应用随机过程

张 波　　商 豪

中国人民大学　统计学院

第1章　预备知识

- 1.1 概率空间
- 1.2 随机变量与分布函数;
- 1.3 数字特征、矩母函数与特征函数
- 1.4 收敛性
- 1.5 独立性与条件期望

§1.1 概率空间

随机试验是概率论的基本概念，试验的结果事先不能准确地预言，但具有如下三个特性:

(1)可以在相同的条件下重复进行;

(2)每次试验的结果不止一个,但预先知道试验的所有可能的结果;

(3)每次试验前不能确定哪个结果会出现.

样本点(ω)：	**随机试验的基本结果**
样本空间(Ω)：	**随机试验所有可能结果组成的集合**
基本事件：	**Ω中的样本点ω**
必然事件：	**样本空间Ω**
不可能事件：	**空集$\emptyset$**
事件：	**由基本事件组成的Ω中的子集A**

定义 1.1.1 设Ω是一个样本空间(或任意一个集合),$\mathcal{F}$是Ω的某些子集组成的集合族. 如果满足:

(1) $\Omega \in \mathcal{F}$;

(2) 若$A \in \mathcal{F}$, 则$A^c = \Omega \setminus A \in \mathcal{F}$;

(3) 若$A_n \in \mathcal{F}, n = 1, 2, \cdots$, 则$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$;

则称$\mathcal{F}$为σ代数. $(\Omega, \mathcal{F})$称为**可测空间**, $\mathcal{F}$中的元素称为**事件**.

注：如果$\mathcal{F}$是事件的σ代数，则(1) $\emptyset \in \mathcal{F}$; (2) **当**$A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

以Ω的某些子集为元素的集合称为(Ω上的)集类. 对于Ω上的任一非空集类$\mathcal{C}$，存在包含$\mathcal{C}$的最小σ代数，即$\bigcap\{\mathcal{H}|\mathcal{H}$为包含$\mathcal{C}$的$\sigma$代数$\}$，称为由$\mathcal{C}$生成的$\sigma$代数，记为$\sigma(\mathcal{C})$.

定义 1.1.2 设$\Omega = \mathbb{R}$. 由所有半无限区间$(-\infty, x)$生成的σ代数(即包含集族 $\{(-\infty, x), x \in \mathbb{R}\}$的最小$\sigma$代数)称为$\mathbb{R}$上的 Borel σ**代数**, 记为$\mathcal{B}(\mathbb{R})$, 其中的元素称为 Borel 集合. 类似地，可定义$\mathbb{R}^n$上的Borel σ**代数**$\mathcal{B}(\mathbb{R}^n)$.

定义 1.1.3 设$(\Omega, \mathcal{F})$是可测空间，$P(\cdot)$ 是定义在$\mathcal{F}$上的实值函数. 如果

(1) 任意$A \in \mathcal{F}, 0 \le P(A) \le 1$;

(2) $P(\Omega) = 1$;

(3) 对$\mathcal{F}$中两两互不相容事件$A_1, A_2, \cdots$,

(即当$i \neq j$时, $A_i \cap A_j = \emptyset$)有

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

则称P是$(\Omega, \mathcal{F})$上的**概率**, $(\Omega, \mathcal{F}, P)$称为**概率空间**, $P(A)$称为事件A的概率.

概率常用性质:

(1) 若$A, B \in \mathcal{F}$,则$P(A \cup B) + P(A \cap B) = P(A) + P(B)$.

(2) 若$A, B \in \mathcal{F}$,且$A \subset B$, 则$P(A) \leq P(B)$ (单调性).

(3) 若$A, B \in \mathcal{F}$,且$A \subset B$, 则$P(B - A) = P(B) - P(A)$.

(4) 若$A_n \in \mathcal{F}, n = 1, 2, \ldots$, 则$P(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} P(A_n)$

(5) 若$A_n \in \mathcal{F}$且$A_n \uparrow A$，即 $A_n \subset A_{n+1}, n = 1, 2, \ldots,$ 且 $\bigcup_{n=1}^{\infty} A_n = A$，则$P(A) = \lim_{n\to\infty} P(A_n)$ （下连续）.

(6) 若$A_n \in \mathcal{F}$且$A_n \downarrow A$，即 $A_{n+1} \subset A_n, n = 1, 2, \ldots$ 且$\bigcap_{n=1}^{\infty} A_n = A$,则$P(A) = \lim_{n\to\infty} P(A_n)$ （上连续）.

完备概率空间： P-零集（即零概率事件）的每个子集仍为事件的概率空间$(\Omega, \mathcal{F}, P)$.

概率空间完备化：令$\mathcal{N}$代表Ω的所有P-零集的子集的全体.由$\{\mathcal{F}, \mathcal{N}\}$ 生成的σ代数（即包含$\mathcal{F}$和$\mathcal{N}$的最小σ代数）称为$\mathcal{F}$的完备化，记为 $\overline{\mathcal{F}}$. $\overline{\mathcal{F}}$ 中的每个集合B 都可以表为$B = A \cup N$其中$A \in \mathcal{F}$，$N \in \mathcal{N}$，且$A \cap N = \emptyset$. 定义

$$\bar{P}(B) = \bar{P}(A \cup N) = P(A).$$

则P就被扩张到$\overline{\mathcal{F}}$上.

定义 1.1.4 设$\{A_n, n\geq 1\}$为一集合序列.令

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k; \qquad \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$$

分别称其为$\{A_n\}$的上极限和下极限(上极限有时也记为$\{A_n, i.o.\}$).
显然有

$$\limsup_{n\to\infty} A_n = \{w|w\text{属于无穷多个}A_n\} = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n$$

$$\liminf_{n\to\infty} A_n = \{w|w\text{至多不属于有限多个}A_n\} = \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n$$

从而恒有$\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$.

若$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$,则称$\{A_n\}$的极限存在,并用$\lim_{n\to\infty} A_n$表示,即令$\lim_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$.

特别地，若对每个n，有$A_n \subset A_{n+1}$(相应地，$A_n \supset A_{n+1}$)，则称$\{A_n\}$为单调增(相应地，单调降).对单调增或单调降序列 $\{A_n\}$，我们分别令$A = \bigcup_n A_n$或$A = \bigcap_n A_n$，称A为$\{A_n\}$的极限，通常记为$A_n \uparrow A$或$A_n \downarrow A$.

例 1.1.5 设有某人在反复地投掷硬币，观察硬币朝上的面是正面或反面. $\Omega = \{$所有由投掷结果“正面”和“反面”组成的序列$\}$，$\mathcal{F} = \{\Omega$的所有子集$\}$，记A_n为第n次投掷的是“正面”的事件，则

$$\limsup_{n\to\infty} A_n = \{\text{有无限多个投掷结果是“正面”}\}$$
$$\liminf_{n\to\infty} A_n = \{\text{除有限多个外，投掷结果都是“正面”}\}$$

定义 1.1.6 设$(\Omega, \mathcal{F})$是可测空间，$P(\cdot)$ 是定义在$\mathcal{F}$上的实值函数.如果

(1) $P(\Omega)=1$;

(2) $\forall A\in\mathcal{F}, 0\leq P(A)\leq 1$;

(3) 对两两互不相容事件$A_1, A_2, \cdots$，(即当$i\neq j$时，$A_i\cap A_j=\emptyset$)有

$$P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$$

则称P是$(\Omega,\mathcal{F})$上的**概率**，$(\Omega,\mathcal{F},P)$称为**概率空间**，$\mathcal{F}$中的元素称为**事件**，$P(A)$称为事件A的概率.

事件的概率有如下性质：

(1) 若$A,B\in\mathcal{F}$,则$P(A\cup B)+P(A\cap B)=P(A)+P(B)$.

(2) 若$A,B\in\mathcal{F}$,且$A\subset B$，则$P(B-A)=P(B)-P(A)$(可减性).

(3) 若$A,B\in\mathcal{F}$,且$A\subset B$，则$P(A)\le P(B)$(单调性).

(4) 若$A_n\in\mathcal{F},n\ge 1$, 则$P(\bigcup_{n\ge1}A_n)\le\sum_{n\ge1}P(A_n)$.

(5) 若$A_n\in\mathcal{F}$且$A_n\uparrow A\in\mathcal{F}$，则 $P(A)=\lim_{n\to\infty}P(A_n)$ (从下连续).

(6) 若$A_n\in\mathcal{F}$且$A_n\downarrow A\in\mathcal{F}$，则 $P(A)=\lim_{n\to\infty}P(A_n)$ (从上连续).

如果概率空间$(\Omega,\mathcal{F},P)$的P零测集(即零概率事件) 的每个子集仍为事件，则称之为完备的概率空间. 为了避免P零测集的子集不是事件的情形出现，我们把概率测度完备化. 令$\mathcal{N}$代表Ω的所有P零测集的子集的全体，由$\{\mathcal{F},\mathcal{N}\}$生成的$\sigma$代数 (即包含$\mathcal{F}$和$\mathcal{N}$的最小$\sigma$代数)称为$\mathcal{F}$的完备化，记为 $\overline{\mathcal{F}}$. $\overline{\mathcal{F}}$ 中的每个集合B 都可以表为$B=A\cup N$，其中$A\in\mathcal{F}$, $N\in\mathcal{N}$, 且$A\cap N=\emptyset$. 定义

$$\bar{P}(B)=\bar{P}(A\cup N)=P(A)$$

则P就被扩张到$\overline{\mathcal{F}}$上.

容易验证，$\bar{P}$是$\overline{\mathcal{F}}$上的概率测度，集函数$\bar{P}$称为P的完备化. 本书总假定P是完备的概率测度.

§1.2 随机变量和分布函数

定义 1.2.1 设$(\Omega, \mathcal{F}, P)$是(完备的)概率空间，X是定义在Ω 上取值于实数集 $\mathbb{R}$的函数，如果对任意实数$x \in \mathbb{R}$，$\{\omega : X(\omega) \le x\} \in \mathcal{F}$，则称$X(\omega)$是$\mathcal{F}$上的**随机变量**，简称为随机变量.

$$F(x) = P(\omega : X(\omega) \le x), \quad -\infty < x < \infty$$

称为随机变量X的**分布函数**.

GoBack
FullScreen
Close
Quit

如果存在函数$f(x)$，满足

$$F(x)=\int_{-\infty}^{x} f(t)dt$$

则称$f(x)$为随机变量X或其分布函数$F(x)$的分布密度.

如果X具有分布密度，则称X为连续型随机变量；如果X最多以正概率取可数多个值,则称X为离散型随机变量.

定义 1.2.2 两个随机变量X与Y，如果满足$P(\omega\in\Omega:X(\omega)\neq Y(\omega))=0$，则称它们是等价的.

注：对于两个等价的随机变量，我们视为同一.

定理 1.2.3 下列命题等价:

(1) X是随机变量;

(2) $\{\omega:X(\omega)\geq a\}\in\mathcal{F},\quad\forall\ a\in\mathbb{R}$;

(3) $\{\omega : X(\omega) > a\} \in \mathcal{F}, \quad \forall\ a \in \mathbb{R}$;

(4) $\{\omega : X(\omega) < a\} \in \mathcal{F}, \quad \forall\ a \in \mathbb{R}$.

注：为简单起见，习惯上将$\{\omega : X(\omega) \geq a\}$**记为**$\{X \geq a\}$,

其他类似记号自明.

定理 1.2.4 (1) 若X, Y是随机变量, 则$\{X < Y\}, \{X \leq Y\}, \{X = Y\}$及$\{X \neq Y\}$ 都属于$\mathcal{F}$;

(2) 若X, Y是随机变量，则$X \pm Y$与XY亦然;

(3) 若$\{X_n\}$是随机变量序列,则$\sup_n X_n$, $\inf_n X_n$, $\limsup_{n\to\infty} X_n$ 和 $\liminf_{n\to\infty} X_n$都是随机变量.

映射$\mathbf{X} : \Omega \to \mathbb{R}^d$**，表示为**$\mathbf{X} = (X_1, \cdots, X_d)$**，若对所有的**$k, 1 \leq k \leq d$, X_k**都是随机变量，则称X为随机向量**.

复值随机变量Z定义为两个实值随机变量X和Y的线性组合$X+iY$.

给定随机变量X，可以生成Ω上的σ代数，即包含所有形如 $\{X\le a\}, a\in\mathbb{R}$的最小$\sigma$代数，记为$\sigma(X)$. 类似可定义由随机变量$X_1,\cdots,X_n$生成的 σ代数$\sigma(X_1,\cdots,X_n)$.

常用的两种类型随机变量：

(1) 离散型随机变量X的概率分布用分布列描述：

$$p_k = P(X=x_k), \quad k=1,2,\cdots$$

其分布函数 $F(x)=\sum_{x_k\le x} p_k$.

(2) 连续型随机变量 X的概率分布用概率密度$f(x)$描述，其分布函数

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

(3) 对于随机向量$X = (X_1, \cdots, X_d)$，它的(d维)分布函数（或联合分布函数）定义为

$$F(x_1, \cdots, x_d) = P(X_1 \le x_1, \cdots, X_d \le x_d)$$

这里d为正整数，$x_k \in \mathbb{R}, k = 1, 2, \ldots, d.$

定理 1.2.5 若$F(x_1, \cdots, x_d)$是联合分布函数，则

(1) $F(x_1, \cdots, x_d)$对每个变量都是单调的;

(2) $F(x_1, \cdots, x_d)$对每个变量都是右连续的;

(3) 对 $i = 1, 2, \cdots, d$

$$\lim_{x_i \to -\infty} F(x_1, \cdots, x_i, \cdots, x_d) = 0,$$

$$\lim_{x_1, x_2, \cdots, x_d \to \infty} F(x_1, x_2, \cdots, x_d) = 1.$$

如果$f(x_1, \cdots, x_d) = \frac{\partial^d F}{\partial x_1 \cdots \partial x_d}$对所有的 $(x_1, \cdots, x_d) \in R^d$ 存在,则称函数$f(x_1, \cdots, x_d)$ 为 $F(x_1, \cdots, x_d)$或

$X=(X_1,\cdots,X_d)$的联合密度函数，并且

$$F(x_1,\cdots,x_d)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d}f(t_1,\cdots,t_d)dt_d\cdots dt_1$$

设$F(x_1,\cdots,x_d)$为$X_1,\cdots,X_d$的联合分布函数

$1\le k_1<\cdots<k_n\le d$，则$X_1,\cdots,X_d$的边际分布 $F_{k_1,\cdots,k_n}(x_{k_1},\cdots,x_{k_n})$
定义为

$$F_{k_1,\cdots,k_n}(x_{k_1},\cdots,x_{k_n})$$
$$=F(\infty,\cdots,\infty,x_{k_1},\infty,\cdots,\infty,x_{k_2},\infty,\cdots,\infty,x_{k_n},\infty,\cdots,\infty)$$

一些常见分布:

1．退化分布:若随机变量X只取常数c，即

$$P\{X=c\}=1$$

则X并不随机，但我们把它看作随机变量的退化情况更为方便，因此称之为退化分布，又称单点分布.

2. Bernoulli分布:在一次试验中，设事件A出现的概率为$p, 0 < p < 1$，不出现的概率为$1-p$，若以X记事件A出现的次数，则X的可能取值仅为$0, 1$，其对应的概率为

$$P\{X = k\} = p^k(1-p)^{n-k}, \quad k = 0, 1$$

3. 二项分布:在n重Bernoulli试验中，设事件A在每次试验中出现的概率均为$p, 0 < p < 1$，以X记事件A出现的次数，X的可能取值为$0, 1, 2, \cdots, n$，其对应的概率为

$$P\{X = k\} = \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0, 1, \cdots, n$$

则称之为以n和p为参数的二项分布，简记为$X \sim B(n, p)$.

4. Poisson 分布：若随机变量X可取一切非负整数

值，且

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

其中$\lambda>0$，则称X服从Poisson分布，记为$X\sim P(\lambda)$.

5. 几何分布:在Bernoulli试验中，设事件A在每次试验中出现的概率均为$p, 0<p<1$，以X记事件A首次出现的试验次数， X的可能取值为$0,1,2,\cdots,$，其对应的概率为

$$P\{X=k\}=p(1-p)^{k-1},\quad k\geq 1$$

则称之为几何分布.

6. Pascal分布:在Bernoulli试验中，设事件A在每次试验中出现的概率均为$p, 0<p<1$，以X记事件A第r次出现的试验次数， X的可能取值为$r, r+1,\cdots$，其对应的概

率为

$$P\{X=k\}=\binom{k-1}{r-1}p^r(1-p)^{k-r}, \qquad k=r,r+1,\cdots$$

则称之为Pascal分布.

7. 负二项分布:对于任意实数$r>0$，称

$$P\{X=k\}=\binom{k+r-1}{k}p^r(1-p)^k, \quad k=0,1,2,\cdots, \quad r>0$$

为负二项分布.

负二项分布通常用于替换Poisson分布.同Poisson分布一样，它也在非负整数上取值，但因为它包含两个参数，相比Poisson分布其变化更灵活. Poisson分布的方差和均值相等，但负二项分布的方差大于均值，这说明当某类数据集观测到的方差大于均值时，负二项分布要比Poisson分

布更合适.

8. 离散均匀分布:如果分布列为

$$p_k = \frac{1}{n}, \quad k = 1, 2, \cdots, n$$

则称之为离散均匀分布.

9. 连续均匀分布(简称均匀分布):如果密度函数为

$$f(x) = \begin{cases} [b-a]^{-1}, & \text{若 } a \le x \le b \\ 0, & \text{其他} \end{cases}$$

其中$a < b$,则称之为区间$[a, b]$上的均匀分布.

10. 正态分布:如果密度函数为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\{-(x-\mu)^2/2\sigma^2\}, x \in \mathbb{R}$$

则称之为参数为μ和σ^2的正态分布,也称为Gauss分布,记为$X \sim N(\mu, \sigma^2)$.

11. d维正态分布: 设$\mu=(\mu_1,\cdots,\mu_d)$，$\mathbf{\Sigma}$是d阶正定对称矩阵，并且其行列式为$|\mathbf{\Sigma}|$. 如果其联合密度函数为

$$f(x_1,\cdots,x_d)=(2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{x}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\}$$

则称之为d维正态分布，记为$X\sim N(\mu,\mathbf{\Sigma})$.

12. Γ 分布： 如果密度函数为

$$f(x)=\begin{cases}\frac{\lambda^s}{\Gamma(s)}x^{s-1}e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

则称之为以$s>0$，$\lambda>0$为参数的Γ分布，其中Γ函数定义为

$$\Gamma(s)=\int_0^\infty e^{-x}x^{s-1}dx,\quad s>0$$

13. 指数分布: 如果在Γ分布中令$s=1$，即密度函数

为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

则称之为指数分布.

14. χ^2分布:如果在Γ分布中取$s=\frac{n}{2}$,n是正整数, $\lambda=\frac{1}{2}$, 即

$$f(x)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}},\quad x>0$$

则称之为自由度是n的χ^2分布.

§1.3 数字特征、矩母函数与特征函数

§1.3.1 Riemann-Stieltjes积分

设$g(x), F(x)$ 为有限区间$(a,b]$上的实值函数，$a = x_0 < x_1 < \cdots < x_n = b$为$(a,b]$的一个分割，令

$$\Delta F(x_i) = F(x_i) - F(x_{i-1}), \xi_i \in [x_{i-1}, x_i], 1 \leq i \leq n,$$

$\lambda = \max_{1\leq i\leq n}(x_i - x_{i-1})$, 如果当$\lambda \to 0$时，极限

$$\lim_{\lambda\to 0}\sum_{i=1}^{n} g(\xi_i)\Delta F(x_i)$$

存在，且与分割的选择以及$\xi_i \in [x_{i-1}, x_i]$的取法无关，则称该极限值为函数$g(x)$关于$F(x)$ 在$(a,b]$上的Riemann-

Stieltjes积分，记为

$$\int_a^b g(x)dF(x) = \lim_{\lambda\to 0}\sum_{i=1}^{n} g(\xi_i)\Delta F(x_i) \qquad (1.3.1)$$

当$F(x)=x$时，Riemann-Stieltjes积分即为Riemann积分．关于Riemann-Stieltjes积分存在的条件，这里不做更进一步的讨论，只给出一个简单的充分条件：若函数$g(x)$连续，$F(x)$单调，则Riemann-Stieltjes积分存在．本书中用到的$g(x)$为连续函数，$F(x)$为分布函数，因此积分的存在性不成问题．为了后面的需要，将积分推广到无限区间上：

$$\int_a^{+\infty} g(x)dF(x) \doteq \lim_{b\to+\infty}\int_a^b g(x)dF(x)$$

$$\int_{-\infty}^{b} g(x)dF(x) \doteq \lim_{a\to-\infty}\int_a^b g(x)dF(x)$$

$$\int_{-\infty}^{+\infty} g(x)dF(x) \doteq \lim_{a\to-\infty,b\to+\infty}\int_a^b g(x)dF(x)$$

与Riemannan积分不同的是

$$\int_{a-}^{a} dF(x) = \lim_{\delta\to 0+}\int_{a-\delta}^{a} dF(x) = F(a) - F(a-)$$

当$F(x)$在$x=a$处有跳跃时，上式的值等于$F(x)$在a点的跃度.当$F(x)$是一个阶梯函数时，Riemann-Stieltjes积分成为一个级数，即设$F(x)$在$x=x_i$处有跃度$p_i, i=1,2,\cdots$,则

$$\int_{-\infty}^{+\infty} g(x)dF(x) = \sum_{i=1}^{+\infty} g(x_i)p_i$$

Riemann-Stieltjes积分的一些基本性质：

(1) 线性性质:

$$\int_a^b [\alpha g_1(x)\pm\beta g_2(x)]dF(x) = \alpha\int_a^b g_1(x)dF(x)\pm\beta\int_a^b g_2(x)dF(x)$$

(2) **区间可加性：**

$$\int_a^b g(x)dF(x) = \int_a^c g(x)dF(x) + \int_c^b g(x)dF(x)$$

(3) $\int_a^b dF(x) = F(b) - F(a)$，**其中**$a, b$**均可为有限数或无穷大**.

(4) $\int_a^b g(x)d[\alpha F_1(x) + \beta F_2(x)] = \alpha \int_a^b g(x)dF_1(x) + \beta \int_a^b g(x)dF_2(x)$

(5) **若**$g(x) \geq 0$，$F(x)$**单调不减**,$b > a$,**则**$\int_a^b g(x)dF(x) \geq 0$.

§1.3.2 数字特征

定义 1.3.1 (1) 取值为$\{s_k\}$的离散型随机变量X的数学期望 (简称为期望) $E[X]$定义为

$$E[X] = \sum_k s_k p_k = \sum_k s_k P(X = s_k)$$

如果$\sum |s_k| p_k < \infty$.

(2) 连续型随机变量X的数学期望 $E[X]$定义为

$$E[X] = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx$$

如果$\int_{-\infty}^{\infty} |x| dF(x) < \infty$, 这里$F(x)$是$X$的分布函数,$f(x)$是其密度函数.

利用Riemann-Stieltjes积分，我们可以对离散型随机

变量和连续型随机变量的期望给出一个统一的表达式:

$$E[X] = \int_{-\infty}^{+\infty} x dF(x)$$

(3) 设X为任一随机变量, 对正整数k, 称$m_k = E[X^k]$为X的k阶原点矩. 数学期望是一阶原点矩.

(4) 设X为任一随机变量, 对正整数k, 称$c_k = E[X - E[X]]^k$为X的k阶中心矩.方差是二阶原点矩.

(5) 设X, Y为两个随机变量, 对正整数k, l, 称$E[X - E[X]]^k[Y - E[Y]]^l$为$X$的$k + l$阶混合中心矩.协方差是二阶混合中心矩.

§1.3.3 关于概率测度的积分

定义 1.3.2 设$(\Omega, \mathcal{F})$为一可测空间, $\mathbb{R}$为实数域, $\bar{\mathbb{R}} =$

$\mathbb{R}\cup\{-\infty,\infty\}$，分别用 $\mathcal{B}(\mathbb{R})$及$\mathcal{B}(\bar{\mathbb{R}})$表示$\mathbb{R}$及$\bar{\mathbb{R}}$上的Borel$\sigma$代数. 令$f$为$\Omega$到$\mathbb{R}$中的一个映射，如果$f^{-1}(\mathcal{B}(\bar{\mathbb{R}}))\subset\mathcal{F}$，则称$f$为**Borel**可测函数，简称可测函数.若进一步f只取实值，则称f为实值可测函数.

如果存在实数$a_k, 1\le k\le n$和Ω的分割$A_k\in\mathcal{F}, 1\le k\le n$ (即$\bigcup_{k=1}^{n}A_k=\Omega$，且 $A_i\cap A_j=\emptyset$, $i\neq j$, $1\le i,j\le n$)，使得 $f(\omega)=\sum_{k=1}^{n}a_kI_{A_k}(\omega)$，则称$f$为简单可测函数.这里$I_A(\cdot)$表示集合 A的示性函数. 若f还可以表示为$f=\sum_{j=1}^{m}b_jI_{B_j}$，则当$B_j\cap A_k\neq\emptyset$时，$b_j=a_k$.于是，通过将分割中的集合合并可以得到$f$的最简单表达式，即表达式中的系数 a_k互不相同.

以下给定一个完备的概率空间$(\Omega,\mathcal{F},P)$. 用$\mathcal{S}^+$表示Ω上非负简单可测函数全体，$\mathcal{L}^+$表示Ω上非负可测函

数全体，$\mathcal{L}$表示Ω上可测函数全体.

首先我们定义非负简单可测函数关于概率测度的积分.

定义 1.3.3 令$\mathbf{S}$表示所有非负简单可测函数$h:\Omega\to R_+$. 如果$h\in\mathbf{S}$, 则定义其关于P的积分为

$$\int hdP=\sum_{k=1}^{n}a_kP(A_k)=E[h]$$

易知$\int_\Omega fdP$不依赖于f的具体表达.称$\int_\Omega fdP$为f关于概率测度P的积分.

定理 1.3.4 (简单函数逼近定理)设$(\Omega,\mathcal{F})$为一可测空间，$f\in\mathcal{L}$.

(1)存在一简单可测函数序列$(f_n,n\geq 1)$，使得对一切$n\geq 1$，有$|f_n|\leq|f|$，且$\lim_{n\to\infty}f_n=f$.

(2)若f非负，则存在非负简单可测函数的增序列(f_n)，

GoBack
FullScreen
Close
Quit

使得$\lim_{n\to\infty} f_n = f$.

接下来定义非负可测函数的积分.

定义 1.3.5 设$f \in \mathcal{L}^+$.任取$f_n \in \mathcal{S}^+$使得$f_n \uparrow f$，令

$$\int_\Omega f dP = \lim_{n\to\infty} \int_\Omega f_n dP$$

上述右端极限存在，且不依赖于序列(f_n)的选取，称$\int_\Omega f dP$为f关于概率测度P的积分.

最后定义可测函数的积分.

定义 1.3.6 设$f \in \mathcal{L}$.令

$$f^+ = \max\{f, 0\} \quad \text{及} \quad f^- = -\min\{f, 0\}$$

注意到$f^+ \geq 0, f^- \geq 0, f = f^+ - f^-, |f| = f^+ + f^-$.如果$\int_\Omega f^+ dP < \infty$或$\int_\Omega f^- dP < \infty$，则称$f$关于$P$的积分存

在.令
$$\int_\Omega f dP = \int_\Omega f^+ dP - \int_\Omega f^- dP$$
如果$\int_\Omega f^+ dP < \infty$且$\int_\Omega f^- dP < \infty$，则称$f$关于概率测度$P$可积.

于是，当随机变量X可积时，它的期望就可以定义为
$$E[X] = \int_\Omega X dP$$

令$\mathcal{L}^p(\Omega), p \geq 1$表示所有使得$E[|X|^p] < \infty$的随机变量(等价类)全体，简记为$\mathcal{L}^p$.

关于Ω 中样本点的某种性质Π，如果使得Π不成立的点的集合的概率是零，则称Π 几乎必然(almost surely) 或以概率1(with probability one)成立，记为$a.s.$ 或$w.p.1$.

定理 1.3.7 设f, g积分存在.

(1) $\forall a \in \mathbb{R}$，af的积分存在，且$\int_\Omega afdP = a\int_\Omega fdP$.

(2) 若$f+g$处处有定义，且$\int_\Omega fdP+\int_\Omega gdP$处处有意义(即不出现$\infty-\infty$)，则$f+g$的积分存在，且有$\int_\Omega (f+g)dP = \int_\Omega fdP + \int_\Omega gdP$.

(3) 若取非负整数值的随机变量X的期望存在，则$E[X] = \sum_{k=1}^{\infty} P\{X \geq k\}$.

(4) 若N为一零测集，则$\int_\Omega fI_N dP = 0$.

(5) 若$f \leq g, a.s.$，则$\int_\Omega fdP \leq \int_\Omega gdP$.

(6) $|\int_\Omega fdP| \leq \int_\Omega |f|dP$.

(7) 设$X \in \mathcal{L}^+$，则$E[X]=0$当且仅当$X=0$ a.s.

§1.3.4 矩母函数

定义 1.3.8 若随机变量X的分布函数为$F_X(x)$, 则称

$$\phi_X(t) = E[e^{tX}] = \int_\Omega e^{tX(\omega)} P(d\omega) = \int_{-\infty}^{\infty} e^{tx} dF_X(x) \tag{1.3.2}$$

为X的矩母函数.

注: X的各阶矩与矩母函数的关系（假设对$\phi(t)$求导时，求导运算与求期望运算可以交换次序）

$$\phi'(t) = E(Xe^{tX})$$

$$\phi''(t) = E(X^2e^{tX})$$

$$\vdots$$

$$\phi^{(n)}(t) = E(X^ne^{tX})$$

令$t=0$，得到 $\phi^{(n)}(0)=E[X^n]$，$n\geq 1$. 当矩母函数存在时，它唯一地决定分布，因此我们能够用矩母函数刻画随机变量的概率分布.但有时随机变量的矩母函数不一定存在，在这种情况下，更方便的是特征函数.

§1.3.5 特征函数

定义 1.3.9 若随机变量X的分布函数为$F_X(x)$，则称

$$\psi_X(t)=E[e^{itX}]=\int_\Omega e^{itX(\omega)}P(d\omega)=\int_{-\infty}^{\infty}e^{itx}dF_X(x)$$

为X的特征函数. 如果F_X有密度$f(x)$，则$\psi_X(t)$就是$f(x)$的Fourier变换

$$\psi_X(t)=\int_{-\infty}^{\infty}e^{itx}f(x)dx$$

特征函数是一个实变量的复值函数，因为$|e^{itx}| = 1$，所以它对一切实数t都有定义.特征函数有如下常用性质：

(1) 有界性：$|\psi(t)| \le 1 = \psi(0)$;

(2) 共轭对称性：$\psi(-t) = \overline{\psi(t)}$;

(3) 一致连续性：

$$|\psi(t+h) - \psi(t)| \le \int_{-\infty}^{\infty} |e^{ihx} - 1| dF(x);$$

(4)线性变换的作用：设$Y = aX + b$,则Y的特征函数是$\psi_Y(t) = e^{ibt}\psi_X(at)$;

(5)两个相互独立的随机变量之和的特征函数等于它们的特征函数之积.

(6) 非负定性: 对于任意的正整数n, 任意实数$t_1, \cdots, t_n$及复数$\lambda_1, \cdots, \lambda_n$，有 $\sum_{k=1}^{n} \sum_{j=1}^{n} \psi(t_k - t_j)\lambda_k\overline{\lambda_j} \ge 0$.

(7) 设随机变量X有n阶矩存在，则它的特征函数可微

分n次，且当$k \le n$时，有

$$\psi^{(k)}(0) = i^k E[X^k]$$

特别地，特征函数可作如下带皮阿诺型余项的Taylor展开：

$$\psi(t) = 1 + itE[X] + \frac{(it)^2}{2!}E[X^2] + \cdots + \frac{(it)^n}{n!}E[X^n] + o(t)$$

例 1.3.10 求标准正态分布$N(0,1)$的特征函数.

解: 由定义 $\psi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx}e^{-x^2/2}dx$

GoBack
FullScreen
Close
Quit

从而

$$\begin{aligned}\psi'(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ixe^{itx}e^{-x^2/2}dx\\ &= \frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx}d(-e^{-x^2/2})\\ &= -\frac{i}{\sqrt{2\pi}}(e^{itx-x^2/2}|_{-\infty}^{\infty} + t\int_{-\infty}^{\infty} e^{itx}e^{-x^2/2}dx)\\ &= -t\psi(t)\end{aligned}$$

解初值问题$\psi'(t)+t\psi(t)=0,\ \psi(0)=1$ 得 $\psi(t)=e^{-\frac{1}{2}t^2}$.

例 1.3.11 求正态分布 $N(\mu,\sigma^2)$ 的特征函数.

解:设 $Y\sim N(\mu,\sigma^2)$, $X\sim N(0,1)$, 则有 $Y=\sigma X+\mu$.

利用性质(4)有 $\psi_Y(t)=e^{it\mu}\psi_X(\sigma t)=e^{i\mu t-\frac{1}{2}\sigma^2t^2}$.

由于特征函数只与分布函数有关，所以称为分布的特征函数. 另一方面，有下述定理.

定理 1.3.12 (唯一性定理) 分布函数由其特征函数唯一决定.

从而说明特征函数与分布函数是相互唯一确定的.

若随机向量$(X_1,\cdots,X_n)$的分布函数为$F(x_1,\cdots,x_n)$，与随机变量相仿，类似地定义它的特征函数

$$\psi(t_1,\cdots,t_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{i(t_1x_1+\cdots+t_nx_n)}dF(x_1,\cdots,x_n)$$

可以类似于一元的场合，建立起n元特征函数的理论.

例如，我们考虑在上1.2节中定义的 d维正态分布，简称为多元正态分布，这个定义在推导很多性质时并不方便，同时还不能考虑Σ不是正定的情形，为此，我们采用下面的定义. 如果随机变量X 的特征函数为

$$\psi_X(t)=\exp\{i\mathbf{t}'\mu-\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}\}$$

则称**X**服从多元正态分布$N(\mu,\mathbf{\Sigma})$.其中$\mathbf{X}=(X_1,\cdots,X_c)'$，$(\cdot)'$表示$(\cdot)$的转置，$\mathbf{t}=(t_1,\cdots,t_d)',\mu=(\mu_1,\cdots,\mu_d)'$，$\mathbf{\Sigma}$是非负定$d\times d$矩阵. 当$\mu=\mathbf{0}$， $\mathbf{\Sigma}=\mathbf{I}_d$时，称为标准多元正态分布，记为$X\sim N(\mathbf{0},\mathbf{I}_d)$.

利用特征函数方法不难证明下面几个命题：

命题 1.3.13 若$\mathbf{X}\sim N(\mathbf{0},\mathbf{I}_d)$，则**X**的任一线性函数$\mathbf{Y}=\mathbf{A}_{n\times d}\mathbf{X}+\mu$ 服从n维正态分布$N(\mu,\mathbf{A}\mathbf{A}')$.

命题 1.3.14 若$\mathbf{Y}\sim N(\mu,\mathbf{\Sigma})$，则$\mathbf{A}\mathbf{Y}+\mathbf{b}\sim N(\mathbf{A}\mu+\mathbf{b},\mathbf{A}\mathbf{\Sigma}\mathbf{A}')$.

§1.4 收敛性

定义 1.4.1 (1) 设$\{X_n, n \geq 1\}$是随机变量序列，若存在随机变量X使得

$$P\{\omega \in \Omega : X(\omega) = \lim_{n\to\infty} X_n(\omega)\} = 1$$

则称随机变量序列$\{X_n, n \geq 1\}$几乎必然收敛(或以概率1收敛)于X，记为 $X_n \to X, a.s.$或$X_n \xrightarrow{a.s.} X$.

(2) 设$\{X_n, n \geq 1\}$是随机变量序列，若存在随机变量X使得 $\forall \varepsilon > 0$,有

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0$$

则称随机变量序列$\{X_n, n \geq 1\}$依概率收敛于X，记为$X_n \xrightarrow{P} X$.

(3) 设随机变量序列$\{X_n\} \subset \mathcal{L}^p, p \geq 1, X \in \mathcal{L}^p$，若有

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0$$

则称随机变量序列$\{X_n, n \geq 1\}$ p次平均收敛于X, 或称$\{X_n\}$在$\mathcal{L}^p$中强收敛于X. 当$p = 2$时，称为均方收敛.

(4)设$\{F_n(x)\}$是分布函数列，如果存在一个单调不减函数$F(x)$,使得在$F(x)$的所有连续点x上均有

$$\lim_{n\to\infty} F_n(x) = F(x)$$

则称$\{F_n(x)\}$弱收敛于$F(x)$，记为$F_n(x) \xrightarrow{W} F(x)$.

设随机变量X_n, X的分布函数分别为$F_n(x)$及$F(x)$，若$F_n(x) \xrightarrow{W} F(x)$, 则称$\{X_n\}$依分布收敛于$X$, 记为$X_n \xrightarrow{L} X$.

定理 1.4.2 (1) 随机变量序列$X_n \xrightarrow{a.s.} X$的充分必要

条件是 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\{\sup_{m\geq n} |X_m - X| \geq \varepsilon\} = 0$$

(2) 随机变量序列$X_n \xrightarrow{P} X$的充分必要条件是$\{X_n\}$的任意子序列都包含几乎必然收敛于X的子序列.

随机变量序列的这4种收敛性之间的关系可以总结为下面的关系图:

几乎必然收敛$\Longrightarrow$依概率收敛$\Longrightarrow$依分布收敛;

p次平均收敛$\Longrightarrow$ 依概率收敛$\Longrightarrow$依分布收敛.

注: 几乎必然收敛与p阶矩收敛之间没有蕴含关系.

例 1.4.3 取$\Omega = (0,1]$, $\mathcal{F}$为$(0,1]$中全体Borel子集所构成的σ-代数, P为Lebesgue测度,我们可以构造出2个随机

变量序列,其中之一是r阶矩收敛的,但是不几乎必然收敛;另外一个则几乎必然收敛, 但不是r阶矩收敛的.

令

$$Y_{11}=1;\quad Y_{21}=\begin{cases}1, & \omega\in(0,\frac{1}{2}]\\ 0, & \omega\in(\frac{1}{2},1]\end{cases}$$

$$Y_{22}=\begin{cases}0, & \omega\in(0,\frac{1}{2}]\\ 1, & \omega\in(\frac{1}{2},1]\end{cases}$$

一般地，将$(0,1]$分成k 个等长区间，并且令

$$Y_{ki}=\begin{cases}1, & \omega\in(\frac{i-1}{k},\frac{i}{k}]\\ 0, & \omega\notin(\frac{i-1}{k},\frac{i}{k}]\end{cases}\quad i=1,\cdots,k,\quad k=1,2,\cdots.$$

定义随机变量序列

$$X_1=Y_{11},\ X_2=Y_{21},\ X_3=Y_{22},\ X_4=Y_{31},\ X_5=Y_{32},\cdots.$$

对任意$\varepsilon > 0$, 由于

$$E|Y_{ki} - 0|^r = \frac{1}{k} \to 0, (k \to \infty)$$

可见$\{X_n\}$为r阶矩收敛，但是对任意固定的$\omega \in \Omega$, 任一自然数k，恰有一个i，使得$Y_{ki}(\omega) = 1$, 而对其余的j有$Y_{kj}(\omega) = 0$. 由此知 $\{X_n(\omega)\}$中有无穷多个1及无穷多个0，于是$\{X_n(\omega)\}$对每个$\omega \in \Omega$ 都不收敛.

如果取

$$Z_n = \begin{cases} n^{\frac{1}{r}}, & 若\omega \in (0, \frac{1}{n}] \\ 0\ \ , & 若\omega \notin (0, \frac{1}{n}] \end{cases}$$

$$Z = 0, \quad \forall \omega \in \Omega$$

易见，$Z_n(\omega) \to Z(\omega) \ \ \forall \omega \in \Omega$, 所以

$$Z_n \to Z, \quad a.s.$$

但是$E(|Z_n - Z|^r) = n \cdot \frac{1}{n} = 1 \not\to 0$.

下面我们给出积分号下取极限的三大基本定理.

定理 1.4.4 (单调收敛定理) 设$f_n \in \mathcal{L}$的积分存在，$n \geq 1$，则:

(1)设$f_n \uparrow f, a.s.$,若$\int_\Omega f_1 dP > -\infty$，则$f$的积分存在，且$\int_\Omega f_n dP \uparrow \int_\Omega f dP$;

(2)设$f_n \downarrow f, a.s.$,若$\int_\Omega f_1 dP < \infty$，则$f$的积分存在，且$\int_\Omega f_n dP \downarrow \int_\Omega f dP$.

定理 1.4.5 (Fatou引理) 设随机变量序列$\{X_n\}$的期望存在,$n \geq 1$,则

$$E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n] \leq \limsup_{n\to\infty} E[X_n] \leq E[\limsup_{n\to\infty} X_n]$$

定理 1.4.6 (Lebesgue控制收敛定理) 设$f_n, f \in$

$\mathcal{L}$，$f_n \xrightarrow{a.s.} f$或$f_n \xrightarrow{P} f$.若存在一非负可积函数g，使得$\forall n \geq 1$有$|f_n| \leq g, a.s.$，则f可积，且有$\lim_{n\to\infty} \int_\Omega f_n dP = \int_\Omega f dP$.

§1.5 独立性与条件期望

§1.5.1 独立性

定义 1.5.1 (1) 设A, B为两个事件，若$P(A \bigcap B) = P(A)P(B)$，则称$A$与$B$独立.更一般地，设$A_1, A_2, \cdots, A_n$为$n$个事件，如果对任何$m \leq n$及$1 \leq k_1 < k_2 < \cdots < k_m \leq n$，有

$$P\left(\bigcap_{j=1}^{m} A_{k_j}\right) = \prod_{j=1}^{m} P(A_{k_j})$$

则称$A_1, A_2, \cdots, A_n$相互独立. $A_1, A_2, \cdots, A_n$两两独立不一定相互独立.

(2) 设$\{A_i, i \in I\}$是一族事件，若 对I的任意有限子集$\{i_1, \cdots, i_k\} \neq \emptyset$有

$$P(\bigcap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j}) \qquad (1.5.1)$$

则称$\{A_i, i \in I\}$是相互独立的.

(3) 设$\{\mathcal{A}_i, i \in I\}$是一族事件类，如果对$I$的任意有限子集$\{i_1, \cdots, i_k\} \neq \emptyset$，任意$A_{i_j} \in \mathcal{A}_{i_j}$ 有(1.5.1)式成立，则称$\{\mathcal{A}_i, i \in I\}$是独立事件类.

(4) 设$\{X_i, i \in I\}$是Ω上一族随机变量，如果σ代数族$\{\sigma(X_i), i \in I\}$ 是独立事件类，则称$\{X_i, i \in I\}$相互独立.

容易证明随机变量$X_1,\cdots,X_n$独立的充分必要条件是它们的联合分布函数可以分解为

$$F(x_1,\cdots,x_n)=F_{X_1}(x_1)\cdots F_{X_n}(x_n)$$

定理 1.5.2 (1)设随机变量$X_1,\cdots,X_n\in\mathcal{L}^1$是独立的，则 $E[\Pi_{k=1}^n X_k]=\Pi_{k=1}^n E[X_k]$.

(2)设随机变量$X_1,\cdots,X_n\in\mathcal{L}^2$是独立的，则 $Var[\sum_{k=1}^n X_k]=\sum_{k=1}^n Var[X_k]$.

定理 1.5.3 (Borel-Cantelli第一引理)

设$\{A_n,n\geq 1\}$是一列事件，若$\sum_{n=1}^\infty P(A_n)<\infty$，则$P(A_n,i.o.)=0$.

定理 1.5.4 (Borel-Cantelli 第二引理)

设$\{A_n,n\geq 1\}$是一独立的事件列，若$\sum_{n=1}^\infty P(A_n)=$

∞，则$P(A_n, i.o.) = 1$.

定义 1.5.5 设$\{X_n, n \geq 1\}$是随机变量序列，$\mathcal{D}_k = \sigma(X_k, X_{k+1}, \cdots)$是由$X_k, X_{k+1}, \cdots$生成的$\sigma$代数，则$\{\mathcal{D}_k\}$是非增的列，它们的交$\mathcal{D} = \cap_{n\geq 1}\mathcal{D}_n$称为序列$\{X_n, n \geq 1\}$ 的尾σ代数，$\mathcal{D}$中的元素称为$\{X_n, n \geq 1\}$的尾事件.

定理 1.5.6 (Kolmogorov 0-1律)
独立随机变量序列的尾事件的概率或为0或为1.

§1.5.2 独立随机变量和的分布

设随机变量X_1, X_2相互独立，F_1, F_2分别为它们的分布函数.令$X = X_1 + X_2$,其分布函数记为 $F_X(x)$. 则由独立

性,有

$$F_X(x) = P\{X_1 + X_2 \leq x\}$$
$$= \int_{-\infty}^{\infty} P\{X_1 + X_2 \leq x | X_1 = t\} dF_1(t)$$
$$= \int_{-\infty}^{\infty} F_2(x - t) dF_1(t) \tag{1.5.2}$$

上述第三个等号右端称作分布函数F_1, F_2的卷积，记为$F_1 * F_2(x)$.一般地对有界函数$g(x)$和一个单调函数$F(x)$，都可以定义F与g的卷积:

$$F * g(x) \triangleq \int_{-\infty}^{\infty} g(x - t) dF(t)$$

这里需要注意的是$F * g$的顺序，$g * F$可能没有意义.但是当F和g都是分布函数时，卷积可以交换顺序. 只需注意到

卷积中的随机变量X_1和X_2的地位是对等的即可得到

$$F_1 * F_2(x) = F_2 * F_1(x)$$

不但如此，容易看出，对于分布函数，卷积还满足结合律和分配律. 即设F, G, H为分布函数，则有

$$(F * G) * H(x) = F * (G * H)(x)$$

$$F * (G + H)(x) = F * G(x) + F * H(x)$$

于是，更进一步还有，设$X_k, k = 1, 2, \cdots, n$是独立同分布F的随机变量，令 $S_0 = 0, S_n = \sum_{k=1}^{n} X_k, n = 1, 2, \cdots,$ S_n的分布记作F_n，则有

$$F_0(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

$$F_n(x) = F * F_{n-1}(x), n = 1, 2, \cdots$$

称F_n为F的n重卷积.

§1.5.3 条件期望

设B是一个事件，且$P(B) > 0$.则事件B发生的条件下事件A发生的条件概率为

$$P(A|B) = P(A \cap B)/P(B)$$

定理 1.5.7 **(全概率公式)** 设$\{B_n\}$是Ω的一个分割，且使得$P(B_n) > 0$，$\forall n$.如果$A \in \mathcal{F}$，则

$$P(A) = \sum_n P(B_n)P(A|B_n)$$

定理 1.5.8 **(Bayes公式)** 设$\{B_n\}$是Ω的一个分割，且使得$P(B_n) > 0$，$\forall n$，如果$P(A) > 0$，则

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_n P(B_n)P(A|B_n)}, \quad n \geq 1$$

如果X与Y是离散型随机变量，对一切使得$P\{Y=y\}>0$的y，给定$Y=y$时，X的条件概率定义为：

$$P\{X=x|Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}$$

X的条件分布定义为：

$$F(x|y)=P\{X\leq x|Y=y\}$$

X的条件期望定义为：

$$E[X|Y=y]=\int xdF(x|y)=\sum_{x}xP\{X=x|Y=y\}$$

如果X与Y有联合概率密度函数$f(x,y)$，则对一切使得

$f_Y(y)\geq 0$的y，给定$Y=y$时，X的条件概率密度函数定义为:

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}$$

X的条件分布定义为：

$$F(x|y) = P\{X \le x|Y = y\} = \int_{-\infty}^{x} f(x|y)dx$$

X的条件期望定义为：

$$E[X|Y = y] = \int xdF(x|y) = \int xf(x|y)dx$$

我们以$E[X|Y]$表示随机变量Y的函数，它在$Y = y$时，取值为$E[X|Y = y]$.条件期望的一个重要性质是对一切随机变量X和Y，当期望存在时，有

$$E[X] = E[E[X|Y]] = \int E[X|Y = y]dF_Y(y) \quad (1.5.3)$$

当Y为一个离散随机变量时，(1.5.3)式为

$$E[X] = \sum_{y} E[X|Y = y]P\{Y = y\}$$

当Y为一个连续随机变量时，(1.5.3)式为

$$E[X]=\int_{-\infty}^{+\infty}E[X|Y=y]f(y)dy$$

例 1.5.9 (随机个随机变量之和)设$X_1,X_2,\cdots$是一列与X独立同分布的随机变量；设N为一非负整值随机变量，且与序列$X_1,X_2,\cdots$独立.求$Y=\sum_{i=1}^{N}X_i$的均值和方差.

解　首先在对N取条件的情况下来计算$Y=\sum_{i=1}^{N}X_i$的矩母函数，即

$$E[\exp\{t\sum_{i=1}^{N}X_i\}|N=n]=E[\exp\{t\sum_{i=1}^{N}X_i\}]=(\phi_X(t))^n$$

其中$\phi_X(t)$是随机变量X的矩母函数，因此

$$E[\exp\{t\sum_{i=1}^{N}X_i\}|N]=(\phi_X(t))^N$$

从而

$$\phi_Y(t) = E[\exp\{t\sum_{i=1}^{N} X_i\}] = E[(\phi_X(t))^N]$$

现在对$\phi_Y(t)$求导得

$$\phi_Y^{'}(t) = E[N(\phi_X(t))^{N-1}\phi_X^{'}(t)]$$

再求一次导数得

$$\phi_Y^{''}(t) = E[N(N-1)(\phi_X(t))^{N-2}(\phi_X^{'}(t))^2+N(\phi_X(t))^{N-1}\phi_X^{''}(t)]$$

计算在$t=0$点的值，得

$$E[Y] = E[NE[X]] = E[N]E[X] \qquad (1.5.4)$$

及

$$E[Y^2] = E\left[N(N-1)(E[X])^2 + NE[X^2]\right] = E[N]Var[X]+E[N^2](E[x])^2$$

因此有

$$Var[Y]=E[Y^2]-(E[Y])^2=E[N]Var[X]+(E[x])^2Var[N] \tag{1.5.5}$$

最后我们将条件期望推广到一般随机变量及σ代数情形.

设X是随机变量，B是事件且$P(B)>0$，则给定事件B，随机变量X的条件期望 定义为

$$E[X|B]=\int XdP_B=[P(B)]^{-1}\int_B XdP=[P(B)]^{-1}E[XI_B]$$

定义 1.5.10 设X是随机变量且$E[|X|]<\infty$.若对每个子σ代数$\mathcal{G}\subset\mathcal{F}$，存在唯一的(几乎必然相等的意义下)随机变量$X^*$，有$E[|X^*|]<\infty$，使得$X^*$ 是$\mathcal{G}$可测随机变量(即

对任何的 $a \in \mathbb{R}$ ，有$\{X^* \le a\} \in \mathcal{G}$)，且

$$E[X^* I_B] = E[X I_B], \qquad \forall B \in \mathcal{G}$$

则称随机变量X^*为X在给定$\mathcal{G}$下的条件期望，记为$X^* = E[X|\mathcal{G}]$.即

$$\int_B E[X|\mathcal{G}]dP = \int_B X dP, \quad \forall B \in \mathcal{G}$$

定理 1.5.11 条件期望有如下基本性质:

(1) $E[E[X|\mathcal{G}]] = E[X]$.

(2) 若X是$\mathcal{G}$可测，则$E[X|\mathcal{G}] = X, a.s.$.

(3) 设$\mathcal{G} = \{\emptyset, \Omega\}$，则$E[X|\mathcal{G}] = E[X], a.s.$.

(4) $E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}], a.s.$.

(5) 若$X \le Y, a.s.$，则$E[X|\mathcal{G}] \le E[Y|\mathcal{G}], a.s.$.

(6) 若a, b为实数，$X, Y, aX+bY$的期望存在，则

$$E[aX+bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}], a.s.$$

如果右边和式有意义.

(7) $|E[X|\mathcal{G}]| \le E[|X||\mathcal{G}], a.s.$.

(8) 设$0 \le X_n \uparrow X, a.s.$，则$E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}], a.s.$.

(9) 设X及XY的期望存在，且Y为$\mathcal{G}$可测，则

$$E[XY|\mathcal{G}] = YE[X|\mathcal{G}], a.s.$$

(10) 若X与$\mathcal{G}$相互独立(即$\sigma(X)$与$\mathcal{G}$相互独立)，则有

$$E[X|\mathcal{G}] = E[X], a.s.$$

(11) 若$\mathcal{G}_1$,$\mathcal{G}_2$ 是两个子σ代数，使得$\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$，则

$$E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1], a.s.$$

(12)　若X,Y是两个独立的随机变量，函数$g(x,y)$使得$E[|g(X,Y)|]<+\infty$，则有

$$E[g(X,Y)|Y]=E[g(X,y)]|_{y=Y},\quad a.s.$$

这里$E[g(X,y)]|_{y=Y}$的意义是，先将y视为常数，求得数学期望$E[g(X,y)]$ 后再将随机变量Y代入到y的位置.

定义 1.5.12　设$f(x_1,\cdots,x_d)$是随机变量$X_1,\cdots,X_d$的联合密度函数，则$X_1,\cdots,X_k$在给定 $X_{k+1},\cdots,X_d$时的条件密度 $f_{1,\cdots,k}(u_1,\cdots,u_k|x_{k+1},\cdots,x_d)$ 定义为

$$\begin{aligned}&f_{1,\cdots,k}(u_1,\cdots,u_k|x_{k+1},\cdots,x_d)\\=&\frac{f(u_1,\cdots,u_k,x_{k+1},\cdots,x_d)}{\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}f(y_1,\cdots,y_k,x_{k+1},\cdots,x_d)dy_1\cdots dy_k}\end{aligned}$$