0$.
- 设 $r.v.X, Y$ 具有二阶矩, 若 $E(X|Y) = Y, E(Y|X) = X$, 证明 $X = Y$.

- 定义: 设 X 的分布函数 (d.f.)为 $F(x)$, 则 X 的特征函数 (c.f.) 定义为

$$\varphi_X(t) := Ee^{itX} = \int_{\mathbb{R}} e^{itx} dF(x), \quad t \in \mathbb{R}.$$

- 注:
  - $\varphi_X(t) = \int_{\mathbb{R}} \cos tx dF(x) + i \int_{\mathbb{R}} \sin tx dF(x)$;
  - 若 X 为离散型 r.v., 且分布律为 $P(X = x_j) = p_j$, $j = 1, 2, \ldots$, 则

$$\varphi_X(t) = \sum_j e^{itx_j} p_j;$$

  - 若 X 为连续型 r.v., 且密度函数为 f, 则

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx.$$

- 重要分布的特征函数:
  - $b(1,p)$ 的特征函数为
$$\varphi(t) = pe^{it} + q;$$
  - $\pi(\lambda)$ 的特征函数为
$$\varphi(t) = e^{\lambda(e^{it}-1)};$$
  - $N(0,1)$ 的特征函数为
$$\varphi(t) = e^{-t^2/2}.$$

■ 特征函数的性质:

- $\varphi_X(0) = 1,\ |\varphi_X(t)| \le 1,\ \varphi_X^*(t) = \varphi_X(-t)$;
- $\varphi_X(\cdot)$ 在 $\mathbb{R}$ 上一致连续;
- 若 $X_1, \ldots, X_n$ 相互独立, 则

$$\varphi_{\sum_{j=1}^n X_j}(t) = \varphi_{X_1}(t) \ldots \varphi_{X_n}(t);$$

- 若 $E|X|^n < \infty$, 则 $\varphi_X(t)$ 关于 $t \in \mathbb{R}$ 为 n 阶可导, 且对于 $k \le n$

$$EX^k = i^{-k} \varphi_X^{(k)}(0);$$

- 令 $Y = aX + b$, 则

$$\varphi_Y(t) = e^{itb} \varphi_X(at).$$

例: 求 $b(n,p), N(\mu,\sigma^2)$ 的特征函数.

■ 唯一性定理: 对于 r.v.s. X_1, X_2, 其分布函数相等等价于特征函数相等, i.e.

$$F_{X_1}(x) \equiv F_{X_2}(x) \Leftrightarrow \varphi_{X_1}(x) \equiv \varphi_{X_2}(x).$$

■ 逆转公式: 若 r.v. X 的特征函数 φ 绝对可积, i.e.

$$\int_{\mathbb{R}} |\varphi(t)| dt < +\infty,$$

则 X 为连续型 r.v., 其分布函数 F 处处可导, 导函数 f 有界连续, 且

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt.$$

■ 分布函数的再生性:

- $X_1 \sim b(n_1,p), X_2 \sim b(n_2,p)$, 且 X_1 与 X_2 相互独立, 则

$$X_1 + X_2 \sim b(n_1 + n_2, p).$$

- $X_1 \sim \pi(\lambda_1), X_2 \sim \pi(\lambda_2)$, 且 X_1 与 X_2 相互独立, 则

$$X_1 + X_2 \sim \pi(\lambda_1 + \lambda_2).$$

- $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2))$, 且 X_1 与 X_2 相互独立, 则

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

■ 多元特征函数: 令

$$\vec{X}=\begin{pmatrix} X_1\\ X_2\\ \vdots\\ X_n \end{pmatrix},\quad \vec{x}=\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n \end{pmatrix},\quad \vec{t}=\begin{pmatrix} t_1\\ t_2\\ \vdots\\ t_n \end{pmatrix}.$$

$\vec{X}$ 的特征函数为

$$\begin{aligned}\varphi_X(\vec{t})&=\varphi(t_1,t_2,\cdots t_n)\\ &=Ee^{i\vec{t}\vec{X}}=Ee^{i\sum_{k=1}^{n}t_iX_i}\\ &=\int\int\cdots\int_{\mathbb{R}^n}e^{i\vec{t}\vec{x}}dF(\vec{x}).\end{aligned}$$

- 多元特征函数的性质:
  - $\varphi(t_1,\cdots,t_n)$ 在 $\mathbb{R}^n$ 中一致连续, 且

$$\left|\varphi(t_1,\cdots t_n)\right| \le \varphi(0,\cdots,0)=1, \quad \varphi_X^*(t_1,\cdots t_n)=\varphi_X(-t_1,\cdots,-t_n).$$

  - 若 $\vec{X}$ 的特征函数为 $\varphi_{\vec{X}}(t_1,\cdots,t_n)$, 则 $\vec{Y}=C_{m\times n}\vec{X}$ 的特征函数为

$$\varphi_{\vec{Y}}(\vec{s})=\varphi_{\vec{X}}(C^{\tau}\vec{s}).$$

  - 若 $\vec{X}$ 的特征函数为 $\varphi_{\vec{X}}(t_1,\cdots,t_n)$, 各分量$X_k$ 的特征函数为 $\varphi_{X_k}(t_k)$, 则

$$X_1,...,X_n \text{ 相互独立} \Leftrightarrow \varphi_{\vec{X}}(t_1,\cdots,t_n)=\prod_{k=1}^{n}\varphi_{X_k}(t_k).$$

- 若 $E|X_1^{k_1}\cdots X_n^{k_n}|<\infty$, 则

$$EX_1^{k_1}\cdots X_n^{k_n}=i^{-\sum_{j=1}^{n}k_j}\frac{\partial^{k_1+\cdots+k_n}\varphi(t_1,\cdots,t_n)}{\partial t_1^{k_1}\cdots\partial t_n^{k_n}}\bigg|_{t_1=\cdots=t_n=0}.$$

7. 作业

(1) 若 $X \sim N(\mu, \sigma^2)$, $Y \sim \pi(\lambda)$ 且相互独立, 求

- $E|X-\mu|^k$, $k \in \mathbb{N}$, 以及 $Ee^{\lambda X}$;
- $X+Y$ 的特征函数.

(2). 对于取值于区间 (a,b) 内的 $r.v.$ X, 试证明

$$a \le EX \le b, \quad DX \le \frac{(b-a)^2}{4}.$$

(3). 令 $r.v.$ X 的分布函数为 F, 且数学期望 EX 存在, 试利用分布积分证明

$$EX = \int_0^{\infty} 1 - F(x)dx - \int_{-\infty}^{0} F(x)dx.$$

特别地, 对于非负 $r.v.$ X, 有

$$EX = \int_0^{\infty} 1 - F(x)dx.$$

(4). 令 (X_1, X_2) 的概率密度为

$$f(x_1, x_2) = 2e^{-x_1 - x_2}, \quad 0 < x_1 < x_2.$$

求 $E\left(X_1 e^{X_2^2} \middle| X_2\right)$.

(5). 某公共汽车站在 $[0, T]$ 时间内来到的乘客批数 $\xi \sim \pi(\lambda t)$. 且每批来到的乘客数是随机变量, 满足来 n 个的概率为 p_n, $n = 0, 1, 2, \ldots$ 求 $[0, T]$ 内来到的乘客数 Y 的期望与方差.

(6). 若r.v. X 的特征函数 $\varphi(t) = \frac{e^{it}(1-e^{int})}{n(1-e^{it})}$, 求 X 的分布律.

(7). 设r.v. X 服从柯西分布, 密度函数为

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

- 试求 X 的特征函数;
- 证明柯西分布具有再生性.