第 1 章　预备知识与随机过程基本概念

1.1　概率论补充知识

1.1.1　概率空间

在概率论中，称随机试验（以下简称试验）的每一可能结果为**样本点**，记为 ω，称样本点的全体为**样本空间**，记为 Ω.

一般而言，事件 A 是样本空间 Ω 的子集，即 $A\subset\Omega$. 说事件 A 在一次试验中发生，当且仅当 A 中的一个样本点在该次试验中出现. 所谓“概率”就是对 A 发生的可能性大小的度量.

由于并不是在所有的 Ω 的子集上都能方便地定义概率，因而仅限于在满足一定条件下的集类上，来研究概率及其性质. 为此引进如下的“事件域”概念.

定义 1.1.1 设 $\mathscr{F}$ 是以样本空间 Ω 的一些子集为元素的集合（谓之**集类**），若它满足条件

(1) $\Omega\in\mathscr{F}$；

(2) 若 $A,B\in\mathscr{F}$，则 $A-B\in\mathscr{F}$；

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

则称 $\mathscr{F}$ 为**事件域**，称 $\mathscr{F}$ 中的元素为**事件**，并且称 Ω 为**必然事件**，$\varnothing$ 为**不可能事件**.

不难验证，事件域 $\mathscr{F}$ 对可列次交、并、差等运算封闭，即，$\mathscr{F}$ 中的任何元素经可列次运算后仍属于 $\mathscr{F}$. 事件域 $\mathscr{F}$ 又称作 **σ-代数**或 **σ-域**.

定义 1.1.2 设 Ω 为样本空间，$\mathscr{F}$ 为 Ω 上的事件域，$P(\cdot)$是一个定义在 $\mathscr{F}$ 上的集函数，若它满足下列条件

(1) **非负性**　$\forall A\in\mathscr{F}$，$P(A)\geq 0$；

(2) **规范性**　$P(\Omega)=1$；

(3) **可列可加性**　当 $A_i\in\mathscr{F}$ ($i=1,2,\cdots$)，且 $A_iA_j=\varnothing$ ($\forall i\neq j$) 有

$$P\left(\sum_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i),$$

则称 $P(\cdot)$为 $\mathscr{F}$ 上的**概率测度**，称$(\Omega,\mathscr{F},P)$为**概率空间**，对 $A\in\mathscr{F}$，称 $P(A)$为**事件 A 的概率**.

在本科概率论中，通常都认为概率空间$(\Omega,\mathscr{F},P)$是预先给定的，在此基础上，来展开对各种概率问题的讨论. 实际上，构造概率空间要视具体问题而定，并无统一模式.

例 1.1.3 设某试验的样本空间 Ω 为有限集：$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$. 取事件域 $\mathscr{F}$ 为 Ω 的幂集 2^{Ω}（即，Ω 的所有子集的全体. 此处的集类 2^{Ω} 中含 2^n 个元.）. 在 $\mathscr{F}$ 上按如下方式给出概率测度 $P(\cdot)$：先取定一组正实数 p_1，p_2，$\cdots$，p_n，且 $\sum\limits_{k=1}^{n}p_k=1$，再令

$$P(\{\omega_k\})=p_k,\quad k=1,2,\cdots,n,$$

$$P(A)=\sum_{\omega_k\in A}P(\{\omega_k\})=\sum_{\omega_k\in A}p_k,\quad \forall A\in\mathscr{F},$$

容易验证如上定义的 $P(\cdot)$是 $\mathscr{F}$ 上的概率测度，因而$(\Omega,\mathscr{F},P)$为概率空间. □

注 (1) 在上例中，对每一样本点 ω_k 赋予正实数 p_k $(P(\{\omega_k\})=p_k)$是问题的要点，但究竟应

取怎样的 p_1，p_2，…，p_n，这要由试验的具体情况来定．例如，若认为每一样本点 ω_k 的出现机会均等，则可取

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

此时

$$P(\{\omega_k\})=\frac{1}{n},\quad k=1,2,\cdots,n,$$

$$P(A)=\frac{m}{n},\quad \forall A\in\mathscr{F},$$

其中 m 是事件 A 所含的样本点的个数．如此构造的(Ω，$\mathscr{F}$，P)即为古典概型问题的概率空间．

(2) 上例构造概率空间的方法可推广到 $\Omega=\{\omega_1，\omega_2，\cdots\}$为可列集的这种场合．

(3) 在以后的讨论中，如无特别需要，均认为概率空间(Ω，$\mathscr{F}$，P)是预先给定的．

延伸阅读

如果某试验的样本空间 Ω 为不可列集，那么通常要用测度论的方法才能构造出相应的概率空间(Ω，$\mathscr{F}$，P)．请看下面的例子．

例 1.1.4 设某试验的样本空间为 $\Omega=(0,1)$，取事件域 $\mathscr{F}$ 为$(0,1)\cap\mathscr{B}^1$，其中 $\mathscr{B}^1$ 是一维 Borel 域（参见本章附录中关于 Borel 域的描述），称 $\mathscr{F}$ 为 $\mathscr{B}^1$ **在集**$(0,1)$**上的限制**．此外，$\mathscr{F}$ 也等于由半环 $\mathscr{C}=\{(a,b]:\ 0\le a<b<1\}$ 生成的 σ-代数，即 $\mathscr{F}=\sigma(\mathscr{C})$．在样本空间 Ω 上引入了事件域 $\mathscr{F}$ 之后，再来构造 $\mathscr{F}$ 上的概率测度．为此，取定一非负可积函数 $f(\cdot)$，它在$(0,1)$以外的地方恒为 0 且 $\int_0^1 f(x)\mathrm{d}x=1$．令 $F(x)=\int_{-\infty}^x f(t)\mathrm{d}t$，易知 $F(\ )$是单调不减、连续函数．在半环 $\mathscr{C}$ 上定义如下的集函数

$$P((a,b])=F(b)-F(a),\quad \forall(a,b]\in\mathscr{C}.$$

由测度扩张定理，P 可扩张为 $\sigma(\mathscr{C})$上的概率测度，至此，本例的概率空间(Ω，$\mathscr{F}$，P)构造完毕．□

注 在本例中，如果认为每个样本点 ω 的出现机会均等，那么可取 $f(\cdot)$为常值，易知，$f(x)=1$，$0<x<1$，而 $F(x)=x$，$0\le x\le 1$．此时，

$$P((a,b])=b-a,\quad \forall(a,b]\in\mathscr{C}.$$

由此构造出的(Ω，$\mathscr{F}$，P)即为区间$(0,1)$上的几何概型问题的概率空间．

在构成概率空间的三要素中，除了样本空间 Ω 是由试验目的及试验条件所决定外，在 Ω 上引入怎样的事件域 $\mathscr{F}$ 要由问题的具体情况来决定．有时可引入最小的事件域 $\mathscr{F}_1=\{\phi,\Omega\}$；而有时又可引入最大的事件域 $\mathscr{F}_2=2^{\Omega}$（即 Ω 的幂集）；一般情况下引入的事件域介于两者之间．例如，取 Ω 的非空真子集 A，令 $\mathscr{F}=\{A,A^c,\varnothing,\Omega\}$，则 $\mathscr{F}$ 是事件域且 $\mathscr{F}_1\subset\mathscr{F}\subset\mathscr{F}_2$．

通常称（Ω，$\mathscr{F}$）为**可测空间**，称 $\mathscr{F}$ 中的元 A 为**可测集**．对可测空间（Ω，$\mathscr{F}$）装备测度 μ，就构成**测度空间**（Ω，$\mathscr{F}$，μ）．若所装备的测度还满足 $\mu(\Omega)=1$，则称（Ω，$\mathscr{F}$，μ）为**概率测度空间**，简称概率空间．按概率论的记法，以 P 替换 μ，记作（Ω，$\mathscr{F}$，P）．

对可测空间（Ω，$\mathscr{F}$）装备概率测度 P，本质上是选择一个定义在 $\mathscr{F}$ 上取值于[0,1]并符合概率三条公理的集函数 P．一般情况下，这样的集函数不会只有一个．因而，概率空间应当根据实际的需要来构造．

由于概率测度 P 只是一种特殊的测度，因而它具有测度应有的那些性质.

概率的所有性质都是在其满足的非负性、规范性及可列可加性这三条公理的基础上演绎出来的.

定理 1.1.5 若（Ω，$\mathscr{F}$，P）为给定的概率空间，则概率（测度）P 具有如下性质：

(1) $P(\varnothing)=0.$

(2) **有限可加性** 若 $A_k\in\mathscr{F}\ (k=1,2,\cdots,n)$，且 $A_iA_j=\varnothing\ \forall\, i\neq j$，则

$$P\left(\sum_{k=1}^{n}A_k\right)=\sum_{k=1}^{n}P(A_k).$$

(3) **可减性与单调性** 设 $A,B\in\mathscr{F}$，若 $A\subset B$，则

$$P(B-A)=P(B)-P(A),$$
$$P(A)\le P(B).$$

(4) 若 $A\in\mathscr{F}$，则 $P(A^c)=1-P(A).$

(5) **Jordan 公式与半有限可加性** 对$\forall\ A_1,A_2,\cdots,A_n\in\mathscr{F}$，有

$$P\left(\bigcup_{k=1}^{n}A_k\right)=\sum_{k=1}^{n}P(A_k)-\sum_{1\le i<j\le n}P(A_iA_j)+\sum_{1\le i<j<k\le n}P(A_iA_jA_k)-\cdots+(-1)^{n+1}P(A_1A_2\cdots A_n).$$

$$P\left(\bigcup_{k=1}^{n}A_k\right)\le\sum_{k=1}^{n}P(A_k).$$

(6) **连续性** 设 $\{A_n,n=1,2,\cdots\}\subset\mathscr{F}$，若$\{A_n，n=1,2,\cdots\}$是单调集列①，则

$$P\left(\lim_{n\to\infty}A_n\right)=\lim_{n\to\infty}P(A_n).$$

证明 性质(1)~(5)的证明留给读者作练习，下面证明性质(6). 先设集列$\{A_n，n=1,2,\cdots\}$是单调增的，即，$A_n\subset A_{n+1}\ (n\ge1)$，此时，$\lim\limits_{n\to\infty}A_n=\bigcup\limits_{k=1}^{\infty}A_k$，令 $B_k=A_k-\bigcup\limits_{i=1}^{k-1}A_i=A_k-A_{k-1}\ (k\ge1)$，其中约定：$A_0=\phi$. 显然，事件列$\{B_n，n=1,2,\cdots\}$两两互斥且$\bigcup\limits_{k=1}^{\infty}A_k=\sum\limits_{k=1}^{\infty}B_k$，故

————————

① 关于集列的单调性与集列的极限概念参见本章附录中的相关内容.

$$
\begin{aligned}
P(\lim_{n\to\infty} A_n) &= P\left(\sum_{k=1}^{\infty} B_k\right) \\
&= \sum_{k=1}^{\infty} P(B_k) \qquad \text{（概率的可列可加性）} \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \left[P(A_k) - P(A_{k-1})\right] \\
&= \lim_{n\to\infty} P(A_n).
\end{aligned}
$$

这一性质称为**概率的下连续性**.

再设集列$\{A_n，n=1,2,\cdots\}$是单调减的，即，$A_n \supset A_{n+1}\ (n\ge 1)$，此时，$\lim\limits_{n\to\infty} A_n = \bigcap\limits_{k=1}^{\infty} A_k$．因

$A_n^c \subset A_{n+1}^c\ (n\ge 1)$，即集列$\{A_n^c, n=1,2,\cdots\}$是单调增的，由上面所证知

$$
P\left(\lim_{n\to\infty} A_n^c\right) = P\left(\bigcup_{k=1}^{\infty} A_n^c\right) = \lim_{n\to\infty} P(A_n^c),
$$

而

$$
P\left(\bigcup_{k=1}^{\infty} A_k^c\right) = 1 - P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1 - P(\lim_{n\to\infty} A_n),
$$
$$
\lim_{n\to\infty} P(A_n^c) = \lim_{n\to\infty}\left[1 - P(A_n)\right] = 1 - \lim_{n\to\infty} P(A_n),
$$

故

$$
P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n).
$$

这一性质称为**概率的上连续性**．概率既有下连续性又有上连续性，即它具有连续性．□

注　(1) 在应用概率的减法公式$P(B-A)=P(B)-P(A)$时，务必注意条件$A\subset B$是否满足，若不然，则结论未必成立．此时，可采用一般情形下的减法公式，

$$
P(B-A) = P(B) - P(AB), \quad \forall A, B \in \mathscr{F}. \tag{1.1.1}
$$

(2)　在 Jordan 公式中取 n = 2、3，就得到两个常用公式

$$
\begin{aligned}
&P(A\cup B) = P(A) + P(B) - P(AB), \\
&P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC).
\end{aligned} \tag{1.1.2}
$$

(3)　由概率的半有限可加性及下连续性可推出下列不等式

$$
P\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} P(A_k), \quad \forall \{A_k, k\ge 1\} \subset \mathscr{F}. \tag{1.1.3}
$$

这一性质称为概率的**半可列可加性**.

利用概率的上述性质便不难证明以下的定理.

定理 1.1.6　(1) (**Borel-Cantelli 引理**) 设$\{A_n, n=1,2,\cdots\}\subset \mathscr{F}$，若$\sum\limits_{n=1}^{\infty} P(A_n) < \infty$，则

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\right)=0. \tag{1.1.4}$$

(2) 设$\{A_n,n=1,2,\cdots\}\subset\mathscr{F}$，若 $A_1, A_2,\cdots, A_n,\cdots$相互独立且$\sum_{n=1}^{\infty}P(A_n)=\infty$，则

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\right)=1. \tag{1.1.5}$$

证明 (1) 事件列$\left\{\bigcup_{k=n}^{\infty}A_k，n=1,2,\cdots\right\}$是单调减的，由概率的上连续性及半可列可加性，得

$$0\le P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\right)=P\left(\lim_{n\to\infty}\bigcup_{k=n}^{\infty}A_k\right)=\lim_{n\to\infty}P\left(\bigcup_{k=n}^{\infty}A_k\right)\le\lim_{n\to\infty}\sum_{k=n}^{\infty}P(A_k)=0,$$

最后一个等号成立是因为级数$\sum_{n=1}^{\infty}P(A_n)<\infty$.

(2) 与(1)的证明的前半部分类似，可得

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\right)=P\left(\lim_{n\to\infty}\bigcup_{k=n}^{\infty}A_k\right)=\lim_{n\to\infty}P\left(\bigcup_{k=n}^{\infty}A_k\right)=\lim_{n\to\infty}\left[1-P\left(\bigcap_{k=n}^{\infty}A_k^c\right)\right].$$

再由独立性及定理条件，知

$$\begin{aligned}
0&\le P\left(\bigcap_{k=n}^{\infty}A_k^c\right)\\
&=\prod_{k=n}^{\infty}\left[1-P(A_k)\right]\\
&\le\prod_{k=n}^{\infty}e^{-P(A_k)} \qquad (\text{对 } x\ge 0，1-x\le e^{-x})\\
&=e^{-\sum_{k=n}^{\infty}P(A_k)}\\
&=0 \qquad (\text{对任意给定的 } n\ge 1，\textstyle\sum_{k=n}^{\infty}P(A_k)=\infty\ ).
\end{aligned}$$

证毕．□

定理 1.1.6 中的事件$\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k$ 称为事件列$\{A_n，n\ge1\}$的上极限事件，它表示“有无穷多个A_n出现”的事件，即

$$\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k=\{\omega:\text{使无穷多个}A_n\text{出现}\}. \tag{1.1.6}$$

另外，事件$\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}A_k$ 称为事件列$\{A_n，n\ge1\}$的下极限事件，表示“至多有有穷个 A_n 不出现”的

事件，亦即

$$\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}A_k=\{\omega:使至多有穷个A_n不出现\}. \tag{1.1.7}$$

(1.1.6)、(1.1.7)两式均可用集合的相互包含的方法来证明（留给读者做练习）. 此外，从集合运算的角度看，上极限事件是“由外向里”收缩，而下极限事件是“由里向外”扩张. 直观上，“至多有有穷个 A_n 不出现”导致“有无穷多个 A_n 出现”，因而下极限事件是上极限事件的子事件. 若两者重合，则表明事件列$\{A_n,\ n\geq1\}$当 $n\to\infty$时有极限.

例 1.1.7 给定事件列$\{A_n, n\geq1\}$，设 $P(A_n)=\frac{1}{n^{\alpha}}, n\geq 1, \alpha>1$. 因 $\sum_{n=1}^{\infty}\frac{1}{n^{\alpha}}<\infty$，由定理 1.1.6(1)知 $P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\right)=0$，亦即 $P\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}A_k^c\right)=1$，故对事件列$\{A_n,\ n\geq1\}$而言，几乎可以肯定至多有有限个 A_n 发生.

例 1.1.8 设$\{X_n,\ n\geq1\}$为独立同 0-1 分布的随机变量序列，因 $P(X_n=1)=p$，$P(X_n=0)=1-p\ (0<p<1)$，又 $\sum_{n=1}^{\infty}P(X_n=1)=\infty$，$\sum_{n=1}^{\infty}P(X_n=0)=\infty$，由定理 1.1.6(2)知，几乎必然地有无穷多个 n 使$(X_n=1)$发生，也几乎必然地有无穷多个 n 使$(X_n=0)$发生，故有 $P\left(\lim_{n\to\infty}X_n不存在\right)=1$。

1.1.2 随机变量及其概率分布

给定概率空间$(\Omega,\ \mathscr{F},\ P)$，设 $X(\cdot)$是定义在 Ω 上，取值于实数集 $\boldsymbol{R}^1$ 的映射，若它满足对任意 $x\in\boldsymbol{R}^1$，都有 $\{\omega: X(\omega)\leq x\}\in\mathscr{F}$，则称 $X(\cdot)$为**随机变量**(**random variable**，缩写 ***r.v.***)，称函数 $F(x)=P\{\omega: X(\omega)\leq x\}$ 为 $X(\cdot)$的**分布函数**.

习惯上常省去 ω，将 $X(\omega)$记作 X，将事件$\{\omega: X(\omega)\leq x\}$记作$\{X\leq x\}$，从而将分布函数记作 $F(x)=P\{X\leq x\}$.

延伸阅读

对给定的概率空间$(\Omega,\ \mathscr{F},\ P)$，若 X 是由可测空间$(\Omega,\ \mathscr{F})$到可测空间$(\boldsymbol{R}^1,\ \mathscr{B}^1)$的一个可测映射，即

$$\{\omega: X(\omega)\in B\}\in\mathscr{F},\quad \forall B\in\mathscr{B}^1,$$

则称 X 为随机变量，称 $P\{\omega: X(\omega)\in B\}$ 为 X 的**概率分布**.

实际上，引入 X 就是对每一样本点 $\omega\in\Omega$ 赋予实数 $X(\omega)$，从而**将试验结果数量化**. 为了将这种赋值纳入概率公理体系，需要对其提出一定的要求，“可测”的要求即表示：当 X 落入任一 Borel 集 B 时，其原象 $X^{-1}(B)\triangleq\{\omega: X(\omega)\in B\}$ 都是事件，从而可以计算相应的概率. 而概率分布 $P\{\omega: X(\omega)\in B\}$ 正是描述 X 取值于 B 的可能性.

记

$$\mu_X(B) \triangleq P \circ X^{-1}(B) \triangleq P\left\{X^{-1}(B)\right\} = P\{\omega : X(\omega) \in B\}, \quad \forall B \in \mathscr{B}^1$$

可以证明 μ_X 是概率测度，称之为由 X **诱导的测度**．将其装备到可测空间($\boldsymbol{R}^1$，$\mathscr{B}^1$)上，就得到新的概率空间($\boldsymbol{R}^1$，$\mathscr{B}^1$，μ_X)．

事实上，通过引入 X 可以将原先的比较抽象的概率空间(Ω，$\mathscr{F}$，P)变换到相对直观、具体的概率空间($\boldsymbol{R}^1$，$\mathscr{B}^1$，μ_X)．

在概率分布中取 $B = (-\infty, x] \in \mathscr{B}^1$，就得到分布函数

$$F(x) = P\{\omega : X(\omega) \in (-\infty, x]\} = P\{\omega : X(\omega) \le x\}, \quad \forall x \in \boldsymbol{R}^1$$

分布函数 $F(\cdot)$的值 $F(x)$刻画 X 取值于 $(-\infty, x]$ 的可能性．

数学上已经证明:

$$\{\omega: X(\omega) \in B\} \in \mathscr{F}, \quad \forall B \in \mathscr{B}^1 \iff \{\omega: X(\omega) \le x\} \in \mathscr{F}, \forall x \in \boldsymbol{R}^1 .$$

这为我们判断一个定义在 Ω 取值于 $\boldsymbol{R}^1$ 的映射 X 是否为随机变量带来方便．

例 1.1.9　给定概率空间(Ω，$\mathscr{F}$，P)，设 $A \in \mathscr{F}$ 是一事件，令

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

证明 X 是定义在 (Ω，$\mathscr{F}$，P)上的随机变量．

证明　对任意实数 x，因

$$\{\omega : X \le x\} = \begin{cases} \varnothing, & x < 0, \\ \bar{A}, & 0 \le x < 1, \\ \Omega, & x \ge 1, \end{cases}$$

故 $\{\omega : X \le x\} \in \mathscr{F}$，即 X 是定义在 (Ω，$\mathscr{F}$，P)上的随机变量．　□

由上例可知，对任一事件 $A \in \mathscr{F}$，总能用随机变量 I_A（称 $I_A(\cdot)$为集合 A 的示性函数）来描述它．反之，定义在概率空间(Ω，$\mathscr{F}$，P)上的任一随机变量 X，通过它取等号或者不等号甚至落入某 Borel 集的方式，可以描述各种各样的事件，它并不是仅由一个事件所决定的．这说明随机变量概念较之随机事件概念更一般，前者包容后者．

由概率的性质可以证明，随机变量 X 的分布函数 $F(\cdot)$具有下列性质

(1) $F(x)$是 x 的单调不减函数；

(2) $F(x)$是 x 的右连续函数；

(3) 若记 $F(-\infty) \triangleq \lim\limits_{x \to -\infty} F(x)$，$F(+\infty) \triangleq \lim\limits_{x \to +\infty} F(x)$，则 $F(-\infty) = 0, F(+\infty) = 1$．

反之，我们有

定理 1.1.10　若实函数 $F(x)\ (x \in \boldsymbol{R}^1)$ 是单调不减、右连续的，且 $F(-\infty) = 0, F(+\infty) = 1$，

则存在一个概率空间及其上的随机变量 X，使 $F_X(x) = F(x)$．

注　定理 1.1.10 称为随机变量的**存在定理**．该定理表明：一个普通实函数只要具备分布

函数的三条性质，那么它一定能作为某随机变量的分布函数．该定理的证明要用到测度论的知识．

证明 考虑可测空间($\boldsymbol{R}^1$，$\mathscr{B}^1$)．先在半环$\mathscr{C}=\{(a,b]:\ a,b\in\boldsymbol{R}^1,a\le b\}$上定义如下的集函数

$$P\big((a,b]\big)=F(b)-F(a),\ \ \forall(a,b]\in\mathscr{C},$$

再由测度扩张定理，P 可扩张为 $\mathscr{B}^1$ 上的概率测度，如此，便构造出概率空间($\boldsymbol{R}^1$，$\mathscr{B}^1$，P)．最后在($\boldsymbol{R}^1$，$\mathscr{B}^1$，P)上定义随机变量 $X(\omega)=\omega$，$\omega\in\boldsymbol{R}^1$，则 $F(\cdot)$即为 X 的分布函数． □

注 如果在上面的证明过程中用任意的单调不减右连续函数 $F(\cdot)$(称这种 $F(\cdot)$为**准分布函数**）来构造 $\mathscr{B}^1$ 上的测度 P 的话，那么此处的 P 就未必是概率测度，称如此构造的测度为 **Lebesgue-Stieltjes 测度**（简称 L-S 测度），且有

$$P(A)=\int_A \mathrm{d}F(x),\ \ \forall A\in\mathscr{B}^1.$$

其中上式右端积分为 Lebesgue-Stieltjes 积分．

若 $F(\cdot)$是准分布函数又满足条件：$F(-\infty)=0,F(+\infty)=1$，则称其为分布函数．由分布函数 $F(\cdot)$构造的 $\mathscr{B}^1$ 上的 L-S 测度是 $\mathscr{B}^1$ 上的概率测度．

直观上，随机变量可分为两种类型：离散型和连续型．

若随机变量 X 的所有可能值是至多可列的：x_1，x_2，…，记

$$p_k\triangleq P\{X=x_k\},\ k=1,2,\cdots,$$

则称 X 为**离散型随机变量**，称$\{p_k,k=1,2,\cdots\}$为 X 的**分布律**或**概率函数**．

显然，分布律（概率函数）具有如下性质

(1) $p_k>0,\ k=1,2,\cdots;$

(2) $\sum\limits_k p_k=1.$

对于离散型随机变量 X，其分布函数 $F(\cdot)$与分布律$\{p_k,k=1,2,\cdots\}$之间存在如下关系

$$F(x)=\sum_{\{k:\,x_k\le x\}}p_k,\ \ \forall x\in\boldsymbol{R}^1;$$

$$p_k\triangleq P\{X=x_k\}=F(x_k)-F(x_k-0).$$

若 X 是非离散型随机变量，其分布函数为 $F(\cdot)$，又存在非负可积函数 $f(\cdot)$，使

$$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t,\quad \forall x\in\boldsymbol{R}^1,$$

则称 X 为**连续型随机变量**，称 $f(\cdot)$为 X 的**概率密度**．

易知，概率密度具有下列性质

(1) $f(x)\ge 0,\ \ -\infty<x<+\infty;$

(2) $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1.$

对于连续型随机变量 X，其分布函数 $F(\cdot)$具有如下性质：

(1)　$F'(x)=f(x)$（几乎处处成立）；

(2)　对任意实常数c，$P\{X=c\}=F(c)-F(c-0)=0$；

(3)　$F(b)-F(a)=\int_a^b f(x)\mathrm{d}x$

注　(1)　在性质(1)中等式“几乎处处成立”的含义是：在数直线 $\boldsymbol{R}^1$ 上挖去一个 Lebesgue 测度为零的集之后，等式处处成立．

(2) 性质(2)表明，连续型随机变量 X 取任何定值的概率皆为 0，并且，其分布函数是左连续的．因任何分布函数都是右连续的，故连续型随机变量的分布函数是连续函数．

(3) 因连续型随机变量 X 具有性质(2)，故它落入以 a 为左端点 b 为右端点的开区间、闭区间、半开半闭区间的概率都相同．

延伸阅读

借助测度的 Lebesgue 分解理论，可以对随机变量的分类问题给出更为全面、细致的解释．

设 $F(\cdot)$是一个分布函数．若存在至多可列集$\{x_n, n=1,2,\cdots\}\subset\boldsymbol{R}^1$及定义在其上的正数列$\{p_n, n=1,2,\cdots\}$，使

$$F(x)=\sum_{\{n:\,x_n\le x\}} p_n\,,\qquad -\infty<x<+\infty$$

则称 $F(\cdot)$为**离散型分布函数**．

若存在非负的 Lebesgue 可测函数 $f(\cdot)$，使

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t,\qquad \forall x\in\boldsymbol{R}^1$$

则称 $F(\cdot)$为**连续型分布函数**．

若 $F(\cdot)$是连续函数，且由 $F(\cdot)$构造的 L-S 测度与 Lebesgue 测度奇异，则称 $F(\cdot)$为**奇异型分布函数**．

可以证明，任何分布函数 $F(\cdot)$都能唯一地分解为下列形式：

$$F(x)=a_dF_d(x)+a_cF_c(x)+a_sF_s(x),\quad -\infty<x<+\infty \tag{1.1.8}$$

其中，a_d, a_c, a_s 为非负常数且 $a_d+a_c+a_s=1$，$F_d(\cdot)$、$F_c(\cdot)$、$F_s(\cdot)$分别为离散型、连续型、奇异型分布函数．

在上述分解式中，当 $a_d=1$ 时，称对应的随机变量为**离散型随机变量**．当 $a_c=1$ 时，称对应的随机变量为**连续型随机变量**．当 $a_s=1$ 时，称对应的随机变量为**奇异型随机变量**．即，随机变量有三种基本类型：离散型、连续型和奇异型．对一般的随机变量而言，因其概率分布是由不同类型的分布混合而成，故称之为混合型随机变量．

例 1.1.11　设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ 1-e^{-\lambda x}, & 0\le x<L,\quad (L>0).\\ 1, & L\le x.\end{cases}$$

显然，$F(\cdot)$不是离散型分布函数（因为它不是阶梯函数），也不是连续型或奇异型分布函数（因为它不是连续函数）．事实上，$F(\cdot)$可分解为

$$F(x)=e^{-\lambda L}F_d(x)+(1-e^{-\lambda L})F_c(x),\quad -\infty<x<+\infty,$$

其中

$$F_d(x)=\begin{cases}0, & x<L,\\ 1, & x\ge L,\end{cases}$$

$$F_c(x)=\begin{cases}0, & x<0,\\ \dfrac{1-e^{-\lambda x}}{1-e^{-\lambda L}}, & 0\le x<L,\\ 1, & L\le x.\end{cases}$$

亦即，X 的概率分布是由离散型分布与连续型分布混合而成，因而它是混合型随机变量.

1.1.3 随机向量及其概率分布

若 $X_1, \cdots$，X_n 是定义在概率空间$(\Omega$，$\mathscr{T}$，$P)$上的 n 个随机变量，则称 $\boldsymbol{X} = (X_1, \cdots, X_n)$为 n 维**随机向量**（**random vector**，缩写 *r.v.*），称 n 元函数

$$F(x_1,\cdots,x_n)\triangleq P\{X_1\le x_1,\cdots,X_n\le x_n\}$$

为 $\boldsymbol{X}$ 的**分布函数**或 $X_1, \cdots$，X_n 的**联合分布函数**.

延伸阅读

给定概率空间$(\Omega$，$\mathscr{T}$，$P)$，若 $\boldsymbol{X} = (X_1, \cdots, X_n)$是由可测空间$(\Omega$，$\mathscr{T})$到可测空间$(\boldsymbol{R}^n,\mathscr{B}^n)$的可测映射，即，

$$\boldsymbol{X}^{-1}(B^{(n)})=\left\{\omega:\left(X_1(\omega),\cdots,X_n(\omega)\right)\in B^{(n)}\right\}\in\mathscr{F},\quad \forall B^{(n)}\in\mathscr{B}^n,$$

则称 $\boldsymbol{X}$ 为 n 维随机向量.

由测度论可知:

$$\boldsymbol{X}^{-1}(B^{(n)})=\left\{\omega:\left(X_1(\omega),\cdots,X_n(\omega)\right)\in B^{(n)}\right\}\in\mathscr{F},\quad \forall B^{(n)}\in\mathscr{B}^n$$

当且仅当对每一 k $(1\le k\le n)$

$$X_k^{-1}(B^{(1)})=\left\{\omega:X_k(\omega)\in B^{(1)}\right\}\in\mathscr{F},\quad \forall B^{(1)}\in\mathscr{B}^1.$$

因而判断一个定义在 Ω 取值于 $\boldsymbol{R}^n$（即，向量值）的映射 $\boldsymbol{X}$ 是否为随机向量，等价于判断它的每一分量是否为随机变量. 这正是本科概率论中随机向量定义的来由.

n 维随机向量 $\boldsymbol{X}$ 的分布函数 F 具有下列性质：

(1) $F(x_1,\cdots, x_n)$关于每一变元 x_i $(1\le i\le n)$是单调不减的.

(2) $F(x_1,\cdots, x_n)$关于每一变元 x_i $(1\le i\le n)$是右连续的.

(3) 若记

$$F(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_n)=\lim_{x_i\to-\infty}F(x_1,\cdots,x_n),$$

$$F(+\infty,\cdots,+\infty)=\lim_{x_1\to+\infty,\cdots,x_n\to+\infty}F(x_1,\cdots,x_n),$$

则

$$F(x_1,\cdots,x_{i-1},-\infty,x_{i+1},\cdots,x_n)=0,$$
$$F(+\infty,\cdots,+\infty)=1.$$

(4) 若 $x_i \le y_i\ (i=1,2,\cdots,n)$，则

$$\Delta_1\Delta_2\cdots\Delta_n F(x_1,\cdots,x_n)=P\left\{x_1<X_1\le y_1,\cdots,x_n<X_n\le y_n\right\}\ge 0, \tag{1.1.9}$$

其中 $\Delta_i F(x_1,\cdots,x_n)\triangleq F(x_1,\cdots,x_{i-1},y_i,x_{i+1},\cdots x_n)-F(x_1,\cdots,x_n)$ 表示 F 对第 i 个变元($1\le i\le n$)作一阶差分.

上述性质的证明留给读者作练习.

注 (1) 性质(4)不能由性质(1), (2), (3)推出. 一般称具有性质(4)的 n 元函数为 ***n* 元单调不减函数**. n 元单调不减函数并不蕴含它关于每一变元是单调不减的.

(2) 在随机向量场合也有所谓的存在定理，我们不加证明地陈述于后："给定 n 元函数 F，若它满足性质(2)、(3)、(4)，则存在概率空间(Ω，$\mathscr{F}$，P)及定义在该空间上的 n 维随机向量 $\boldsymbol{X}$，使得 $F_{\boldsymbol{X}}(x_1,\cdots,x_n)=F(x_1,\cdots,x_n)$."

常见的随机向量也分为两种：离散型和连续型. 对于后者，是指存在非负可积的 n 元函数 f，使

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n\,,\quad -\infty<x_1,\cdots,x_n<+\infty.$$

称 f 为 $\boldsymbol{X}$ 的**概率密度**或 $X_1,\cdots,X_n$ 的**联合概率密度**.

n 维连续型随机向量的概率密度 f 具有如下性质：

(1) $f(x_1,\cdots,x_n)\ge 0,\quad -\infty<x_1,\cdots,x_n<+\infty$；

(2) $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=1$；

(3) $\dfrac{\partial^n F(x_1,\cdots,x_n)}{\partial x_1\cdots\partial x_n}=f(x_1,\cdots,x_n)$ (几乎处处成立)；

(4) $P\left\{\boldsymbol{X}\in B^{(n)}\right\}=\int\cdots\int_{B^{(n)}} f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\,,\quad \forall B^{(n)}\in\mathscr{B}^n$.

在 n 元分布函数 F 中保留 k ($1\le k<n$)个变元，比如 $x_1,\cdots,x_k$，而令其余变元都趋于$+\infty$，得到 ***k* 维边缘分布函数**

$$F(x_1,\cdots,x_k,+\infty,\cdots,+\infty)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_k}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_k\mathrm{d}y_{k+1}\cdots\mathrm{d}y_n.$$

记 $G(x_1,\cdots,x_k)=F(x_1,\cdots,x_k,+\infty,\cdots,+\infty)$. 由上面等式可知 k 维边缘分布函 G 也是连续型的，其概率密度为

$$g(x_1,\cdots,x_k)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1,\cdots,x_k,x_{k+1},\cdots,x_n)\mathrm{d}x_{k+1}\cdots\mathrm{d}x_n\,.$$

由分布函数可以唯一地决定边缘分布函数，反之则未必成立.

例 1.1.10 设两个连续型分布函数 $F(x, y)$和 $G(x, y)$的概率密度分别为

$$f(x,y)=\begin{cases}\dfrac{1}{8}(x+y), & 0\le x,y\le 2,\\ 0, & \text{其他};\end{cases}$$

$$g(x,y)=\begin{cases}\dfrac{1}{16}(1+x)(1+y), & 0\le x,y\le 2,\\ 0, & \text{其他}.\end{cases}$$

显然，$f(\ ;\ )$与$g(\ ;\ )$不同，但它们的边缘概率密度

$$\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\int_0^2 \frac{x+y}{8}\mathrm{d}y=\frac{1+x}{4}\quad (0\le x\le 2),$$

$$\int_{-\infty}^{+\infty} g(x,y)\mathrm{d}y=\int_0^2 \frac{(1+x)(1+y)}{16}\mathrm{d}y=\frac{1+x}{4}\quad (0\le x\le 2)$$

相同，另一对边缘概率密度也相同．由此可见，在知道了 X、Y 的边缘概率密度分别为 $(1+x)/4$，$0\le x\le 2$，$(1+y)/4$，$0\le y\le 2$ 条件下并不能唯一地确定(X,Y)的联合概率密度．

1.1.4 独立性

独立性是概率论中重要的基本概念之一．

(1) 称随机变量 $X_1,\cdots,X_n\,(n\ge 2)$是相互独立的，如果对任意实数 $x_1,\cdots,x_n$，有

$$P\{X_1\le x_1,X_2\le x_2,\cdots,X_n\le x_n\}=P\{X_1\le x_1\}P\{X_2\le x_2\}\cdots P\{X_n\le x_n\}.$$

由此可知，若 $X_1,\cdots,X_n$ 相互独立，则 $X_{i_1},\cdots,X_{i_k}$ 也相互独立，其中$1\le i_1<\cdots<i_k\le n$，$2\le k\le n$.

(2) 称随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$是相互独立的，如果对任意自然数$k\ge 2$，$X_1,X_2,\cdots,X_k$相互独立．

(3) 称随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$与 $\boldsymbol{Y}=(Y_1,\cdots,Y_m)$相互独立，如果对任意实数 $x_1,\cdots,x_n,y_1,\cdots,y_m$，有

$$\begin{aligned}&P\{X_1\le x_1,\cdots,X_n\le x_n,Y_1\le y_1,\cdots,Y_m\le y_m\}\\&=P\{X_1\le x_1,\cdots,X_n\le x_n\}P\{Y_1\le y_1,\cdots,Y_m\le y_m\}.\end{aligned}$$

注意，当向量$\boldsymbol{X}=(X_1,\cdots,X_n)$与向量$\boldsymbol{Y}=(Y_1,\cdots,Y_m)$相互独立时，$X_1,\cdots,X_n$未必相互独立．同理，$Y_1,\cdots,Y_m$也未必相互独立．

类似地，可给出 $n\,(n\ge 3)$个随机向量的相互独立性概念．读者可尝试写出相应的定义来．

(4) 称随机变量序列$\{X_n,n\ge 1\}$与$\{Y_m,m\ge 1\}$是相互独立的，如果对任意自然数 n，m，向量$(X_1,\cdots,X_n)$与$(Y_1,\cdots,Y_m)$相互独立．

例 1.1.11 已知 X_1，X_2，X_3 相互独立，Y_1，Y_2 相互独立，又设向量(X_1,X_2,X_3)与(Y_1,Y_2)独立，证明 X_1，X_2，X_3，Y_1，Y_2 相互独立．

证明 对$\forall x_1,x_2,x_3,y_1,y_2\in\boldsymbol{R}^1$，因

$$\begin{aligned}&P\{X_1\le x_1,X_2\le x_2,X_3\le x_3,Y_1\le y_1,Y_2\le y_2\}\\&=P\{X_1\le x_1,X_2\le x_2,X_3\le x_3\}P\{Y_1\le y_1,Y_2\le y_2\}\\&=P\{X_1\le x_1\}P\{X_2\le x_2\}P\{X_3\le x_3\}P\{Y_1\le y_1\}P\{Y_2\le y_2\},\end{aligned}$$

故 X_1，X_2，X_3，Y_1，Y_2 相互独立． □

定理 1.1.12　若 $X_1, \cdots, X_n$ 相互独立，将其分为 $k\ (2\le k\le n)$个不相重叠的组，记

$$\boldsymbol{Y}_1=(X_1,\cdots,X_{n_1}), \boldsymbol{Y}_2=(X_{n_1+1},\cdots,X_{n_2}), \cdots, \boldsymbol{Y}_k=(X_{n_{k-1}+1},\cdots,X_n),$$

其中$1\le n_1<n_2<\cdots<n_{k-1}<n$．设 g_j 为n_j-n_{j-1}元 Borel 可测函数① $(1\le j\le k$，这里约定 $n_0=0$，$n_k=n)$，则

(1)　$\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_k$相互独立；

(2)　$g_1(\boldsymbol{Y}_1), g_2(\boldsymbol{Y}_2), \cdots, g_k(\boldsymbol{Y}_k)$相互独立．

上述定理表明，如果 n 个随机变量相互独立，把它们分成若干个彼此互不重叠的向量，那么这些向量之间相互独立；这些向量的函数之间也相互独立．

该定理的结论(1)可由随机变量及随机向量的独立性定义直接得到．结论(2)的证明因要用测度论方法，故从略．

1.1.5　随机变量及随机向量的变换

给定概率空间$(\Omega，\mathscr{F}，P)$及定义在$(\Omega，\mathscr{F}，P)$上的随机变量 X，若 g 是 Borel 可测函数，则 $Y=g(X)$是$(\Omega，\mathscr{F}，P)$上的随机变量．

事实上，对任意 Borel 集$B^{(1)}\in\mathscr{B}^1$，因 g 是 Borel 可测函数，故$g^{-1}\left(B^{(1)}\right)\in\mathscr{B}^1$，从而有

$$\left\{\omega: g(X)\in B^{(1)}\right\}=\left\{\omega: X\in g^{-1}\left(B^{(1)}\right)\right\}\in\mathscr{F},$$

即 $Y=g(X)$是$(\Omega，\mathscr{F}，P)$上的随机变量．

对随机变量 X 作变换是概率论中经常会遇到的问题．为了保证对 X 作变换后，新变量 $Y=g(X)$仍然是随机变量，需要对变换 g 作限定，而 Borel 可测就是对变换 g 提出的限定条件．

Borel 可测函数是很广泛的一种函数，在自然科学及工程技术领域中遇到的函数大多为 Borel 可测函数．理论上可以证明，连续函数、单调函数甚至分段单调函数等都是 Borel 可测函数．

Borel 可测函数具有对代数运算和极限运算封闭的特性．即若 f、g 是 Borel 可测函数，则

$$\alpha f+\beta g,\ fg,\ f/g\ (g\ne 0),\ |f|,\ \max\{f,g\},\ \min\{f,g\}$$

都是 Borel 可测函数．若$\{f_n, n\ge 1\}$是一列Borel可测函数，则$\underline{\lim}_{n\to\infty} f_n$，$\overline{\lim}_{n\to\infty} f_n$，$\lim_{n\to\infty} f_n$　（如果相应的极限存在且有限的话）也都是 Borel 可测函数．

设 $\boldsymbol{X}=(X_1, \cdots, X_n)$为 n 维随机向量，

(1)　若 g 是 n 元 Borel 可测函数，则 $Y=g(\boldsymbol{X})$是随机变量；

(2)　若 $g_j\ (1\le j\le m)$是 n 元 Borel 可测函数，则 $\boldsymbol{Y}=(Y_1, \cdots, Y_m)$是 m 维随机向量，其中

$Y_j=g_j(\boldsymbol{X}),\ 1\le j\le m.$

对于连续型随机向量的变换，我们不加证明地给出如下定理：

定理 1.1.13　设 n 维连续型随机向量 $\boldsymbol{X}=(X_1, \cdots, X_n)$的概率密度为 $f_{\boldsymbol{X}}(x_1, \cdots, x_n)$，又设 n 元函数 $y_j(x_1, \cdots, x_n)\ (j=1, 2, \cdots, n)$满足条件

(1)　存在唯一反函数 $x_j(y_1, \cdots, y_n)$，即，方程

---

①　有关 Borel 可测函数概念可参见附录中的相关内容．

$$\begin{cases} y_1(x_1,\cdots,x_n)=y_1, \\ \quad\vdots \\ y_n(x_1,\cdots,x_n)=y_n, \end{cases} \tag{1.1.8}$$

存在唯一实数解 $x_j(y_1,\cdots,y_n)$ $(j=1,2,\cdots,n)$；

(2) $y_j(x_1,\cdots,x_n)$及$x_j(y_1,\cdots,y_n)$都连续，$\dfrac{\partial x_i}{\partial y_j},\dfrac{\partial y_i}{\partial x_j}$存在且连续，$i,j=1,2,\cdots,n$，记

$$J \triangleq \frac{\partial(x_1,\cdots,x_n)}{\partial(y_1,\cdots,y_n)}=\begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix},$$

则 n 维随机向量 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)$（其中 $Y_j=y_j(x_1,\cdots,x_n)$, $j=1,2,\cdots,n$）的概率密度为

$$f_{\boldsymbol{Y}}(y_1,\cdots,y_n)=f_{\boldsymbol{X}}\left(x_1(y_1,\cdots,y_n),\cdots,x_n(y_1,\cdots,y_n)\right)|J|. \tag{1.1.9}$$

例 1.1.14 设 X, Y 相互独立，它们的概率密度分别为 f_X，f_Y，求 $U=X+Y$ 的概率密度.

解 令 $u=x+y$，补充令 $v=x-y$，两方程联立，解得

$$x=\frac{u+v}{2},\ \ y=\frac{u-v}{2}.$$

向量(X, Y)变换到向量(U, V)的 Jacobi 行列式为

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix}=-1/2.$$

由(1.1.9)式，向量(U, V)的概率密度为

$$f_{(U,V)}(u,v)=f_{(X,Y)}\left(\frac{u+v}{2},\ \frac{u-v}{2}\right)\left|-\frac{1}{2}\right|=\frac{1}{2}f_X\left(\frac{u+v}{2}\right)f_Y\left(\frac{u-v}{2}\right),$$

故 $U=X+Y$ 的概率密度为

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{+\infty} f_{(U,V)}(u,v)\mathrm{d}v \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} f_X\left(\frac{u+v}{2}\right)f_Y\left(\frac{u-v}{2}\right)\mathrm{d}v \\ &= \int_{-\infty}^{+\infty} f_X(x)f_Y(u-x)\mathrm{d}x. \end{aligned}$$

最后一个等号是将积分变元 v 替换为 $x=(u+v)/2$ 而得到的．若替换为 $y=(u-v)/2$，则得

$$f_U(u)=\int_{-\infty}^{+\infty} f_X(u-y)f_Y(y)\mathrm{d}y,$$

即

$$f_U(u)=\int_{-\infty}^{+\infty} f_X(x)f_Y(u-x)\mathrm{d}x=\int_{-\infty}^{+\infty} f_X(u-y)f_Y(y)\mathrm{d}y.$$

这就是众所周知的求独立随机变量之和的概率密度的卷积公式． □

用本例的方法，还顺便得到 $V=X-Y$ 的概率密度

$$\begin{aligned}f_V(v)&=\int_{-\infty}^{+\infty}f_{(U,V)}(u,v)\mathrm{d}u\\&=\int_{-\infty}^{+\infty}f_X\left(\frac{u+v}{2}\right)f_Y\left(\frac{u-v}{2}\right)\frac{1}{2}\mathrm{d}u\\&=\int_{-\infty}^{+\infty}f_X(x)f_Y(x-v)\mathrm{d}x\\&=\int_{-\infty}^{+\infty}f_X(y+v)f_Y(y)\mathrm{d}y.\end{aligned}$$

在本例的解题中，如果选择补充 $v=x$，那么按照相同的思路，也可以导出 $U=X+Y$ 的概率密度的计算公式．读者不妨一试．

如果(1.1.8)式有多个解

$$\begin{cases}x_1^{(k)}=x_1^{(k)}(y_1,\cdots,y_n)\\ \cdots\cdots\\ x_n^{(k)}=x_n^{(k)}(y_1,\cdots,y_n)\end{cases}\quad(k=1,2,\cdots),$$

那么定理的结论相应地修正为

$$f_{\boldsymbol{Y}}(y_1,\cdots,y_n)=\sum_k f_{\boldsymbol{X}}\left(x_1^{(k)}(y_1,\cdots,y_n),\cdots,x_n^{(k)}(y_1,\cdots,y_n)\right)\left|J^{(k)}\right|,\tag{1.1.10}$$

其中

$$J^{(k)}=\frac{\partial(x_1^{(k)},\cdots,x_n^{(k)})}{\partial(y_1,\cdots,y_n)},\ k=1,2,\cdots.$$

例 1.1.15 设 X，Y 是两个独立的标准正态变量，试证：$U=X^2+Y^2$ 与 $V=\dfrac{X}{Y}$ 相互独立．

证明 (X,Y)的概率密度为

$$f_{(X,Y)}(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\}\quad(-\infty<x,\ y<+\infty).$$

由方程组 $\begin{cases}u=x^2+y^2\\ v=\dfrac{x}{y}\end{cases}$，解得

$$\begin{cases}x^{(1)}=\dfrac{v\sqrt{u}}{\sqrt{1+v^2}}\\ y^{(1)}=\dfrac{\sqrt{u}}{\sqrt{1+v^2}}\end{cases}\quad(u\ge 0,\ -\infty<v<+\infty),\qquad \begin{cases}x^{(2)}=-\dfrac{v\sqrt{u}}{\sqrt{1+v^2}}\\ y^{(2)}=-\dfrac{\sqrt{u}}{\sqrt{1+v^2}}\end{cases}\quad(u\ge 0,\ -\infty<v<+\infty).$$

算出

$$J^{(1)}=\frac{\partial(x^{(1)},y^{(1)})}{\partial(u,v)}=-\frac{1}{2(1+v^2)},\quad J^{(2)}=\frac{\partial(x^{(2)},y^{(2)})}{\partial(u,v)}=-\frac{1}{2(1+v^2)},$$

因而(U,V)的概率密度为

$$f_{(U,V)}(u,v)=\sum_{k=1}^{2} f_{(X,Y)}\left(x^{(k)}(u,v), y^{(k)}(u,v)\right)\left|J^{(k)}\right|$$
$$=\frac{1}{2\pi(1+v^2)}e^{-\frac{u}{2}} \quad (u\ge 0,\ -\infty<v<+\infty),$$

易知 U 的概率密度为

$$f_U(u)=\begin{cases}\frac{1}{2}e^{-\frac{u}{2}}, & u\ge 0,\\ 0, & u<0.\end{cases}$$

V 的概率密度为

$$f_V(v)=\frac{1}{\pi(1+v^2)} \quad (-\infty<v<+\infty).$$

前者为指数分布，后者为 Cauchy 分布．且有

$$f_{(U,V)}(u,v)=f_U(u)f_V(v) \quad (u\ge 0,\ -\infty<v<+\infty),$$

故 U 与 V 相互独立．□

1.1.6 随机变量的数字特征

设随机变量 X 的分布函数为 F_X，若 $\int_{-\infty}^{+\infty}|x|\mathrm{d}F_X(x)<+\infty$，则称

$$EX \triangleq \int_{-\infty}^{+\infty} x\,\mathrm{d}F_X(x) \tag{1.1.11}$$

为 X 的**数学期望**．X 的数学期望也简称为**期望**或**均值**．

在工程应用中，(1.1.11)式的积分大多理解为 Riemann-Stieltjes 积分(以下简称 R-S 积分)．

根据 R-S 积分性质，若随机变量 X 为离散型，其分布律为 $P\{X=x_k\}$, $k=1,2,\cdots$，则(1.1.11)式化为

$$EX=\sum_k x_k P\{X=x_k\}.$$

若随机变量 X 为连续型，其概率密度为 f_X，则(1.1.11)式化为

$$EX=\int_{-\infty}^{+\infty} x\, f_X(x)\mathrm{d}x.$$

一般地，设 g 是一元 Borel 可测函数，若 $\int_{-\infty}^{+\infty}|g(x)|\mathrm{d}F_X(x)<+\infty$，则

$$E\left[g(X)\right]=\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x)$$
$$=\begin{cases}\sum_k g(x_k)P\{X=x_k\}, & \text{当} X \text{为离散型},\\ \int_{-\infty}^{+\infty} g(x)f_X(x)\mathrm{d}x, & \text{当} X \text{为连续型}.\end{cases}$$

在上式中，分别取 $g(x)=(x-EX)^2$, x^k, $(x-EX)^k$ (其中 $k>0$ 为实常数)，就得到方差、k 阶原点矩及 k 阶中心矩的计算公式．

为了便于后面的学习，下面对期望的性质作简要回顾并补充．在这些性质中，记号 X、Y、X_i ($i=1, 2, \cdots, n$)等代表在给定概率空间(Ω, $\mathscr{F}$, P)上的随机变量，C、C_i ($i=1, 2, \cdots, n$)等代表

常数，g、h 等代表 Borel 可测函数，并且假定所提及的随机变量的期望都是存在的.

(1) $E(C)=C$.

(2) $E\left(\sum_{i=1}^{n} C_i X_i\right)=\sum_{i=1}^{n} C_i EX_i$.

(3) 若 X, Y 独立，则 $E[g(X)h(Y)]=Eg(X)\cdot Eh(Y)$. 特别， $E(XY)=EX\cdot EY$.

此性质可推广到 n 个相互独立的随机变量的情形.

(4) 若 $X\geq 0$，则 $EX\geq 0$. 进而，若 $X\geq Y$，则 $EX\geq EY$.

(5) $|EX|\leq E|X|$.

(6) (**Cauchy-Schwarz 不等式**) 若 $E|X|^2$，$E|Y|^2$ 存在，则 $E|XY|$存在，且有

$$|E(XY)|^2\leq [E|XY|]^2\leq E|X|^2\, E|Y|^2,$$

等价地

$$|E(XY)|\leq E|XY|\leq \sqrt{E|X|^2\, E|Y|^2}.$$

(7) 若 $E(X^k)$存在（其中 $k>0$ 为实常数），则对任意的 $0\leq r<k$，$E(X^r)$也存在.

(8) 若 $X\geq 0$，则 $EX=0$ 的充分必要条件是 $P\{X=0\}=1$.

考虑到上述性质大多已在本科概率论中学习过，而对那些新补充的性质的严格论证又要用到测度论的方法，故不在这里一一地证明.

延伸阅读

在近代概率论中，随机变量 X 的期望被定义为在给定的概率空间(Ω，$\mathscr{F}$，P)上求 X 的抽象积分.

设 X 是概率空间(Ω，$\mathscr{F}$，P)上的随机变量，若$\int_\Omega |X(\omega)|\mathrm{d}P<+\infty$，则称$\int_\Omega X(\omega)\mathrm{d}P$ 为 X 的期望，记作 EX，即

$$EX=\int_\Omega X(\omega)\mathrm{d}P \tag{1.1.12}$$

一般地，设 X 是概率空间(Ω，$\mathscr{F}$，P)上的随机变量，$y=g(\cdot)$为一元 Borel 可测函数，若$\int_\Omega \left|g\left(X(\omega)\right)\right|\mathrm{d}P<+\infty$，则称$\int_\Omega g\left(X(\omega)\right)\mathrm{d}P$ 为 $Y=g(X)$的期望，记作 $EY=E[g(X)]$，即

$$EY=E[g(X)]=\int_\Omega g\left(X(\omega)\right)\mathrm{d}P$$

由此可见，无论是 X 本身还是 X 的函数，它们的期望可统一地定义为在 Ω 上关于概率测度 P 的抽象积分.

在 Ω 上关于概率测度 P 的抽象积分毕竟与普通的 Riemann 积分差别甚大也不便于计算，为此，我们介绍下面的积分转化定理.

定理 1.1.16（积分转化定理） 设 X 是概率空间(Ω，$\mathscr{F}$，P)上的随机变量，其分布函数为 F_X，若 $g(\cdot)$为一元 Borel 可测函数，则

$$\int_\Omega g\left(X(\omega)\right)\mathrm{d}P=\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x). \tag{1.1.13}$$

(1.1.13)式的意义是如果一方可积，那么另一方也可积且两端积分相等.

在(1.1.13)式中取 $g(x)=x$，则有

$$\int_{\Omega} X(\omega)\mathrm{d}P=\int_{-\infty}^{+\infty} x\,\mathrm{d}F_X(x).$$

这表明 EX 的两种定义方式((1.1.11)式与(1.1.12)式)并不矛盾.

(1.1.13)式的右端为 Lebesgue-Stieltjes 积分(简称 L-S 积分). 积分转化定理的意义在于，将概率空间(Ω，$\mathscr{F}$，P)上 $g(X)$关于概率测度 P 的抽象积分转化为在($\boldsymbol{R}^1$，$\mathscr{B}^1$，μ_X)上 g 关于 F_X 的 L-S 积分.

不过，求 g 的 L-S 积分仍不方便. 人们经研究发现，在一定条件下，若 g (关于 F_X)是 Riemann-Stieltjes 可积(简称 R-S 可积)，则它必是(关于 F_X) L-S 可积且两种积分相等，即

$$(L\text{-}S)\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x)=(R\text{-}S)\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x). \qquad (1.1.14)$$

而 R-S 积分与普通的 Riemann 积分相近，也较容易计算.

利用分布函数的分解式(1.1.6)，可得

$$(L\text{-}S)\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x)=(L\text{-}S)\int_{-\infty}^{+\infty} g(x)a_d\mathrm{d}F_d(x)+(L\text{-}S)\int_{-\infty}^{+\infty} g(x)a_c\mathrm{d}F_c(x)$$
$$+(L\text{-}S)\int_{-\infty}^{+\infty} g(x)a_s\mathrm{d}F_s(x).$$

如果 F_X 中不含奇异型分布（即 $a_s=0$），又 g 关于 F_d 、F_c 是 R-S 可积，那么根据(1.1.14)式及 R-S 积分的性质，上式化为

$$(L\text{-}S)\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x)=a_d\sum_k g(x_k)\left[F_d(x_k)-F_d(x_k-0)\right]+a_c\int_{-\infty}^{+\infty} g(x)f_c(x)\mathrm{d}x \qquad (1.1.15)$$

当随机变量 X 的分布仅含离散型分布或仅含连续型分布时，由(1.1.15)式可知，计算 $g(X)$ 的期望有如下的简单公式

$$E\left[g(X)\right]=\begin{cases}\displaystyle\sum_k g(x_k)P\{X=x_k\}, & \text{当}a_d=1,\ \text{即}X\text{ 是离散型},\\ \displaystyle\int_{-\infty}^{+\infty} g(x)f_c(x)\mathrm{d}x, & \text{当}a_c=1,\ \text{即}X\text{ 是连续型}.\end{cases}$$

这与本科概率论中，求随机变量函数的期望的计算公式一致.

例 1.1.17 计算例 1.1.9 中的 X 的期望 EX.

解 例 1.1.9 中的 X 为混合型随机变量 . 由(1.1.15)式，得

$$EX=e^{-\lambda L}\cdot L\cdot 1+(1-e^{-\lambda L})\int_0^L x\frac{\lambda e^{-\lambda x}}{1-e^{-\lambda L}}\mathrm{d}x=\frac{1-e^{-\lambda L}}{\lambda}. \quad \square$$

如果随机变量 X 的分布中含有奇异型分布，那就没有什么好的简便方法能算出积分 $\int_{-\infty}^{+\infty} g(x)\mathrm{d}F_X(x)$，也就很难求得 $g(X)$ 的期望. 这就是本科概率论只介绍离散型与连续型随机变量的缘故.

1.1.7 条件数学期望

条件数学期望是概率论中很重要的基本概念之一，对于随机过程理论尤其如此. 让我们先从条件概率说起.

给定概率空间(Ω，$\mathscr{F}$，P)，设 $A, B\in\mathscr{F}$为两个事件，当 $P(B)>0$ 时，定义

$$P(A\,|\,B)\triangleq\frac{P(AB)}{P(B)},$$

称 $P(A\,|\,B)$为在 B 发生的条件下，A 发生的**条件概率**.

设 X 为(Ω，$\mathscr{F}$，P)上的随机变量，对事件 B（$P(B)>0$），称

$$F(x\mid B)\triangleq P\{X\le x\mid B\},\quad \forall x\in \boldsymbol{R}^1$$

为在 B 发生的条件下，X 的**条件分布函数**．为简明起见，将 $F(x\mid B)$称为 X 关于事件 B 的条件分布函数．

条件分布函数 $F(x\mid B)$也可分为离散型与连续型．对于后者，若存在非负可积函数 $f(x\mid B)$，使

$$F(x\mid B)=\int_{-\infty}^{x} f(t\mid B)\mathrm{d}t,\quad \forall x\in \boldsymbol{R}^1,$$

则称 $f(x\mid B)$为在 B 发生的条件下，X 的**条件概率密度**(简称 X 关于 B 的条件概率密度)．

与数学期望的定义方式类似，定义 X 关于 B 的**条件数学期望**（简称**条件期望**）为

$$E(X\mid B)\triangleq\int_{-\infty}^{+\infty} x\,\mathrm{d}F(x\mid B). \tag{1.1.16}$$

当 X 关于 B 有条件分布律 $P\{X=x_k\mid B\}(k=1,2,\cdots)$ 时，(1.1.16)化为

$$E(X\mid B)=\sum_k x_k P\{X=x_k\mid B\}.$$

当 X 关于 B 有条件概率密度 $f(x\mid B)$时，(1.1.16)化为

$$E(X\mid B)=\int_{-\infty}^{+\infty} x f(x\mid B)\mathrm{d}x.$$

例 1.1.18 设 $X\sim N(\mu,\sigma^2)$，$B=\{X>x_0\}$，求 X 关于 B 的条件概率密度 $f(x\mid B)$及条件期望 $E(X\mid B)$.

解 因 $F(x\mid B)=P\{X\le x\mid B\}=\dfrac{P\{x_0<X\le x\}}{P\{X>x_0\}}$

$$=\begin{cases}\dfrac{F_X(x)-F_X(x_0)}{1-F_X(x_0)}, & x>x_0,\\ 0, & x\le x_0,\end{cases}$$

故

$$f(x\mid B)=F'(x\mid B)=\frac{f_X(x)}{1-F_X(x_0)},\quad x>x_0 .$$

而

$$\begin{aligned}E(X\mid B)&=\int_{-\infty}^{+\infty} x f(x\mid B)\mathrm{d}x=\frac{1}{1-F_X(x_0)}\int_{x_0}^{+\infty} x f_X(x)\mathrm{d}x\\
&=\frac{1}{1-F_X(x_0)}\int_{x_0}^{+\infty} x\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x\\
&=\mu+\frac{\sigma}{\sqrt{2\pi}\left[1-\Phi\left(\dfrac{x_0-\mu}{\sigma}\right)\right]}e^{-\frac{(x_0-\mu)^2}{2\sigma^2}},\end{aligned}$$

其中 Φ 为标准正态分布的分布函数．□

在上例中，若取 $x_0=\mu$，则有

$$f(x \mid X>\mu)=2 f_X(x)=\frac{2}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x>\mu,$$

$$E(X \mid X>\mu)=\mu+\sigma\sqrt{\frac{2}{\pi}}.$$

现在我们来考虑一个随机变量关于另一个随机变量的条件期望问题. 设 X, Y 为随机变量，X 关于 $Y=y$ 的条件期望定义为

$$E(X \mid Y=y) \triangleq \int_{-\infty}^{+\infty} x \mathrm{d} F_{X \mid Y}(x \mid y) = \begin{cases} \sum_k x_k P\{X=x_k \mid Y=y\}, & \text{当 } X,Y \text{为离散型}, \\ \int_{-\infty}^{+\infty} x f_{X \mid Y}(x \mid y) \mathrm{d} x, & \text{当 } X,Y \text{为连续型}. \end{cases} \tag{1.1.17}$$

其中

$$F_{X \mid Y}(x \mid y) \triangleq P\{X \leq x \mid Y=y\}$$

是 X 关于 $Y=y$ 的条件分布函数.

$$P\{X=x_k \mid Y=y\}=\frac{P\{X=x_k, Y=y\}}{P\{Y=y\}}, \quad k=1,2,\cdots$$

是 X 关于 $Y=y$ 的条件分布律.

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

是 X 关于 $Y=y$ 的条件概率密度.

注 (1) 当(X, Y)为离散型随机向量时，对 Y 的每一可能值 y，因 $P\{Y=y\}>0$，故 X 关于 $Y=y$ 的条件分布函数及条件分布律均无歧义且两者之间有如下关系：

$$F_{X \mid Y}(x \mid y)=\sum_{\{k:\, x_k \leq x\}} P\{X=x_k \mid Y=y\}, \quad \forall x \in \boldsymbol{R}^1,$$

$$P\{X=x_k \mid Y=y\}=F_{X \mid Y}(x_k \mid y)-F_{X \mid Y}(x_k-0 \mid y), \; k=1,2,\cdots$$

(2) 当(X, Y)为连续型随机向量时，因对任意实数 y，有 $P\{Y=y\}=0$，故 X 关于 $Y=y$ 的条件分布函数 $F_{X \mid Y}(x \mid y)$不能理解为条件概率 $P\{X \leq x \mid Y=y\}$，而应理解为如下的极限

$$F_{X \mid Y}(x \mid y) \triangleq \lim_{\delta \to 0^+} P\{X \leq x \mid y-\delta<Y \leq y+\delta\}.$$

当然，这是在右端的极限存在时. 如果极限不存在就规定其为 0.

因

$$P\{X \leq x \mid y-\delta<Y \leq y+\delta\}=\frac{\int_{-\infty}^{x} \mathrm{d} u \int_{y-\delta}^{y+\delta} f(u, v) \mathrm{d} v}{\int_{y-\delta}^{y+\delta} f_Y(v) \mathrm{d} v} = \frac{\int_{-\infty}^{x} \mathrm{d} u \frac{1}{2\delta} \int_{y-\delta}^{y+\delta} f(u, v) \mathrm{d} v}{\frac{1}{2\delta} \int_{y-\delta}^{y+\delta} f_Y(v) \mathrm{d} v},$$

故在几乎处处的意义下，有

$$F_{X|Y}(x|y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}\mathrm{d}u.$$

从而记 $f_{X|Y}(x|y)\triangleq f(x,y)/f_Y(y)$，并称之为 X 关于 $Y=y$ 的条件概率密度.

例 1.1.19 设随机向量 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$，即$(X,Y)$具有概率密度

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\},$$

则 X 关于 $Y=y$ 的条件概率密度为

$$\begin{aligned}f_{X|Y}(x|y)&=\frac{f(x,y)}{f_Y(y)}\\&=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right]^2\right\}.\end{aligned}$$

这说明 X 关于 $Y=y$ 的条件分布仍为正态分布，两个参数分别是 $\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\ \sigma_1^2(1-\rho^2)$，

记作 $X|Y=y\sim N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\ \sigma_1^2(1-\rho^2)\right)$. 于是，$X$ 关于 $Y=y$ 的条件期望为

$$E\left(X|Y=y\right)=\int_{-\infty}^{+\infty}x f_{X|Y}(x|y)\mathrm{d}x=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2).$$

由 $E(X\mid Y=y)$的定义及上例可知，$E(X\mid Y=y)$是 y 的函数. 现在，通过这一函数来引入随机变量 $E(X\mid Y)$，其对应法则为：当 $Y=y$ 时，$E(X\mid Y)$的取值为 $E(X\mid Y=y)$. 或者我们换一种表述方式：

$$E(X|Y)\triangleq E(X|Y=y)\big|_{y=Y}.$$

称 $E(X\mid Y)$为 X 关于 Y 的条件期望.

拿上面的例子来说，如果要求 X 关于 Y 的条件期望 $E(X\mid Y)$，那么

$$E(X|Y)=E(X|Y=y)\big|_{y=Y}=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(Y-\mu_2).$$

由此可见，对二元正态向量(X,Y)，X 关于 Y 的条件期望 $E(X\mid Y)$是 Y 的线性函数. 这是多元正态分布的一条重要性质.

下面介绍条件期望的性质，在这些性质当中均假定所提及的随机变量是绝对可积的（即 $E|X|<+\infty$），g、h 是 Borel 可测函数.

(1) $E[g(X)|X]=g(X)$ *a.s.*①. 特别地，$E(X|X)=X$ *a.s.*.

(2) $E\left[g(X)h(Y)|X\right]=g(X)E\left[h(Y)|X\right]$ *a.s.*. 特别地，$E(XY|X)=XE(Y|X)$ *a.s.*.

(3) 若 X 与 Y 独立，则 $E(X|Y)=EX$ *a.s.*. 特别地，若 C 为常数，则 $E(C|Y)=C$ *a.s.*.

(4) $E(aX+bY|Z)=aE(X|Z)+bE(Y|Z)\quad a.s.$，其中 a, b 为常数.

(5) 若 $X_1 \le X_2$，则 $E(X_1|Y) \le E(X_2|Y)$ $a.s.$. 特别地，$|E(X|Y)| \le E(|X||Y)$ $a.s.$.

(6) 全期望公式

$$E\left[E(X|Y)\right]=E(X). \tag{1.1.18}$$

上述性质的严格论证要用到测度论的方法. 这里仅就所涉及的随机变量均为离散型的情形给予一一验证.

证明 (1) 因 $E\left[g(X)|X=x\right]=\sum_k g(x_k)P\left\{X=x_k|X=x\right\}=g(x)\cdot 1=g(x)$，这里 x 是 X 的一个可能值（若不然，则规定 $E\left[g(X)|X=x\right]$ 为 0. 在其后的证明中也都遵循这一规定，不再一一申明.），故 $E\left[g(X)|X\right]=E\left[g(X)|X=x\right]\big|_{x=X}=g(X)$ $a.s.$.

等式 $E\left[g(X)|X=x\right]=g(x)$ 的直观意义是比较明显的.

(2) 因

$$\begin{aligned} E\left[g(X)h(Y)|X=x\right] &= \sum_j\sum_k g(x_k)h(y_j)P\left\{X=x_k,Y=y_j|X=x\right\} \\ &= g(x)\sum_j h(y_j)P\left\{Y=y_j|X=x\right\} \\ &= g(x)E(Y|X=x), \end{aligned}$$

故 $E\left[g(X)h(Y)|X\right]=g(x)E\left[h(Y)|X=x\right]\big|_{x=X}=g(X)E\left[h(Y)|X\right]$ $a.s.$.

① 对一个与 ω 有关的命题 p(ω)，若有 $P\{\omega$：使 p(ω)成立$\}=1$，则称命题 p(ω)几乎必然成立，记作，命题 p(ω)成立 $a.s.$，这里的 $a.s.$是 almost surely 的缩写.

(3) 由

$$\begin{aligned} E(X|Y=y) &= \sum_k x_k P\left\{X=x_k|Y=y\right\} \\ &= \sum_k x_k P\left\{X=x_k\right\} \qquad (\because X,Y\text{独立}) \\ &= EX \end{aligned}$$

得 $E(X|Y)=E(X|Y=y)\big|_{y=Y}=EX$ $a.s.$.

因常数 C 与任何随机变量都独立，故 $E(C|Y)=C\quad a.s.$.

(4) 在 $Z=z$ 的条件下，X、Y 以及(X,Y)的条件分布律为

$$P\left\{X=x_k|Z=z\right\},\ P\left\{Y=y_j|Z=z\right\},\ P\left\{X=x_k,Y=y_j|Z=z\right\},\quad k,j=1,2,\cdots.$$

因为

$$P\left\{X=x_k \middle| Z=z\right\}=\sum_j P\left\{X=x_k, Y=y_j \middle| Z=z\right\}, \quad k=1,2,\cdots,$$

$$P\left\{Y=y_j \middle| Z=z\right\}=\sum_k P\left\{X=x_k, Y=y_j \middle| Z=z\right\}, \quad j=1,2,\cdots,$$

又因

$$\begin{aligned}
E\left(aX+bY \middle| Z=z\right) &= \sum_k \sum_j (ax_k+by_j) P\left\{X=x_k, Y=y_j \middle| Z=z\right\} \\
&= \sum_k \sum_j ax_k P\left\{X=x_k, Y=y_j \middle| Z=z\right\} + \sum_k \sum_j by_j P\left\{X=x_k, Y=y_j \middle| Z=z\right\} \\
&= a\sum_k x_k P\left\{X=x_k \middle| Z=z\right\} + b\sum_j y_j P\left\{Y=y_j \middle| Z=z\right\} \\
&= aE(X \mid Z=z)+bE(Y \mid Z=z),
\end{aligned}$$

所以

$$\begin{aligned}
E\left(aX+bY \middle| Z\right) &= E\left(aX+bY \middle| Z=z\right)\Big|_{z=Z} \\
&= aE(X \mid Z=z)\big|_{z=Z} + bE(Y \mid Z=z)\big|_{z=Z} \\
&= aE(X \mid Z)+bE(Y \mid Z).
\end{aligned}$$

(5) 令 $X=X_2-X_1$，由已证的性质(4)可知，只需证明“若 $X\ge 0$，则 $E(X \mid Y)\ge 0$ $a.s.$.”即可．当 $Y=y$ 时，因

$$\begin{aligned}
E(X \mid Y=y) &= \sum_k x_k P\left\{X=x_k \middle| Y=y\right\} \\
&\ge 0 \quad (\because X=x_k\ge 0, k=1,2,\cdots),
\end{aligned}$$

故

$$E(X \mid Y)=E\left(X \middle| Y=y\right)\Big|_{y=Y}\ge 0 \ \ a.s..$$

由 $-|X|\le X\le |X|$ 及性质(5)已证明的结论，得 $-E\left(|X| \middle| Y\right)\le E(X \mid Y)\le E\left(|X| \middle| Y\right)$ $a.s.$，即

$$\left|E(X \mid Y)\right|\le E(|X| \mid Y) \ \ a.s..$$

(6) 设 X, Y 的分布律为 $P\left\{X=x_k\right\}, k=1,2,\cdots$ 和 $P\left\{Y=y_j\right\}, j=1,2,\cdots$．因 $E(X \mid Y)$ 是 Y 的函数，由随机变量函数的数学期望公式及(1.1.17)式，有

$$\begin{aligned}
E\left[E(X \mid Y)\right] &= \sum_j E\left(X \middle| Y=y_j\right) P\left\{Y=y_j\right\} \\
&= \sum_j \left[\sum_k x_k P\left\{X=x_k \middle| Y=y_j\right\}\right] P\left\{Y=y_j\right\} \\
&= \sum_j \sum_k x_k P\left\{X=x_k, Y=y_j\right\} \\
&= \sum_k x_k \left[\sum_j P\left\{X=x_k, Y=y_j\right\}\right] \\
&= \sum_k x_k P\left\{X=x_k\right\} = EX.
\end{aligned}$$

至此，条件期望的性质(1)~(6)验证完毕．□

全期望公式(1.1.18)可改写为如下便于应用的形式

$$E(X)=\int_{-\infty}^{+\infty} E(X \mid Y=y)\mathrm{d}F_Y(y). \tag{1.1.19}$$

在上式中，令 $X=1_A$（即，事件 A 的示性函数），注意到

$$E(X)=P(A),$$
$$E(X\mid Y=y)=P(A\mid Y=y),$$

可得

$$P(A)=\int_{-\infty}^{+\infty} P(A\mid Y=y)\mathrm{d}F_Y(y). \tag{1.1.20}$$

称(1.1.20)式为全概率公式的一般形式．由此可见，全期望公式是全概率公式的推广和发展．

例 1.1.20 设$(X,Y)\sim N(0,1;0,1;\rho)$，求$E(X^2Y^2)$.

解 由全期望公式及条件期望的性质(2)，得

$$E(X^2Y^2)=E\left[E(X^2Y^2\mid Y)\right]=E\left[Y^2E(X^2\mid Y)\right].$$

由例 1.1.19 知，$X\mid Y=y\sim N(\rho y,\ 1-\rho^2)$，故

$$E(X^2\mid Y=y)=(1-\rho^2)+(\rho y)^2,$$
$$E(X^2\mid Y)=(1-\rho^2)+(\rho Y)^2.$$

注意到

$$E(Y^2)=D(Y)+(EY)^2=1,$$
$$E(Y^4)=\int_{-\infty}^{+\infty} y^4\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\mathrm{d}y=3,$$

从而

$$\begin{aligned}E(X^2Y^2)&=E\left[Y^2E(X^2\mid Y)\right]\\&=E\left[Y^2(1-\rho^2)+\rho^2Y^4\right]\\&=(1-\rho^2)E(Y^2)+\rho^2E(Y^4)\\&=1-\rho^2+3\rho^2=1+2\rho^2.\quad\square\end{aligned}$$

例 1.1.21 设 X 为离散型随机变量，分布律为 $p_k=P\{X=x_k\},\ k=1,2,\cdots,n$. Y 为连续型随机变量，概率密度为 $f_Y(y)$，且 X，Y 相互独立，求 $Z=X+Y$ 的概率密度.

解 因

$$
\begin{aligned}
F_Z(z) &= P\{X+Y\le z\} \\
&= \int_{-\infty}^{+\infty} P\{X+Y\le z \mid X=x\}\mathrm{d}F_X(x) \quad （由(1.1.20)式） \\
&= \sum_{k=1}^{n} P\{X+Y\le z \mid X=x_k\} p_k \qquad （由X为离散型r.v.） \\
&= \sum_{k=1}^{n} P\{Y\le z-x_k \mid X=x_k\} p_k \\
&= \sum_{k=1}^{n} P\{Y\le z-x_k\} p_k \qquad （由X与Y独立）
\end{aligned}
$$

故

$$f_Z(z)=F_Z'(z)=\sum_{k=1}^{n} f_Y(z-x_k)p_k . \quad \square$$

与条件期望的定义类似，将 X 关于 $Y=y$ 的条件方差定义为

$$
\begin{aligned}
D(X\mid Y=y) &= \int_{-\infty}^{+\infty}\left[x-E(X|Y=y)\right]^2 \mathrm{d}F_{X|Y}(x\mid y) \\
&= \begin{cases} \sum_{k}\left[x_k-E(X|Y=y)\right]^2 P\{X=x_k\mid Y=y\}, & 当X,Y为离散型, \\ \int_{-\infty}^{+\infty}\left[x-E(X|Y=y)\right]^2 f_{X|Y}(x\mid y)\mathrm{d}x, & 当X,Y为连续型. \end{cases}
\end{aligned}
$$

一般情况下，$D(X|Y=y)$的计算结果是 y 的函数．以 Y 替换 y，称这一新的随机变量为 X 关于 Y 的**条件方差**，记作 $D(X\mid Y)$，即

$$D(X\mid Y)\triangleq D(X\mid Y=y)\big|_{y=Y} .$$

对于条件方差 $D(X\mid Y)$，下面的这条性质是比较重要的，

$$D(X)=E\left[D(X\mid Y)\right]+D\left[E(X\mid Y)\right]. \qquad (1.1.21)$$

称上式为**全方差公式**.

延伸阅读

在近代概率论中，条件期望是以公理的方式来定义的．给定的概率空间(Ω，$\mathscr{F}$，P)，X 是其上的随机变量，且 $E|X|<+\infty$，设 $\mathscr{G}$ 是 $\mathscr{F}$ 的子 σ-代数，即，$\mathscr{G}$ 是 σ-代数且 $\mathscr{G}\subset\mathscr{F}$，称随机变量 $E(X\mid\mathscr{G})$ 是 X 关于 $\mathscr{G}$ 的条件期望，如果它满足以下两个条件:

(1) $E(X\mid\mathscr{G})$为 $\mathscr{G}$ 可测;

(2) 对任何 $A\in\mathscr{G}$，有

$$\int_A E(X\mid\mathscr{G})\mathrm{d}P=\int_A X\,\mathrm{d}P.$$

如此定义的条件期望是否会落空呢？我们来分析一下．对任何 $A\in\mathscr{G}$，令

$$\mu(A)=\int_A X\,\mathrm{d}P.$$

可以验证，μ 是定义在 $\mathscr{G}$ 上的有限符号测度(符号测度又称为广义测度，它与测度的区别就在于值域范围，前者是$[-\infty,+\infty]$，后者是$[0,+\infty]$)．μ 与限制在 $\mathscr{G}$ 上的测度 P 有如下关系:

$\mu \ll P$，即 μ 关于 P 是绝对连续的．根据 Radon-Nikodym 定理，此时，存在 Radon-Nikodym 导数 $\dfrac{d\mu}{dP}$，该导数就满足上面的那两个条件．由此可知，在所述条件下 X 关于 $\mathscr{G}$ 的条件期望 $E(X\,|\,\mathscr{G})$ 是一定存在的．

对 $A\in\mathscr{F}$，取 $X=1_A$，记 $P(A|\mathscr{G})\triangleq E(1_A\,|\,\mathscr{G})$，称 $P(A|\mathscr{G})$ 为 A 关于 $\mathscr{G}$ 的条件概率．

1.1.8　概率分布的几种变换

1. 概率母函数

设 X 为取非负整数值的随机变量，分布律为 $p_k=P\{X=k\},\ k=0,1,2,\cdots$，则 X 的概率母函数（**probability generating function**）定义为

$$G(z)\triangleq\sum_{k=0}^{\infty}p_k z^k,\quad -1\le z\le 1.$$

X 的概率母函数也称作对分布律 $\{p_k,\ k=0,1,2,\cdots\}$ 的 **Z 变换**或**几何变换**．它是由 Laplace 引入到概率论中并最先被系统应用的一种变换方法．

利用幂级数的性质不难证明：概率母函数在[−1,1]上一定存在，且与分布律一一对应．

由 X 的概率母函数可以确定其分布律，

$$p_0=G(0),$$
$$p_k=\frac{G^{(k)}(0)}{k!},\ k=1,2,\cdots.$$

这便是概率母函数名称的来由．

当 X 的期望与方差存在时，因

$$G'(z)\big|_{z=1}=\sum_{k=1}^{\infty}kp_k=EX,$$
$$G''(z)\big|_{z=1}=\sum_{k=2}^{\infty}k(k-1)p_k=E\big(X(X-1)\big)=E(X^2)-EX,$$

故

$$EX=G'(1),$$
$$DX=E(X^2)-(EX)^2=G''(1)+G'(1)-[G'(1)]^2.$$

例 1.1.22　设 $X\sim P(\lambda)$，则 X 的概率母函数为

$$G(z)=\sum_{k=0}^{\infty}\frac{1}{k!}\lambda^k e^{-\lambda}z^k=e^{\lambda(z-1)}.$$

因 $G'(z)=\lambda e^{\lambda(z-1)}$，$G''(z)=\lambda^2 e^{\lambda(z-1)}$，故 $EX=G'(1)=\lambda$，$DX=G''(1)+G'(1)-[G'(1)]^2=\lambda$．

设 $X_1,\cdots,X_n$ 是 n 个相互独立、取非负整数值的随机变量，记 G_k 为 X_k（$1\le k\le n$）的概率母函数，若 $Y=X_1+\cdots+X_n$，G_Y 为 Y 的概率母函数，则

$$G_Y(z)=G_1(z)G_2(z)\cdots G_n(z).$$

用归纳法及独立变量之和的卷积公式可证明该结论，留给读者作练习.

例 1.1.23　设 $X_1 \sim P(\lambda_1)$，$X_2 \sim P(\lambda_2)$，且 X_1 与 X_2 独立，$Y = X_1+ X_2$，G_Y 为 Y 的概率母函数，由上述结论和例 1.1.22，得

$$G_Y(z) = G_1(z)G_2(z) = e^{\lambda_1(z-1)}e^{\lambda_2(z-1)} = e^{(\lambda_1+\lambda_2)(z-1)}.$$

再根据概率母函数与分布律是一一对应的，故 Y 仍服从 Poisson 分布，参数为 $\lambda_1+\lambda_2$，即 $Y \sim P(\lambda_1+\lambda_2)$.

例 1.1.24　将 4 颗匀称骰子投掷一次，求点数之和为 18 的概率.

解　令 X_k 为第 k 颗骰子掷出的点数，则 X_1, X_2, X_3, X_4 相互独立，$Y = X_1+ X_2 + X_3+ X_4$ 表示 4 颗骰子的点数和. 因 X_k 的概率母函数为 $\frac{1}{6}(z+z^2+\cdots+z^6)$，故 Y 的概率母函数为

$$\begin{aligned}G_Y(z) &= \frac{1}{6^4}(z+z^2+\cdots+z^6)^4 = \frac{z^4}{6^4}(1-z^6)^4(1-z)^{-4} \\ &= \frac{z^4}{6^4}(1-4z^6+6z^{12}-\cdots)\sum_{k=0}^{\infty}\binom{-4}{k}(-z)^k.\end{aligned}$$

由概率母函数的定义知，所求概率 $P\{Y=18\}$正是其概率母函数展开式中 z^{18} 前的系数，于是

$$P\{Y=18\} = \frac{1}{6^4}\left[\binom{-4}{14}(-1)^{14} - 4\binom{-4}{8}(-1)^8 + 6\binom{-4}{2}(-1)^2\right] = \frac{80}{6^4}.\quad \square$$

2．矩母函数

随机变量 X 的矩母函数（**moment generating function**）定义为

$$\psi(t) \triangleq E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx}\mathrm{d}F_X(x).$$

X 的矩母函数也称作对分布函数 F_X 的**指数变换**.

由 X 的矩母函数可以确定其各阶矩. 若 $E(X^k)$ $(k=1,2,\cdots)$存在，则

$$E(X^k) = \psi^{(k)}(0),\ k = 1,2,\cdots.$$

矩母函数因此得名.

注　矩母函数 ψ 的变动范围是由分布函数决定的. 对有些分布函数，不存在矩母函数.

可以证明：若 X 的矩母函数存在，则 X 的矩母函数与分布函数一一对应.

与概率母函数类似，对矩母函数也成立如下命题：设 $X_1, \cdots, X_n$ 是 n 个相互独立的随机变量，记 ψ_k 为 $X_k\,(1\le k\le n)$的矩母函数，若 $Y=X_1+\cdots+X_n$，ψ_Y 为 Y 的矩母函数，则

$$\psi_Y(t) = \psi_1(t)\psi_2(t)\cdots\psi_n(t).$$

例 1.1.25　设 $X \sim \mathrm{Exp}(\lambda)$，即 $f_X(x) = \lambda e^{-\lambda x}$ $(x>0)$，则 X 的矩母函数为

$$\psi(t) = E(e^{tX}) = \int_0^{+\infty} e^{tx}\lambda e^{-\lambda x}\mathrm{d}x = \frac{\lambda}{\lambda - t}\ \ (-\infty < t < \lambda).$$

因

$$\psi'(t)=\frac{\lambda}{(\lambda-t)^2},$$

$$\psi''(t)=\frac{2\lambda}{(\lambda-t)^3},$$

故

$$EX=\psi'(0)=\frac{1}{\lambda},$$

$$DX=E(X^2)-(EX)^2=\psi''(0)-\left(\psi'(0)\right)^2=\frac{1}{\lambda^2}.$$

3. 特征函数

随机变量 X 的特征函数（**characteristic function**）定义为

$$\varphi_X(t)\triangleq E(e^{itX})=\int_{-\infty}^{+\infty}e^{itx}\mathrm{d}F_X(x)$$
$$=\begin{cases}\int_{-\infty}^{+\infty}e^{itx}f_X(x)\mathrm{d}x, & \text{当}X\text{为连续型},\\ \sum_k e^{itx_k}P\{X=x_k\}, & \text{当}X\text{为离散型}.\end{cases}\tag{1.1.22}$$

注 (1)因为$|e^{itX}|\le 1$，所以随机变量的特征函数总是存在的.

(2) 由(1.1.22)式可知，连续型随机变量 X 的特征函数是其概率密度 f_X 的 Fourier 变换.

下面，介绍特征函数具有的一些重要性质.

(1) $\varphi(0)=1,\ |\varphi(t)|\le\varphi(0),\ \varphi(-t)=\overline{\varphi(t)}$, 并且 $\varphi(t)$ 在$(-\infty,+\infty)$上一致连续.

(2) $\varphi(t)$是非负定的，即对任给 n 个实数 t_k 及复数 z_k（$1\le k\le n$），有

$$\sum_{k=1}^{n}\sum_{l=1}^{n}\varphi(t_k-t_l)z_k\overline{z}_l\ge 0.$$

(3) 若 $X_1, \cdots, X_n$ 相互独立，则 $\varphi_{X_1+\cdots+X_n}(t)=\varphi_{X_1}(t)\varphi_{X_2}(t)\cdots\varphi_{X_n}(t)$.

(4) 若 X 的 n 阶矩 $E(X^n)$存在，则

$$E(X^k)=\frac{\varphi_X^{(k)}(0)}{i^k}\quad(k\le n),$$

其中$i=\sqrt{-1}$.

(5) 随机变量的分布函数与特征函数是一一对应的.

上述性质的证明请参阅文献[6]，[8].

例 1.1.26 设 $X\sim\chi^2(n)$，其概率密度为

$$f(x;n)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

则 X 的特征函数为

$$\begin{aligned}\varphi_X(t)&=\int_0^{+\infty}e^{itx}\frac{1}{2^{n/2}\Gamma(n/2)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}\mathrm{d}x\\&=\frac{1}{\Gamma(n/2)}\int_0^{+\infty}u^{\frac{n}{2}-1}e^{-(1-2it)u}\mathrm{d}u\quad(\text{令}\,u=x/2)\\&=\frac{1}{\Gamma(n/2)}\cdot\frac{\Gamma(n/2)}{s^{n/2}}\quad(\text{查 Laplace 变换表，其中}\,s=1-2it)\\&=(1-2it)^{-n/2}.\end{aligned}$$

由

$$\varphi_X'(t)=ni(1-2it)^{-\frac{n}{2}-1},$$
$$\varphi_X''(t)=-n(n+2)(1-2it)^{-\frac{n}{2}-2},$$

得

$$\begin{aligned}EX&=\frac{\varphi_X'(0)}{i}=n,\\DX&=E(X^2)-(EX)^2\\&=\frac{\varphi_X''(0)}{i^2}-\left(\frac{\varphi_X'(0)}{i}\right)^2=2n.\end{aligned}$$

若 $X\sim\chi^2(n_1)$，$Y\sim\chi^2(n_2)$，且 X 与 Y 独立，则由特征函数的性质(3)，得

$$\varphi_{X+Y}(t)=\varphi_X(t)\varphi_Y(t)=(1-2it)^{-\frac{n_1+n_2}{2}}.$$

因特征函数与分布函数是一一对应的，故 $X+Y\sim\chi^2(n_1+n_2)$．用独立变量之和的卷积公式也能得出同样的结论，但不如上述方法简便.

设 n 维随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)$的分布函数为 $F(x_1,\cdots,x_n)$，则 $\boldsymbol{X}$ 的 n 元特征函数定义为

$$\varphi_{\boldsymbol{X}}(t_1,\cdots,t_n)\triangleq Ee^{i(t_1X_1+\cdots+t_nX_n)}=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}e^{i(t_1x_1+\cdots+t_nx_n)}\mathrm{d}F(x_1,\cdots,x_n)$$

记 $\boldsymbol{t}\triangleq(t_1,\cdots,t_n)^{\mathrm{T}}$，$\boldsymbol{x}\triangleq(x_1,\cdots,x_n)^{\mathrm{T}}$，$\boldsymbol{X}\triangleq(X_1,\cdots,X_n)^{\mathrm{T}}$，则 n 元特征函数可改记为

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t})=E(e^{i\boldsymbol{X}^{\mathrm{T}}\boldsymbol{t}})=\int_{\boldsymbol{R}^n}e^{i\boldsymbol{x}^{\mathrm{T}}\boldsymbol{t}}\mathrm{d}F(\boldsymbol{x}).$$

定理 1.1.27 若 n 维随机向量 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}\sim N(\boldsymbol{a},\ \boldsymbol{B})$，即 $\boldsymbol{X}$ 的概率密度为

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}|\boldsymbol{B}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\},\ \boldsymbol{x}\triangleq(x_1,\cdots,x_n)^{\mathrm{T}}\in\boldsymbol{R}^n,$$

则 $\boldsymbol{X}$ 的 n 元特征函数是

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t})=\exp\left\{i\boldsymbol{a}^{\mathrm{T}}\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{t}\right\},\quad\boldsymbol{t}\triangleq(t_1,\cdots,t_n)^{\mathrm{T}}\in\boldsymbol{R}^n.\tag{1.1.23}$$

定理的证明参见文献[6].

多元特征函数也有与一元特征函数相似的一些性质，这里就不全部列举而仅介绍其中的一条性质.

若$E(X_1^{m_1}X_2^{m_2}\cdots X_n^{m_n})$存在，则

$$E(X_1^{m_1}X_2^{m_2}\cdots X_n^{m_n})=i^{-(m_1+m_2+\cdots+m_n)}\cdot\frac{\partial^{m_1+m_2+\cdots+m_n}\varphi_{\boldsymbol{X}}(0,0,\cdots,0)}{\partial t_1^{m_1}\partial t_2^{m_2}\cdots\partial t_n^{m_n}}.$$

例 1.1.28 设随机向量 $\boldsymbol{X}=(X_1,X_2,X_3,X_4)^{\mathrm{T}}\sim N(\boldsymbol{0},\ \boldsymbol{B})$，求 $E(X_1X_2X_3X_4)$.

解 $\boldsymbol{X}$ 的特征函数是 $\varphi_{\boldsymbol{X}}(\boldsymbol{t})=\exp\left\{-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{t}\right\}$，其中 $\boldsymbol{B}$ 是 $\boldsymbol{X}$ 的协方差阵（为对称正定阵），

二次型 $\boldsymbol{t}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{t}=\sum_{j=1}^{4}\sum_{i=1}^{4}b_{ij}t_it_j$，其中 $b_{ij}=b_{ji},i,j=1,\cdots,4$. 由链导法，得

$$\frac{\partial\varphi_{\boldsymbol{X}}}{\partial t_1}=\varphi_{\boldsymbol{X}}(\boldsymbol{t})\left[-\sum_{j=1}^{4}b_{1j}t_j\right],$$

$$\frac{\partial^2\varphi_{\boldsymbol{X}}}{\partial t_1\partial t_2}=\varphi_{\boldsymbol{X}}(\boldsymbol{t})\left[\sum_{j=1}^{4}b_{1j}t_j\right]\left[\sum_{j=1}^{4}b_{2j}t_j\right]+\varphi_{\boldsymbol{X}}(\boldsymbol{t})(-b_{12}),$$

$$\begin{aligned}\frac{\partial^3\varphi_{\boldsymbol{X}}}{\partial t_1\partial t_2\partial t_3}&=\varphi_{\boldsymbol{X}}(\boldsymbol{t})\left[\sum_{j=1}^{4}b_{1j}t_j\right]\left[\sum_{j=1}^{4}b_{2j}t_j\right]\left[-\sum_{j=1}^{4}b_{3j}t_j\right]\\&\quad+\varphi_{\boldsymbol{X}}(\boldsymbol{t})\left[b_{13}\sum_{j=1}^{4}b_{2j}t_j+b_{23}\sum_{j=1}^{4}b_{1j}t_j\right]+\varphi_{\boldsymbol{X}}(\boldsymbol{t})\left[\sum_{j=1}^{4}b_{3j}t_j\right]b_{12},\end{aligned}$$

$$\begin{aligned}\frac{\partial^4\varphi_{\boldsymbol{X}}}{\partial t_1\partial t_2\partial t_3\partial t_4}&=\varphi_{\boldsymbol{X}}(\boldsymbol{t})\left[b_{13}b_{24}+b_{23}b_{14}\right]+\varphi_{\boldsymbol{X}}(\boldsymbol{t})b_{12}b_{34}\\&\quad+\left\{\text{其余的加项，其中每一项都含因子}\sum_{j=1}^{4}b_{ij}t_j\ (1\le i\le 4)\right\}.\end{aligned}$$

因而

$$E(X_1X_2X_3X_4)=i^{-4}\frac{\partial^4\varphi_{\boldsymbol{X}}(0,0,0,0)}{\partial t_1\partial t_2\partial t_3\partial t_4}=b_{12}b_{34}+b_{13}b_{24}+b_{14}b_{23}.$$

由于 $\boldsymbol{X}$ 的均值向量为零向量，从而 $b_{ij}=Cov(X_i,X_j)=E(X_iX_j),\ i,j=1,2,3,4$，故得

$$E(X_1X_2X_3X_4)=E(X_1X_2)E(X_3X_4)+E(X_1X_3)E(X_2X_4)+E(X_1X_4)E(X_2X_3).\ \square$$

1.1.9 随机变量的四种收敛性

以下设随机变量列$\{X_n,n\ge 1\}$及随机变量 X 都定义在同一概率空间上.

1. 依概率收敛（convergence in probability）

若对任给 $\varepsilon>0$，都有

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

则称$\{X_n,n\ge 1\}$**依概率收敛**到 X，记作 $(p)\lim_{n\to\infty}X_n=X$ 或 $X_n\xrightarrow{P}X$.

2．　依分布收敛（convergence in distribution）

设 F_n 和 F 分别为 X_n 和 X 的分布函数，记 $C(F)$ 为 F 的连续点的集合．若

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in C(F),$$

则称 $\{X_n, n \geq 1\}$**依分布收敛**到 X，记作 $(d)\lim_{n\to\infty} X_n = X$ 或 $X_n \xrightarrow{d} X$ ．

注　对依分布收敛而言，由于仅涉及 X_n 与 X 的分布函数的极限，因而随机变量序列 $\{X_n\}$ 及随机变量 X 并不需要定义在同一概率空间．

3．　*r* 阶平均收敛（convergence in moment of order *r*）

对 $r > 0$，若有

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

则称 $\{X_n, n \geq 1\}$***r* 阶平均收敛**到 X，记作 $(L^r)\lim_{n\to\infty} X_n = X$ 或 $X_n \xrightarrow{L^r} X$ ．

在 r 阶平均收敛中，最常见的是 r=2 的情形，此时，称为**均方收敛**（convergence in mean square），记作 $(m.s.)\lim_{n\to\infty} X_n = X$ 或 $X_n \xrightarrow{m.s./L^2} X$ ．

4．　概率 1 收敛

若

$$P\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1,$$

则称 $\{X_n, n \geq 1\}$**概率 1 收敛**到 X．概率 1 收敛到 X 也常称作**几乎必然收敛**（convergent almost surely to X）到 X，记作 $(a.s.)\lim_{n\to\infty} X_n = X$ 或 $X_n \xrightarrow{a.s.} X$ ．

5．　各种收敛性之间的关系

前面介绍了随机变量序列的 4 种收敛性，它们之间存在着一定的强弱关系．

定理 1.1.29　(1) 若 $X_n \xrightarrow{a.s.} X$，则 $X_n \xrightarrow{P} X$ ．

(2) 若 $X_n \xrightarrow{L^r} X$，则 $X_n \xrightarrow{P} X$ ．

(3) 若 $X_n \xrightarrow{P} X$，则 $X_n \xrightarrow{d} X$ ．

证明　(1) 因为 $\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = \bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left(\omega : |X_n(\omega) - X(\omega)| < \frac{1}{m}\right)$，所以

$$\begin{aligned} X_n \xrightarrow{a.s.} X &\Leftrightarrow P\left\{\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left(\omega :\ |X_n(\omega) - X(\omega)| < \frac{1}{m}\right)\right\} = 1, \\ &\Leftrightarrow P\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(\omega :\ |X_n(\omega) - X(\omega)| \geq \frac{1}{m}\right)\right\} = 0, \\ &\Leftrightarrow \forall \varepsilon > 0,\ P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(\omega :\ |X_n(\omega) - X(\omega)| \geq \varepsilon\right)\right\} = 0. \end{aligned}$$

由

$$\{\omega:|X_k(\omega)-X(\omega)|\geq\varepsilon\}\subset\bigcup_{n=k}^{\infty}\left(\omega:|X_k(\omega)-X(\omega)|\geq\varepsilon\right)$$

及

$$\lim_{k\to\infty}P\left\{\bigcup_{n=k}^{\infty}\left(\omega:|X_k(\omega)-X(\omega)|\geq\varepsilon\right)\right\}=P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(\omega:|X_n(\omega)-X(\omega)|\geq\varepsilon\right)\right\}=0,$$

得

$$\lim_{k\to\infty}P\{\omega:|X_k(\omega)-X(\omega)|\geq\varepsilon\}=0.$$

(2) 由 Markov 不等式

$$P\{|X|\geq\varepsilon\}\leq\frac{E(|X|^r)}{\varepsilon^r},$$

得

$$P\{|X_n-X|\geq\varepsilon\}\leq\frac{E(|X_n-X|^r)}{\varepsilon^r}\to 0\ \ (n\to\infty),$$

即 $\lim\limits_{n\to\infty}P\{|X_n-X|\geq\varepsilon\}=0$.

(3) 设 $s<x$，对任意 $n\geq 1$，由

$$\{X\leq s\}=\{X_n\leq x,X\leq s\}\cup\{X_n>x,X\leq s\}\subset\{X_n\leq x\}\cup\{X_n>x,X\leq s\},$$

得

$$F(s)\leq F_n(x)+P\{X_n>x,X\leq s\}.$$

而

$$P\{X_n>x,X\leq s\}\leq P\{|X_n-X|\geq x-s\}\to 0\ \ (\because X_n\xrightarrow{P}X),$$

故

$$F(s)\leq\varliminf_{n\to\infty}F_n(x).$$

又设 $t>x$，用与上面类似的证法，可得

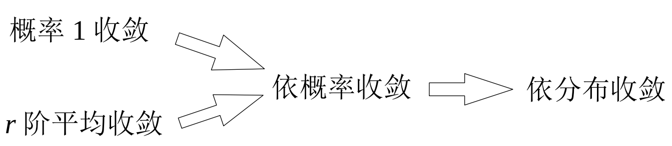
$$\varlimsup_{n\to\infty}F_n(x)\leq F(t).$$

因而，对任意 $s<x<t$，有

$$F(s)\leq\varliminf_{n\to\infty}F_n(x)\leq\varlimsup_{n\to\infty}F_n(x)\leq F(t)$$

当 $x\in C(F)$ 时，在上式中令 $s\to x,\ t\to x$，就有 $\lim\limits_{n\to\infty}F_n(x)=F(x)$，即 $X_n\xrightarrow{d}X$． □

随机变量序列的这四种收敛性可用下面的图来表示：



需要注意的是，上述关系之逆命题一般不成立．另外，概率 1 收敛未必蕴含 r 阶平均收敛；反之，r 阶平均收敛也未必蕴含概率 1 收敛．

6.　收敛概念的应用

在本科概率论中介绍过一些大数定律，实际上，大数定律就用到了依概率收敛这一概念．

例如，在 Bernoulli 试验中，记 $P(A)=p$，S_n 表示前 n 次试验中 A 出现的次数，则 $\frac{S_n}{n}\xrightarrow{P}p$．这就是 Bernoulli 大数定律的内容．又如，若$\{X_n, n\ge 1\}$为独立同分布的随机变量列，$EX_1=\mu<\infty$，则$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{P}\mu$．这是 Khintchine 大数定律告诉我们的．

下面，给出大数定律的一般形式的描述．

设$\{X_n, n\ge 1\}$为随机变量序列，$S_n=\sum_{k=1}^{n}X_k$，$n=1,2,\cdots$，若存在常数列$\{a_n, n\ge 1\}$，使得

(1)　$\frac{S_n}{n}-a_n\xrightarrow{P}0$，则称$\{X_n, n\ge 1\}$遵从大数定律．

(2)　$\frac{S_n}{n}-a_n\xrightarrow{a.s.}0$，则称$\{X_n, n\ge 1\}$遵从强大数定律．

由定理 1.1.29 知，当$\{X_n, n\ge 1\}$遵从强大数定律时必遵从大数定律，所以大数定律又叫做**弱大数定律**（**weak law of large number**）．

这里，我们不加证明地介绍两个强大数定律．

定理 1.1.30 (Kolmogorov 强大数定律)　设$\{X_n, n\ge 1\}$为相互独立的随机变量列，且$\sum_{n=1}^{\infty}\frac{DX_n}{n^2}<\infty$，则

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(X_k-EX_k)=0\right\}=1.$$

定理 1.1.31　(Borel 强大数定律)　在 Bernoulli 试验中，记 $P(A)=p$，S_n 表示前 n 次试验中 A 出现的次数，则

$$P\left\{\lim_{n\to\infty}\frac{S_n}{n}=p\right\}=1.$$

1.2　随机过程基本概念

1.2.1 随机过程概念和记号

1. 实值随机过程

在本科概率论中，主要研究一个或有限多个随机变量的统计规律．随着科学技术的发展，人们还需要观察和研究随机变量随时间的变化过程，也就是随时间演变的随机现象，这便涉及到研究无穷多个随机变量的问题．随机过程理论正是在这种要求下，发端于 20 世纪初期，并伴随物理学、生物学、通讯与控制工程学、管理学等其它学科的需要及促进而逐步发展起来．

下面是需要从整体上来研究无穷多个随机变量的例子．

例 1.2.1 考虑某种群，它由能产生同类后代的个体构成，初始种群的个体总数用 X_0 表示，称为第 0 代的总数．第 0 代的全体后代构成第 1 代，其总数用 X_1 表示．一般地，以 X_n 表示第 n 代的总数．要分析种群生长的变化规律，就需要研究无穷多个随机变量$\{X_n, n=0,1,2,\cdots\}$．

例 1.2.2 研究车站检票口乘客到达的数量．用 $X(t)$表示在时间间隔 $(0,t]$内到达检票口的人数，$X(t)$为取非负整数值的随机变量．要分析客流随时间的演变规律，就需要研究随机变量族$\{X(t), t\geqslant 0\}$．

例 1.2.3 为研究极地气候随时间的变化规律，以 $X(t)$表示我国南极科考站建立后的第 t 天所记录的当地气温值．可以认为 $X(t)$是取负数值的随机变量．对每一天做记录，就得到随机变量族$\{X(t), t=1,2,\cdots\}$．

例 1.2.4 英国植物学家布朗(R. Brown)发现，一个完全浸没于液体中或气体里的花粉微粒作不规则的运动（后人称这种运动为 Brown 运动）．这种运动的起因是花粉不断地受到周围介质分子的碰撞，这些微小碰撞力的总和使花粉作随机运动．若以 $X(t)$，$Y(t)$，$Z(t)$表示 t 时刻花粉微粒的空间坐标，要研究花粉微粒的随机运动规律，就需考虑三维随机变量族$\{(X(t), Y(t), Z(t)), t\geqslant 0\}$．为了简单起见，也可以只考察其横坐标随时间的变化规律，即研究随机变量族$\{X(t), t\geqslant 0\}$．

上述例子都涉及到依赖于参数而变化的随机变量（一族随机变量）．在这种随机变量族中，随机变量之间往往不独立，因而，仅对每一个随机变量进行研究是不够的，需要从整体上来研究这一族随机变量．这就引出如下的随机过程概念．

定义 1.2.5 设给定概率空间$(\Omega, \mathscr{F}, P)$及参数集 T．若对每个 $t\in T$，都有定义在$(\Omega, \mathscr{F}, P)$上的随机变量 $X(t,\omega)$与之对应，则称$\{X(t,\omega), t\in T\}$为（**实值**）**随机过程**（简称过程）．

称随机变量族$\{X(t,\omega), t\in T\}$为过程是因为参数 t 大多代表时间．但也不尽然，在有些场合，t 也可以代表别的量．比如，序号、高度、距离等．有时，参数 t 也可以是向量．如 t 是 $\boldsymbol{R}^3$ 中的点，考察 t 点处的气压 $X(t)$．当参数 t 为向量时，称对应的随机过程为**随机场**．本书仅限于讨论 $T\subset \boldsymbol{R}^1$，并称 $t\in T$ 为时间．

当参数集 T 为可列集时，称$\{X(t,\omega), t\in T\}$为**离散参数随机过程**（**随机序列，时间序列**）．当参数集 T 为实数空间 $\boldsymbol{R}^1$ 中的有限或无穷区间时，称$\{X(t,\omega), t\in T\}$为**连续参数随机过程**．

从本质上讲，随机过程 $X(\cdot,\cdot)$是关于 $t\in T, \omega\in\Omega$ 的二元映射，即，$X(\cdot,\cdot)\colon T\times\Omega\mapsto R$．

当固定 $t\in T$ 时，$X(t,\cdot)$是一个随机变量．因而，与参数集 T 的所有元素相对应，$\{X(t,\cdot), t\in T\}$代表了一族随机变量．

当进行一次随机试验，出现样本点 $\omega\in\Omega$ 时，相应地得到了一个 t 的函数 $X(\cdot,\omega)$，称之为随机过程的**样本函数**（**路径，轨道，实现**），因此，与试验的所有可能结果相对应，$\{X(\cdot,\omega), \omega\in\Omega\}$代表了一族样本函数．

与 $t\in T, \omega\in\Omega$ 对应的 $X(t,\omega)$是一个实数，称为随机过程的一个**状态**．集合

$$S\triangleq\left\{X(t,\omega), t\in T, \omega\in\Omega\right\}$$

称为随机过程的**状态空间**（**值域，相空间**）．

以后将过程$\{X(t,\omega), t\in T\}$简记为$\{X(t), t\in T\}$并用记号 $\boldsymbol{X}$ 来表示，即 $\boldsymbol{X}\triangleq\{X(t), t\in T\}$．

2. 复值随机过程

在信息与控制工程领域，对随机信号进行时频分析时会涉及复值的情形．为此，需要将随机过程概念由实值推广到复值．

若 X, Y 是定义在概率空间$(\Omega, \mathscr{F}, P)$上两个随机变量，记 $Z\triangleq X+iY$，其中 $i=\sqrt{-1}$，则称

Z 为复值随机变量.

复值随机变量 Z 的概率分布是由随机向量(X, Y)的概率分布所确定，其均值、方差定义为

$$EZ \triangleq EX + iEY,$$
$$DZ \triangleq E|Z - EZ|^2.$$

需注意的是，复值随机变量 Z 是取值为复数的随机变量，一般情况下，其均值 EZ 仍为复值，但方差 $DZ \geq 0$ 为非负实数.

对定义在相同概率空间$(\Omega, \mathscr{F}, P)$上的两个复值随机变量 Z_1, Z_2，它们的协方差定义为

$$Cov(Z_1, Z_2) \triangleq E\left[(Z_1 - EZ_1)\overline{(Z_2 - EZ_2)}\right].$$

与（实值）随机变量情形不同的是 $Cov(Z_2, Z_1) \neq Cov(Z_1, Z_2)$，而应修正为

$$Cov(Z_2, Z_1) = \overline{Cov(Z_1, Z_2)}.$$

当 $Cov(Z_1, Z_2) = 0$ 时，亦称 Z_1 与 Z_2 不相关.

下面来介绍复值随机过程概念.

定义 1.2.6 对给定的参数集 T，若$\{X(t), t\in T\}$、$\{Y(t), t\in T\}$是定义在相同的概率空间$(\Omega, \mathscr{F}, P)$上的两个随机过程，则称$\{X(t) + iY(t), t\in T\}$为复值随机过程，记作

$$\boldsymbol{Z} \triangleq \{X(t) + iY(t),\ t \in T\}.$$

实际上，复值随机过程 $\boldsymbol{Z} = \{X(t) + iY(t),\ t \in T\}$ 代表了定义在概率空间$(\Omega, \mathscr{F}, P)$上的一族复值随机变量.

对复值随机过程的研究方法大多与实值随机过程相仿，因此，在以后的讨论中若无特别申明，所说的随机过程均为实值随机过程.

延伸阅读

随机过程概念还可以推广到更一般的情形.

设$(\Omega, \mathscr{F}, P)$为概率空间，$(E, \mathscr{E})$为任一可测空间，X 为从 Ω 到 E 的映射，若对任意 $B \in \mathscr{E}$，都有 $\{\omega : X(\omega) \in B\} \in \mathscr{F}$，则称 X 为取值于 E 的**随机元**.

如果取$(E, \mathscr{E}) = (\boldsymbol{R}^d, \mathscr{B}^d)$，其中 $\mathscr{B}^d$ 为 d 维 Borel 域，那么取值于 E 的随机元 X 就是 d 维随机向量. 特别，当 $d = 1$ 时，随机元 X 即是随机变量.

如果取 E 为复数空间，$\mathscr{E}$ 为复 Borel 域，那么取值于 E 的随机元 X 就是复值随机变量.

在将随机变量概念推广到一般情形下的随机元概念之后，就可以来讨论随机过程概念的推广问题.

给定概率空间$(\Omega, \mathscr{F}, P)$和参数集 T，设$(E, \mathscr{E})$为可测空间，若对每一 $t\in T$，都有取值于 E 的随机元 $X(t)$与之对应，则称$\{X(t), t\in T\}$为 **E 值随机过程**，E 称为**状态空间**或**相空间**.

1.2.2 随机过程的概率特性

1. 有限维分布函数族

掌握了分布函数 $F(x) = P\{X \leqslant x\}$就能全面地了解随机变量 X 的概率规律. 类似地，对于过程 $\boldsymbol{X}=\{X(t), t\in T\}$，因为它代表了彼此可能有关联的一族随机变量，为了描述它的统计特性，需要知道对于任意一个时刻 $t\in T$，$X(t)$的分布函数

$$F_t(x) \triangleq P\{X(t) \le x\}, \tag{1.2.1}$$

称 $F_t(x)$为过程 $\boldsymbol{X}$ 在 t 时刻的**一维分布函数**.

还需要知道对于任意两个不同时刻 t_1、$t_2 \in T$，$X(t_1)$，$X(t_2)$的联合分布函数

$$F_{t_1,t_2}(x_1,x_2) \triangleq P\{X(t_1) \le x_1, X(t_2) \le x_2\}, \tag{1.2.2}$$

称 $F_{t_1,t_2}(x_1,x_2)$ 为过程 $\boldsymbol{X}$ 在 t_1、t_2 时刻的**二维分布函数**.

知道了过程 $\boldsymbol{X}$ 的所有一、二维分布函数，对于研究一个过程来说仍远远不够．一般地，对任意有限个不同时刻 t_1，…，$t_n \in T$，过程 $\boldsymbol{X}$ 的 ***n* 维分布函数**定义为

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) \triangleq P\{X(t_1) \le x_1,\cdots,X(t_n) \le x_n\}$$

过程 $\boldsymbol{X}$ 的所有一维分布函数，所有二维分布函数，…，所有 n 维分布函数，等等，其全体

$$\{F_{t_1,\cdots,t_n}(x_1,\cdots,x_n),\ t_1,\cdots,t_n \in T,\ n=1,2,\cdots\} \tag{1.2.3}$$

称为过程 $\boldsymbol{X}$ 的**有限维分布函数族**.

掌握了随机过程的有限维分布函数族，就能知道 $\boldsymbol{X}$ 中任意有限个随机变量的联合分布，从而就能了解这些随机变量之间的相互依赖关系．从这个意义上讲，过程的有限维分布函数族完整地描述随机过程的概率规律.

不难看出，随机过程的有限维分布函数族具有下列性质：

(1) 对 $1,\cdots,n$ 的任意排列 $i_1,\ \cdots,i_n$，有 $F_{t_1,\cdots,t_n}(x_1,\cdots,x_n) = F_{t_{i_1},\cdots,t_{i_n}}(x_{i_1},\cdots,x_{i_n})$;

(2) 若 $m<n$，则 $F_{t_1,\cdots,t_m}(x_1,\cdots,x_m) = F_{t_1,\cdots,t_m,t_{m+1},\cdots,t_n}(x_1,\cdots,x_m,\infty,\cdots,\infty)$.

与定理 1.1.8 一样，我们不加证明地给出如下定理.

定理 1.2.7 （**Kolmogorov**） 若有参数集 T 及满足上述两条性质的有限维分布函数族(1.2.3)， 则一定存在概率空间$(\Omega,\ \mathscr{F},\ P)$及定义于其上的随机过程 $\boldsymbol{X} = \{X(t), t \in T\}$，使得 $\boldsymbol{X}$ 的有限维分布函数族与已给的有限维分布函数族重合.

Kolmogorov 定理表明，对随机过程的研究可转化为对有限维分布函数族的研究。

2. 随机过程的数字特征

虽然随机过程的有限维分布函数族完整地刻画了过程的概率特性，但在实际中，一方面，有时并不需要了解随机过程全部的概率特性，只需知道它的局部概率特征就够了．另一方面，要确定过程的有限维分布函数族通常是很困难的，有些情况下甚至是不可能的．因而，我们来考虑随机过程的局部统计特征——随机过程的数字特征.

设过程 $\boldsymbol{X} = \{X(t), t \in T\}$．$\boldsymbol{X}$ 的**均值函数**（**mean function**）定义为

$$m_{\boldsymbol{X}}(t) \triangleq E[X(t)] = \int_{-\infty}^{+\infty} x \mathrm{d}F_t(x) \quad (t \in T), \tag{1.2.4}$$

其中 $F_t(x)$为过程 $\boldsymbol{X}$ 在 t 时刻的一维分布函数.

$m_{\boldsymbol{X}}(t)$表示过程 $\boldsymbol{X}$ 的样本函数在 t 时刻的状态的理论均值（又称“集平均”）．几何上，它表示 $\boldsymbol{X}$ 在各个时刻的摆动中心.

$\boldsymbol{X}$ 的**方差函数**（**variance function**）定义为

$$D_{\boldsymbol{X}}(t) \triangleq E|X(t) - m_{\boldsymbol{X}}(t)|^2 \quad (t \in T). \tag{1.2.5}$$

方差函数也可记作 $\mathrm{Var}_{\boldsymbol{X}}(t)$ 或 $\sigma_{\boldsymbol{X}}^2(t)$.

$\sigma_{\boldsymbol{X}}(t) \triangleq \sqrt{D_{\boldsymbol{X}}(t)}$ 称为 $\boldsymbol{X}$ 的**标准差函数**（**standard deviation function**）. $D_{\boldsymbol{X}}(t)$ 和 $\sigma_{\boldsymbol{X}}(t)$都刻画过程的样本函数在 t 时刻对 $m_{\boldsymbol{X}}(t)$的偏离程度.

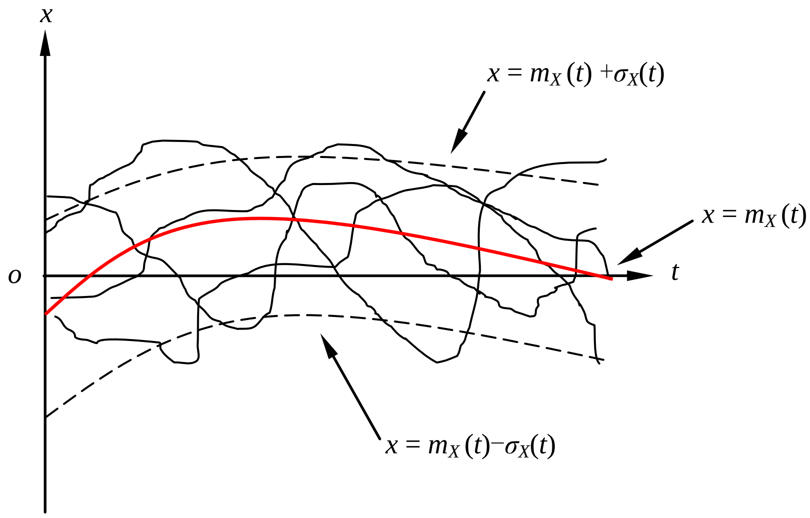


图 1.2.1

$\boldsymbol{X}$ 的**自相关函数**（**autocorrelation function**）定义为

$$R_{\boldsymbol{X}}(t_1,t_2) \triangleq E\left[X(t_1)\overline{X(t_2)}\right] \quad (t_1,t_2 \in T). \tag{1.2.6}$$

$\boldsymbol{X}$ 的**自协方差函数**（**auto-covariance function**）定义为

$$C_{\boldsymbol{X}}(t_1,t_2) \triangleq E\left[\left(X(t_1)-m_{\boldsymbol{X}}(t_1)\right)\overline{\left(X(t_2)-m_{\boldsymbol{X}}(t_2)\right)}\right] \quad (t_1,t_2 \in T). \tag{1.2.7}$$

注 (1) 在过程 $\boldsymbol{X}$ 的方差函数、自相关函数及自协方差函数的定义中出现了取模或取共轭的运算符号，这是为了让相应的定义也适用于复值随机过程的缘故．在以后的讨论中，如果在某些定义、定理以及证明中原本讨论的是实值随机过程却出现取模或取共轭的情形，那么其理由就是本款注释所指出——这些概念及结论可平行推广至复值随机过程．今后就不再一一注明．

(2) 工程上，有时会用到过程 $\boldsymbol{X}$ 的均方值函数，其定义为

$$\Psi_{\boldsymbol{X}}(t) \triangleq E\left[|X(t)|^2\right] \quad (t \in T).$$

(3) 因为

$$C_{\boldsymbol{X}}(t_1,t_2) = R_{\boldsymbol{X}}(t_1,t_2) - m_{\boldsymbol{X}}(t_1)\overline{m_{\boldsymbol{X}}(t_2)},$$
$$D_{\boldsymbol{X}}(t) = C_{\boldsymbol{X}}(t,t) = R_{\boldsymbol{X}}(t,t) - |m_{\boldsymbol{X}}(t)|^2,$$
$$\Psi_{\boldsymbol{X}}(t) = R_{\boldsymbol{X}}(t,t),$$

所以均值函数和自相关函数是随机过程的两个最基本的数字特征.

设 $\boldsymbol{X}=\{X(t), t\in T\}$、$\boldsymbol{Y}=\{Y(t), t\in T\}$为定义在同一概率空间$(\Omega, \mathscr{F}, P)$上且有相同参数集 T 的两个随机过程.

$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的**互相关函数**（**cross correlation function**）定义为

$$R_{\boldsymbol{XY}}(t_1,t_2) \triangleq E\left[X(t_1)\overline{Y(t_2)}\right] \quad (t_1,t_2 \in T). \tag{1.2.8}$$

$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的**互协方差函数（cross covariance function）**定义为

$$C_{XY}(t_1,t_2) \triangleq E\left[\left(X(t_1)-m_X(t_1)\right)\overline{\left(Y(t_2)-m_Y(t_2)\right)}\right] \quad (t_1,t_2 \in T). \qquad (1.2.9)$$

其中 $m_X(\cdot)$、$m_Y(\cdot)$分别是 $\boldsymbol{X}$ 、$\boldsymbol{Y}$ 的均值函数.

例 1.2.8 设过程 $\boldsymbol{X}=\{Vt+b, -\infty<t<+\infty\}$，其中 b 为常数，V 为随机变量且 $V\sim N(0,1)$，求 $\boldsymbol{X}$ 的均值函数和自相关函数.

解 对任意固定的 $t\in(-\infty, +\infty)$,

$$m_X(t) = E(Vt+b) = tEV+b = b.$$

对任意固定的 $t_1, t_2\in(-\infty, +\infty)$,

$$\begin{aligned} R_X(t_1, t_2) &= E[(Vt_1+b)(Vt_2+b)] \\ &= t_1t_2E(V^2) + b(t_1+t_2)EV + b^2 \\ &= t_1t_2 + b^2. \quad \square \end{aligned}$$

例 1.2.9 设过程 $\boldsymbol{X}=\{A\cos(\lambda t+\Phi), -\infty<t<+\infty\}$，其中，$\lambda>0$ 为常数，A 和 Φ 为相互独立的随机变量，且 $A\sim U(0,1)$，$\Phi\sim U(0, 2\pi)$．求 $\boldsymbol{X}$ 的均值函数和自相关函数.

解 对任意固定的 $t\in(-\infty, +\infty)$

$$\begin{aligned} m_X(t) &= E[A\cos(\lambda t+\Phi)] \\ &= E(A)E[\cos(\lambda t+\Phi)] \qquad \text{（由}A\text{与}\Phi\text{独立）} \\ &= \int_0^1 a\,\mathrm{d}a\int_0^{2\pi}\cos(\lambda t+\varphi)\frac{1}{2\pi}\mathrm{d}\varphi \\ &= 0. \end{aligned}$$

对任意固定的 t_1，$t_2\in(-\infty, +\infty)$

$$\begin{aligned} R_X(t_1,t_2) &= E\left[A^2\cos(\lambda t_1+\Phi)\cos(\lambda t_2+\Phi)\right] \\ &= E(A^2)E\left[\cos(\lambda t_1+\Phi)\cos(\lambda t_2+\Phi)\right] \quad \text{（由}A\text{与}\Phi\text{独立）} \\ &= E(A^2)E\left[\frac{1}{2}\left[\cos\left(\lambda(t_2-t_1)\right)+\cos\left(\lambda(t_1+t_2)+2\Phi\right)\right]\right] \\ &= \int_0^1 a^2\mathrm{d}a\int_0^{2\pi}\frac{1}{2}\left[\cos\left(\lambda(t_2-t_1)\right)+\cos\left(\lambda(t_1+t_2)+2\varphi\right)\right]\frac{1}{2\pi}\mathrm{d}\varphi \\ &= \frac{\cos\left[\lambda(t_2-t_1)\right]}{6}. \quad \square \end{aligned}$$

例 1.2.10 设 $\boldsymbol{X}=\{\alpha\cos(\lambda t+\Theta)$，$-\infty<t<+\infty\}$，$\boldsymbol{Y}=\{\beta\cos(\lambda t+\Theta-\varphi)$，$-\infty<t<+\infty\}$，其中 α，β，λ，φ 为常数，Θ 为随机变量且 $\Theta\sim U(0, 2\pi)$．求 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的互相关函数.

解 对任意固定的 t_1，$t_2\in(-\infty, +\infty)$

$$\begin{aligned} R_{XY}(t_1,t_2) &= E\left[\alpha\cos(\lambda t_1+\Theta)\beta\cos(\lambda t_2+\Theta-\varphi)\right] \\ &= \alpha\beta\int_0^{2\pi}\frac{1}{2}\left[\cos\left(\lambda(t_2-t_1)-\varphi\right)+\cos\left(\lambda(t_1+t_2)+2\theta-\varphi\right)\right]\frac{1}{2\pi}\mathrm{d}\theta \\ &= \frac{\alpha\beta}{2}\cos\left(\lambda(t_2-t_1)-\varphi\right). \quad \square \end{aligned}$$

3. 两个随机过程的不相关与相互独立

设 $\boldsymbol{X}=\{X(t), t\in T\}$、$\boldsymbol{Y}=\{Y(t), t\in T\}$为定义在同一概率空间$(\Omega, \mathscr{F}, P)$上两个随机过程．若对任意 t_1，$t_2\in T$，有

$$C_{XY}(t_1, t_2) = 0,$$

则称 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 是**不相关的**.

若对任意 n，$m\in\boldsymbol{N}$ 及任意 t_1，$\cdots$，t_n，t'_1，$\cdots$，$t'_m\in T$，$(X(t_1), \cdots, X(t_n))$与$(Y(t'_1), \cdots, Y(t'_m))$相互独立，则称 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 是**相互独立的**.

如果 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立，且它们的二阶混合矩存在，那么 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 是不相关的．反之，由 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的不相关一般不能得出 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立的结论来.

设过程 $\boldsymbol{X}$，$\boldsymbol{Y}$ 的均值函数与自相关函数为 $m_{\boldsymbol{X}}(\cdot)$，$m_{\boldsymbol{Y}}(\cdot)$与 $R_{\boldsymbol{X}}(\cdot\,;\,\cdot)$，$R_{\boldsymbol{Y}}(\cdot\,;\,\cdot)$．令

$$\boldsymbol{Z} = \{ X(t) + Y(t), t\in T \},$$

则过程 $\boldsymbol{Z}$ 的均值函数与自相关函数为

$$m_{\boldsymbol{Z}}(t) = E[X(t) + Y(t)] = E[X(t)] + E[Y(t)] = m_{\boldsymbol{X}}(t) + m_{\boldsymbol{Y}}(t),$$

$$\begin{aligned} R_{\boldsymbol{Z}}(t_1, t_2) &= E[(X(t_1)+Y(t_1))(X(t_2)+Y(t_2))] \\ &= R_{\boldsymbol{X}}(t_1, t_2) + R_{\boldsymbol{XY}}(t_1, t_2) + R_{\boldsymbol{YX}}(t_1, t_2) + R_{\boldsymbol{Y}}(t_1, t_2). \end{aligned}$$

特别，当 $m_{\boldsymbol{X}}(t)\equiv 0\equiv m_{\boldsymbol{Y}}(t)$且 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 不相关时，有

$$R_{\boldsymbol{Z}}(t_1, t_2) = R_{\boldsymbol{X}}(t_1, t_2) + R_{\boldsymbol{Y}}(t_1, t_2).$$

1.2.3 随机过程的基本类型

随机过程通常按两种方式来分类.

1. 按参数集与状态集分类

按照过程 $\boldsymbol{X}=\{ X(t), t\in T \}$的参数集 T 与状态空间 S 的可列与否，将过程分为 4 类，即

(1) 可列参数、可列状态的随机过程；

(2) 可列参数、不可列状态的随机过程；

(3) 不可列参数、可列状态的随机过程；

(4) 不可列参数、不可列状态的随机过程.

例 1.2.1~1.2.4 即分属上述 4 类过程.

这种分类方法将所有的随机过程明确地分为了 4 类，每一个随机过程属于且仅属于其中的一类．但这种分类法有些流于形式，实际意义并不大．在随机过程理论研究中倾向于用下面的方式来分类.

2. 按随机过程的性质特点分类

按照随机过程的性质特点，即，依据过程的统计特性（如分布族，数字特征）或依据过程内在的概率法则，对过程进行分类.

(1) 独立随机过程

设过程 $\boldsymbol{X}=\{ X(t), t\in T \}$，若对任意 $n>1$ 及 n 个不同时刻 t_1，$\cdots$，$t_n\in T$，随机变量 $X(t_1)$，$\cdots$，$X(t_n)$相互独立，则称 $\boldsymbol{X}$ 为**独立随机过程**.

独立随机过程 $\boldsymbol{X}=\{ X(t), t\in T \}$代表了一族相互独立的随机变量，因而它的有限维分布函数族$\left\{F_{t_1,t_2,\cdots,t_n},\ t_1,t_2,\cdots,t_n\in T,\ n\in\boldsymbol{N}\right\}$满足

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n)=F_{t_1}(x_1)F_{t_2}(x_2)\cdots F_{t_n}(x_n)\ \ (x_1,x_2,\cdots,x_n\in\boldsymbol{R}^1).$$

特别，当参数集 $T=\{1, 2, \cdots\}$时，过程 $\boldsymbol{X}=\{ X(t), t\in T \}$即为独立随机变量序列.

(2) 二阶矩过程

设过程 $\boldsymbol{X}=\{ X(t), t\in T \}$，若对任意 $t\in T$，$E|X(t)|^2<\infty$，则称 $\boldsymbol{X}$ 为**二阶矩过程**.

由 Cauchy-Schwarz 不等式可知，二阶矩过程的均值函数和自相关函数都是存在的.

二阶矩过程是比较宽泛的一类过程，自然科学和工程技术领域中所遇到的随机过程大多属于此范畴.

在二阶矩过程类中有一些比较重要的子类，例如，正态过程，正交增量过程等.

若 $\boldsymbol{X}=\{ X(t), t\in T \}$的任一有限维分布都是正态分布，亦即，对任意 $n\geq 1$ 和任意 n 个不同

的 $t_1, t_2, \cdots, t_n \in T$，$n$ 维随机向量$(X(t_1), X(t_2), \cdots, X(t_n))$的联合概率密度为

$$f_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=(2\pi)^{-\frac{n}{2}}|\boldsymbol{C}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{m})\right\}$$

其中

$$\begin{aligned}
&\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}},\\
&\boldsymbol{m}=\left(m_{\boldsymbol{X}}(t_1),\cdots,m_{\boldsymbol{X}}(t_n)\right)^{\mathrm{T}},\\
&\boldsymbol{C}=\begin{bmatrix} C_{\boldsymbol{X}}(t_1,t_1) & \cdots & C_{\boldsymbol{X}}(t_1,t_n)\\ C_{\boldsymbol{X}}(t_2,t_1) & \cdots & C_{\boldsymbol{X}}(t_2,t_n)\\ \vdots & & \vdots \\ C_{\boldsymbol{X}}(t_n,t_1) & \cdots & C_{\boldsymbol{X}}(t_n,t_n)\end{bmatrix},
\end{aligned}$$

$\boldsymbol{C}$ 为对称正定矩阵，则称 $\boldsymbol{X}$ 为**正态过程**（或**高斯过程**）.

正态过程 $\boldsymbol{X}$ 的有限维分布完全由它的均值函数 $m_{\boldsymbol{X}}(\cdot)$和自相关函数 $R_{\boldsymbol{X}}(\cdot,\cdot)$所确定．于是，人们不仅可以在二阶矩的范围内来研究它，还可以从分布的角度来讨论与过程有关的一些概率问题.

设 $\boldsymbol{X}=\{X(t), 0\le t<+\infty\}$为二阶矩过程，若对任意的 $0\le t_1<t_2\le t_3<t_4$，都有

$$E\left[\left(X(t_2)-X(t_1)\right)\overline{\left(X(t_4)-X(t_3)\right)}\right]=0,$$

则称 $\boldsymbol{X}$ 为**正交增量过程**.

对于正交增量过程$\{X(t), 0\le t<+\infty\}$，如果规定 $X(0)=0$，那么当 $s<t$ 时，

$$\begin{aligned}
E\left[X(s)\overline{X(t)}\right]&=E\left[\left(X(s)-X(0)\right)\overline{\left(X(t)-X(s)+X(s)\right)}\right]\\
&=E|X(s)|^2 \quad \text{（由正交增量性）}\\
&=\Psi_{\boldsymbol{X}}(s),
\end{aligned}$$

即当 $s<t$ 时，$R_{\boldsymbol{X}}(s,t)=\Psi_{\boldsymbol{X}}(s)$．同理，当 $t\le s$ 时，$R_{\boldsymbol{X}}(s,t)=\Psi_{\boldsymbol{X}}(t)$．故 $R_{\boldsymbol{X}}(s,t)=\Psi_{\boldsymbol{X}}(s\wedge t)$，其中记号 $s\wedge t$ 表示从 s 与 t 中取较小的那个值，即 $s\wedge t=\min\{s,t\}$.

另外，当 $s<t$ 时，

$$\begin{aligned}
E|X(t)-X(s)|^2&=R_{\boldsymbol{X}}(t,t)-R_{\boldsymbol{X}}(t,s)-R_{\boldsymbol{X}}(s,t)+R_{\boldsymbol{X}}(s,s)\\
&=\Psi_{\boldsymbol{X}}(t)-\Psi_{\boldsymbol{X}}(s).
\end{aligned}$$

由此可知，正交增量过程的均方值函数 $\Psi_{\boldsymbol{X}}(\cdot)$是单调不减函数.

(3)　独立增量过程

设过程 $\boldsymbol{X}=\{X(t), 0\le t<+\infty\}$，若对任意的 $0\leqslant t_1<t_2<\cdots<t_n$，过程的增量

$$X(t_2)-X(t_1), X(t_3)-X(t_2), \cdots, X(t_n)-X(t_{n-1})$$

相互独立，则称 $\boldsymbol{X}$ 为**独立增量过程**．进一步地，若独立增量过程 $\boldsymbol{X}$ 的任一增量 $X(s+t)-X(s)$ $(s\geqslant 0, t>0)$的概率分布与 s 无关，则称 $\boldsymbol{X}$ 为**平稳独立增量过程**.

一般而言，因独立增量过程未必都存在二阶矩，故它未必是二阶矩过程，反之亦然.

因而，存在二阶矩的那些独立增量过程自然地成为了独立增量过程类中的重要子类．Poisson 过程和 Brown 运动就是其中的典型代表.

(4)　Markov 过程

Markov 过程是一类具有无后效性的随机过程．粗略地说，“无后效性”是指：已知过程

在现在时刻 t 所处的状态，它在将来时刻 u ($u > t$)所处状态的概率与它在过去时刻 s ($s < t$)的状态无关.

设过程 $\boldsymbol{X}=\{X(t), t\in T\}$，其中 $T\subset[0,+\infty)$．若对任意 $n\geq 1$，任意 n 个时刻 $0\leq t_1<t_2<\cdots<t_n$ 以及任意 $s>0$（$t_n+s\in T$），有

$$P\left\{X(t_n+s)\leq x\middle|X(t_1)=x_1,\cdots,X(t_n)=x_n\right\}$$
$$=P\left\{X(t_n+s)\leq x\middle|X(t_n)=x_n\right\}\quad (x,x_1,\cdots,x_n\in\boldsymbol{R}^1),$$

则称 $\boldsymbol{X}$ 为 **Markov 过程**.

对于 Markov 过程 $\boldsymbol{X}$，当它的参数集为 $T=\{0,1,2,\ \cdots\}$，状态空间 S 为可列集时，称 $\boldsymbol{X}$ 为**离散参数 Markov 链**．当它的参数集为 $T=[0,+\infty)$，状态空间 S 为可列集时，称 $\boldsymbol{X}$ 为**连续参数 Markov 链**．如果它的参数集 T 与状态空间 S 都是不可列集，那么称 $\boldsymbol{X}$ 为**一般状态空间上的 Markov 过程**.

(5)　鞅过程

设过程 $\boldsymbol{X}=\{X(t), t\in T\}$，若它满足以下条件

①　对任意 $t\in T$，$E|X(t)|<\infty$；

②　对任意 $t_1<t_2<\cdots<t_n<t_{n+1}$，有 $E\left[X(t_{n+1})\middle|X(t_1),X(t_2),\cdots,X(t_n)\right]=X(t_n)\ \ a.s.$，

则称 $\boldsymbol{X}$ 为鞅（**martingales**）.

习题 1

1.1　设 Ω 为某试验的样本空间，A 为 Ω 的非空子集，令 $\mathscr{F}_1=\{\varnothing,\Omega\}$, $\mathscr{F}_2=\{A,A^c,\varnothing,\Omega\}$, $\mathscr{F}_3=2^{\Omega}$，证明：$\mathscr{F}_1$，$\mathscr{F}_2$，$\mathscr{F}_3$ 皆为事件域且 $\mathscr{F}_1\subset\mathscr{F}_2\subset\mathscr{F}_3$。

1.2　一次投掷三颗匀称骰子并观察可能出现的点数，给出这一试验的概率空间。

1.3　证明 1.1 节定理 1 中的性质(1)~(5) 。

1.4　在给定的概率空间$(\Omega,\mathscr{F},P)$中，设 $A,B,C,A_1,\cdots,A_n\in\mathscr{F}$，证明下列不等式：

(1) $\left|P(AB)-P(A)P(B)\right|\leq\dfrac{1}{4}$ (提示：对 $0\leq x\leq 1$, 恒有 $x-x^2\leq\dfrac{1}{4}$)；

(2) $\left|P(AB)-P(BC)\right|\leq 1-P(AC)$；

(3) 若 $\bigcap\limits_{k=1}^{n}A_k\subset A$，则 $P(A)\geq\sum\limits_{k=1}^{n}P(A_k)-(n-1)$。

1.5　设有 n 张大小形状相同且编号分别为 $1, 2,\cdots, n$ 的卡片。采用有放回随机抽取方式直至抽到一张先前曾取过的卡片为止。设 X 是达成这一目的所需的抽取次数。求 X 的概率分布（提示：先算 $P\{X>k\}$。）。

1.6　设 X 与 Y 为独立随机变量，其中 X 服从参数为(2, 0.5)的二项分布，Y 服从参数为 1 的泊松分布，计算 $P\{Y\geq X\}$。

1.7　用归纳法证明 n 元分布函数的性质(2)：若 $x_i \le y_i (i=1,2,\cdots,n)$ ，则

$$\Delta_1\Delta_2\cdots\Delta_n F(x_1,\cdots,x_n) = P\left\{x_1 < X_1 \le y_1,\cdots,x_n < X_n \le y_n\right\} \ge 0,$$

其中 $\Delta_i F(x_1,\cdots,x_n) \triangleq F(x_1,\cdots,x_{i-1},y_i,x_{i+1},\cdots x_n) - F(x_1,\cdots,x_n)$ 表示 F 对第 i 个变元 $(1 \le i \le n)$作一阶差分。

1.8　设 $h(x,y) = e^{-(x+y)},\ x \ge 0, y \ge 0$；$g(x,y) = \begin{cases} 1, & x+y \ge 0, \\ 0, & x+y < 0. \end{cases}$ 证明：h 是二元单调不减函数但对每一变元却是单调减的；g 对每一变元是单调不减的但不是二元单调不减函数。

1.9　设 X_1，X_2，X_3 为相互独立的随机变量，已知 $\mathrm{E}(X_k) = 0,\ \mathrm{E}(X_k^2) = \sigma^2,\ k=1,2$ 。令

$Y = (X_1X_3 + X_2)\Big/\sqrt{1+X_3^2}$ ，求 Y 的期望与方差。

1.10　设随机变量 X 服从几何分布：$\mathrm{P}\left\{X = k\right\} = p(1-p)^k,\ k = 0,1,2,\cdots,$ 其中 $0 < p < 1$。求 $\mathrm{E}\left|X - 1/p\right|$（提示：$1 \equiv 1_{\{X \le c\}} + 1_{\{X > c\}}$）。

1.11　设随机变量 X 存在有限均值 μ，g 为单调不减函数，证明 $\mathrm{E}\left[g(X)(X-\mu)\right] \ge 0$ 。

1.12　设随机变量 $X \sim Exp(\lambda)$，$t > 0$ 为常数，求

(1) $\min\left\{X, t\right\}$ 的概率分布；

(2) 在 $X > t$ 的条件下，X 的条件概率密度。

1.13　对条件概率亦有相应的乘法公式及全概率公式，试推导下列等式：

(1) 设 $A, B_1, B_2, \cdots, B_n$ 为事件，若 $P\left(AB_1B_2\cdots B_{n-1}\right) > 0$，则

$$P\left(B_1B_2\cdots B_n \middle| A\right) = P\left(B_1 \middle| A\right)P\left(B_2 \middle| AB_1\right)\cdots P\left(B_n \middle| AB_1B_2\cdots B_{n-1}\right);$$

(2) 设 $A_1, A_2, \cdots, A_n, \cdots$ 为两两互斥事件，且事件 $B \subset \bigcup_{k=1}^{\infty} A_k$ ，又 $P(A) > 0$，则

$$P\left(B \middle| A\right) = \sum_{k=1}^{\infty} P\left(A_k \middle| A\right)P\left(B \middle| AA_k\right)。$$

1.14　设 A, B, C 为三个事件，证明 A 与 B 关于 C 条件独立当且仅当 $P\left(A \middle| BC\right) = P\left(A \middle| C\right)$。

1.15　设 X 为连续型随机变量，其概率密度为 $f_X(x) = xe^{-x},\ x > 0$，求 $\mathrm{E}\left(X \mid X > 1\right)$。

1.16　设 X 与 Y 为随机变量，并且 $0 < \mathrm{D}(X) < \infty$，$0 < \mathrm{D}(Y) < \infty$，证明：

$\mathrm{Cov}\left(X, \mathrm{E}(Y|X)\right) = \mathrm{Cov}(X, Y)$。

1.17　已知随机向量(X,Y)的概率密度为：

$$f_{X,Y}(x,y)=\begin{cases}2, & 0<y\le x<1,\\ 0, & \text{其他},\end{cases}$$

求条件期望 $\mathrm{E}(X\mid y)$ (0<y<1) 及 $\mathrm{E}(Y\mid x)$ (0<x<1)。

1.18 如果 X 关于 Y 的条件方差 $\mathrm{D}(X\mid Y)$ 被定义为：$\mathrm{D}(X\mid Y)=\mathrm{E}\left\{\left[X-E(X\mid Y)\right]^2\mid Y\right\}$，试证：

(1) $\mathrm{D}(X\mid Y)=\mathrm{E}(X^2\mid Y)-\left[E(X\mid Y)\right]^2$；

(2) （全方差公式）$\mathrm{D}(X)=\mathrm{E}\left[\mathrm{D}(X\mid Y)\right]+\mathrm{D}\left[\mathrm{E}(X\mid Y)\right]$。

1.19 设 X 与 Y 是相互独立的连续型随机变量，且 Y~$U(0,1)$，令 $Z=X+Y$，证明 Z 的概率密度为 $f_Z(z)=F_X(z)-F_X(z-1)$。

1.20 设 X 与 Y 是相互独立的随机变量，其中 X~$U(0,1)$，Y 服从 0-1 分布，即 $P\{Y=1\}=p$，$P\{Y=0\}=1-p\ (0<p<1)$，求 $X+Y$ 以及 XY 的概率分布。

1.21 设 X_1 与 X_2 是独立同指数分布 $Exp(\lambda)$的随机变量，

(1) 证明 X_1+X_2与$\dfrac{X_1}{X_1+X_2}$ 独立；

(2) 令 $Y_1=X_1+X_2,\ Y_2=X_1-X_2$，求 Y_1, Y_2 的联合概率密度与边缘概率密度。

1.22 设 X_1 与 X_2 为独立同 $N(\mu,\sigma^2)$ 分布的随机变量。令 $Y_1=X_1+X_2$，$Y_2=X_1-X_2$，

(1) 求 Y_1，Y_2 的联合概率密度；

(2) 求 Y_1 与 Y_2 的协方差。

1.23 设 X_1,X_2,X_3 独立同指数分布 $Exp(1)$的随机变量，令 $Y_1=\sqrt{X_1},\ Y_2=\sqrt{X_1+X_2},$

$Y_3=\sqrt{X_1+X_2+X_3}$，求

(1) Y_1,Y_2,Y_3 的联合概率密度；

(2) Y_1,Y_2 的联合概率密度。

1.24 已知随机变量 X 服从参数为 λ 的泊松分布．假若 λ 本身是另一个随机变量，试就以下两种情形，求 X 的概率函数（分布律）。

(1) λ 服从均值 $\mu=1/c$ ($c>0$)的指数分布；

(2) λ 服从参数为 α，c($\alpha>0$, $c>0$)的 Γ 分布

$$f(\lambda)=\begin{cases}c^{\alpha}\dfrac{\lambda^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda c}, & \lambda>0,\\ 0, & \lambda\le 0.\end{cases}$$

1.25 设 X 是取非负整数值的随机变量，其概率母函数为 $G(z)=\sum_{n=0}^{\infty}p_n z^n$。每观测到 X，则

做 X 重贝努里试验，其每次成功的概率为 p。以 Y 表示 X 重贝努里试验中的成功次数，求 Y 的概率母函数。

1.26 已知随机变量 X 的概率母函数为：

$$G_X(z)=\frac{8}{(3-z)(5-z)}$$

求 X 的概率分布 $\{p_k, k=0,1,\cdots\}$。

1.27 设 X_1 和 X_2 为相互独立的随机变量且同服从区间 $[\theta-1,\theta+1]$ 上的均匀分布。证明：

X_1-X_2 的概率密度与 θ 无关（考虑其矩母函数或特征函数）。

1.28 设随机变量 X 服从二项分布 $B(n, p)$，其中 $0 < p < 1$，求 X 的特征函数并借助特征函数计算 X 的期望和方差。

1.29 若 $(p)\lim\limits_{n\to\infty} X_n = X$，又存在常数 c，使得对一切 n，有 $|X_n|\le c$，则 $(ms)\lim\limits_{n\to\infty} X_n = X$。（提示：题设条件蕴含 $|X|\le c,\ a.s.$）

1.30 已知随机变量 $X\sim Exp(1)$，令

$$Y_n=\begin{cases} n, & X\le \dfrac{1}{n}, \\ 0, & X>\dfrac{1}{n}. \end{cases}$$

当 $n\to\infty$ 时，问 Y_n 是否 (1) 依分布收敛；(2) 依概率收敛；(3) 概率 1 收敛。如果收敛的话请指出相应的极限。

1.31 设随机过程 $\boldsymbol{X}=\{e^{-At},\ t>0\}$，其中 A 为具有概率密度 f 的随机变量。

(1) 求 $\boldsymbol{X}$ 的一维和二维概率分布；

(2) 如果 $A\sim Exp(2)$，求 $\boldsymbol{X}$ 的均值函数与自协方差函数。

1.32 设随机过程 $\boldsymbol{X}=\{At+B,\ t>0\}$，其中 A 和 B 是相互独立的标准正态变量，求 $\boldsymbol{X}$ 的一维和二维概率密度。

1.33 如果随机过程 $\boldsymbol{X}=\{X(t), t>0\}$ 只有四条样本曲线：$X(t,\omega_1)=1,\ X(t,\omega_2)=-1,$ $X(t,\omega_3)=\sin t,\ X(t,\omega_4)=\cos t\ \ (t>0)$，并且 $P(\{\omega_k\})=1/4,\ k=1,2,3,4$，求 $\boldsymbol{X}$ 的均值函数与自相关函数。

1.34 设 $\boldsymbol{X}=\{X(t), t\in T\}$ 与 $\boldsymbol{Y}=\{Y(t), t\in T\}$ 是定义在相同概率空间上的两个二阶矩过程，假设 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立，证明 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 不相关。

1.35 设 $\boldsymbol{X}=\{\sin(\lambda t), t=0,\pm1,\pm2,\cdots\}$，$\boldsymbol{Y}=\{\cos(\lambda t), t=0,\pm1,\pm2,\cdots\}$，其中 λ 是服从区间 $(0, 2\pi)$ 上均匀分布的随机变量，证明 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 不相关。