第1章 预备知识

一般来说，随机过程可视为动态的随机变量．对随机过程一般理论的研究通常认为开始于 20 世纪 30 年代，其理论基础被公认为是由柯尔莫哥洛夫（Kolmogorov）和杜布（Doob）奠定的．这一学科最早源于对物理学的研究．吉布斯（Gibbs）、玻尔兹曼（Boltzmann）、庞加莱（Poincare）、爱因斯坦（Einstein）、维纳（Wiener）及莱维（Levy）等人都做了这方面开创性的工作．1931 年，Kolmogorov 发表了《概率论的解析方法》，奠定了马尔可夫（Markov）过程的理论基础．1953 年，Doob 出版了《随机过程论》，系统且严格地叙述了随机过程基本理论．

随机过程是以概率论为基础的，本章对要用到的概率论基本知识做简单的回顾，包括概率的定义，随机变量及分布函数，数字特征，矩母函数和特征函数，n 维高斯分布，最后给出几种收敛的定义．

1.1 概　率

概率论中，随机试验是指在一定条件下，其结果事先不能确定的试验．所有可能的试验结果组成样本空间，通常用 Ω 表示，Ω 中的元素称为样本点，通常用 ω 表示．Ω 的子集称为事件，通常用 A、B、C 等表示，空集 $\varnothing$ 称为不可能事件，Ω 也称为必然事件（注：本书中，∞ 等同于 $+\infty$）．

定义 1.1 设 Ω 是一个集合，$A\subset\Omega$，如果 $P(A)$满足：

（1）$0\leqslant P(A)\leqslant 1$；

（2）$P(\varnothing)=1$；

（3）对两两互不相容事件 $A_1, A_2, \cdots, A_n\subset\Omega$ 有，$P\left(\bigcup\limits_{n=1}^{\infty}A_n\right)=\sum\limits_{n=1}^{\infty}P(A_n)$．

则称 $P(A)$是事件 A 的概率（Probability）．

由定义可知，概率具有如下性质：

（1）$P(\varnothing)=0$；

（2）如果 $A\subset B$，则 $P(A)\leqslant P(B)$；（单调性）

（3）$A_1, A_2, \cdots, A_n\subset\Omega$，有 $P(\bigcup\limits_{n=1}^{\infty}A_n)\leqslant\sum\limits_{n=1}^{\infty}P(A_n)$；（次可加性）

（4）若 $A_1, A_2, \cdots, A_n\subset\Omega$，有

$$P(A_1\cup A_2\cup\cdots\cup A_n)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)\text{；（多除少补原理）}\tag{1.1}$$

（5）若 $A_1\subset A_2\subset\cdots$，有 $\lim\limits_{n\to\infty}P(A_n)=P(\lim\limits_{n\to\infty}A_n)=P\left(\bigcup\limits_{n=1}^{\infty}A_n\right)$；若 $A_1\supset A_2\supset\cdots$，有 $\lim\limits_{n\to\infty}P(A_n)=P(\lim\limits_{n\to\infty}A_n)=P\left(\bigcap\limits_{n=1}^{\infty}A_n\right)$．（连续性）

性质（5）中，如果事件序列$\{A_n, n\geqslant 1\}$满足$A_n\subset A_{n+1}$（$n\geqslant 1$），则称为递增序列；如果满足$A_n\supset A_{n+1}$（$n\geqslant 1$），则称为递减序列．对递增序列$\{A_n, n\geqslant 1\}$，有$A_n=\bigcup_{k=1}^{n}A_k$，从而有$\lim_{n\to\infty}A_n=\bigcup_{n=1}^{\infty}A_n$；对递减序列$\{A_n, n\geqslant 1\}$，有$A_n=\bigcap_{k=1}^{n}A_k$，从而有$\lim_{n\to\infty}A_n=\bigcap_{n=1}^{\infty}A_n$．

下面看性质（4）的一个应用例子．

【例 1.1】（匹配问题）在一次集会上，n个人把他们的帽子放到房间的中央混合在一起，而后每人随机选取一顶，求恰好有k个人拿到自己帽子的概率．

解：记P_k为恰有k个人拿到自己帽子的概率，$k=0,1,2,\cdots,n$，并设A_i为第i个人拿到自己帽子这一事件，i=1, 2, $\cdots$, n，则P_0表示没有人拿到自己帽子的概率，且

$$
\begin{aligned}
P_0&=P(\overline{A}_1\overline{A}_2\cdots\overline{A}_n)=1-P(A_1\cup A_2\cup\cdots\cup A_n)\\
&=1-\left[\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)\right].
\end{aligned}
$$

依题意，A_i表示第i个人拿到自己帽子，因第i个人是等可能地选n个帽子中的任何一顶，故有

$$P(A_i)=\frac{1}{n},$$

同样，对任意的$i<j<k$，有

$$P(A_iA_j)=P(A_i)P(A_j\mid A_i)=\frac{1}{n}\cdot\frac{1}{n-1}=\frac{1}{n(n-1)},$$

$$P(A_iA_jA_k)=P(A_i)P(A_j\mid A_i)P(A_k\mid A_iA_j)=\frac{1}{n(n-1)(n-2)},$$

$$\cdots$$

$$P(A_1A_2\cdots A_n)=\frac{1}{n(n-1)\cdots 1}=\frac{1}{n!},$$

因此，没有一个人拿对自己帽子的概率为

$$
\begin{aligned}
P_0&=1-\left[C_n^1\frac{1}{n}-C_n^2\frac{1}{n(n-1)}+C_n^3\frac{1}{n(n-1)(n-2)}-\cdots+(-1)^{n-1}C_n^n\frac{1}{n!}\right]\\
&=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\cdots+\frac{(-1)^n}{n!}=\sum_{k=0}^{n}\frac{(-1)^k}{k!}.
\end{aligned}
$$

P_k为恰有k个人拿到自己帽子的概率．而前k个人拿到自己帽子，后$n-k$个人没有拿到自己帽子的概率为

$$
\begin{aligned}
P(A_1A_2\cdots A_k\overline{A}_{k+1}\overline{A}_{k+2}\cdots\overline{A}_n)&=P(A_1A_2\cdots A_k)P(\overline{A}_{k+1}\overline{A}_{k+2}\cdots\overline{A}_n\mid A_1A_2\cdots A_k)\\
&=\frac{1}{n(n-1)\cdots(n-k+1)}\cdot\sum_{m=0}^{n-k}\frac{(-1)^m}{m!},
\end{aligned}
$$

这里由P_0的表达式可知，

$$P(\bar{A}_{k+1}\bar{A}_{k+2}\cdots\bar{A}_n \mid A_1A_2\cdots A_k)=\sum_{m=0}^{n-k}\frac{(-1)^m}{m!}.$$

这种不同的排序方法共有 C_n^k 种，因此恰有 k 个人拿到自己帽子的概率为

$$P_k=C_n^k\frac{1}{n(n-1)\cdots(n-k+1)}\cdot\sum_{m=0}^{n-k}\frac{(-1)^m}{m!}=\frac{1}{k!}\cdot\sum_{m=0}^{n-k}\frac{(-1)^m}{m!}，\ k=0,1,2,\ \cdots,n.$$

注意到，显然当 $n\to\infty$ 时，$P_k\to\dfrac{\mathrm{e}^{-1}}{k!}$.

1.2　随机变量与分布函数

定义 1.2　设 $X=X(\omega)$是定义在 Ω 上，取值于 $\mathbf{R}$ 的实函数，如果对任意实数 x，$P\{\omega: X(\omega)\leqslant x\}$ 是某试验结果 $A\subset\Omega$ 的概率，则称 X 是一个**随机变量**（Random variable）．称函数

$$F(x)=P\{\omega：X(\omega)\leqslant x\},\qquad -\infty<x<\infty \tag{1.2}$$

为随机变量 X 的**（累积）分布函数**（Cumulative distribution function）.

注意，我们常常略去 ω 不写.

记 $\bar{F}(x)=1-F(x)=P(X>x)$ ．如果 X 表示机器的寿命，则 $P\{X>x\}$ 表示到 x 时机器还工作的概率，因此称 $\bar{F}(x)$ 为生存函数.

分布函数具有下列性质：

（1）$F(x)$是非减函数，即当 $x_1<x_2$ 时，$F(x_1)\leqslant F(x_2)$;

（2）$0\leqslant F(x)\leqslant 1$，且 $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$，$F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$；

（3）$F(x)$是右连续的，即 $F(x+0)=F(x)$.

可以证明，如果定义在实数域 $\mathbf{R}$ 上的实值函数 $F(x)$满足以上三条性质，则必存在一个随机变量 X，其分布函数为 $F(x)$.

随机变量 X 称为**离散的**（Discrete），如果它取值的集合是有限的或可数的．对离散型随机变量 X，其分布函数为

$$F(x)=\sum_{x_k\leqslant x}P\{X=x_k\}.$$

随机变量 X 称为**连续的**（Continuous），如果存在一非负可积函数 $p(x)$，使其分布函数

$$F(x)=\int_{-\infty}^{x}p(x)\mathrm{d}x.$$

这里 $p(x)$称为随机变量 X 的**概率密度函数**（Probability density function）.

表 1.1 和表 1.2 分别给出了一些常见的离散型和连续型随机变量的概率分布和概率密度函数.

定义 1.3　设 $\boldsymbol{X}=(X_1, X_2,\ \cdots, X_n)$是定义在 Ω 上的 n 维空间 $\mathbf{R}^n$ 中取值的向量函数，如果对任意 $\boldsymbol{x}=(x_1, x_2,\ \cdots, x_n)\in\mathbf{R}^n$，$P\{\omega: X_1\leqslant x_1, X_2\leqslant x_2,\ \cdots, X_n\leqslant x_n\}$是某试验结果 $A\subset\Omega$ 的概率，则称 $\boldsymbol{X}$ 为 n **维随机变量**或**随机向量**（Random vector）．称

$$F(\boldsymbol{x})=F(x_1,x_2,\cdots,x_n)=P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n\} \tag{1.3}$$

为 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)$的**联合分布函数**（Joint distribution function）.

联合分布函数$F(x_1,x_2,\cdots,x_n)$也有与一维随机变量的分布函数类似的性质.

设$F(x_1,x_2,\cdots,x_n)$为（$X_1, X_2, \cdots, X_n$）的联合分布函数，$i_1,i_2,\cdots,i_k$是$1,2,\cdots,n$中任意k个数，若$1\leqslant i_1<i_2<\cdots<i_k\leqslant n$，则$X_{i_1},X_{i_2},\cdots,X_{i_k}$的边际分布函数为

$$F_{i_1,i_2,\cdots,i_k}(x_{i_1},x_{i_2},\cdots,x_{i_k})=F(\infty,\cdots,\infty,x_{i_1},\infty,\cdots,\infty,x_{i_2},\infty,\cdots,\infty,x_{i_k},\infty,\cdots,\infty).$$

定义 1.4 设$X_1, X_2, \cdots, X_n$是定义在Ω上的n个随机变量，$F(x_1,x_2,\cdots,x_n)$和$F_1(x_1),\cdots,F_n(x_n)$分别是（$X_1, X_2, \cdots, X_n$）的联合分布函数和$X_1, X_2, \cdots, X_n$的边际分布函数，如果$\forall x_1,x_2,\cdots,x_n\in\mathbf{R}$，有

$$F(x_1,x_2,\cdots,x_n)=F_1(x_1)F_2(x_2)\cdots F_n(x_n), \tag{1.4}$$

则称随机变量$X_1, X_2, \cdots, X_n$是**相互独立**的（Independent）.

表 1.1 常见的离散型随机变量

分 布	概 率 分 布	期 望	方 差
0-1 分布	$P(X=1)=p,\ P(X=0)=q,\quad 0<p<1,\ p+q=1$	p	pq
二项分布	$P(X=k)=C_n^k p^k q^{n-k},\quad 0<p<1,\ p+q=1,\ k=0,1,\cdots,n$	np	npq
泊松分布	$P(X=k)=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad \lambda>0,\ k=0,1,\cdots$	λ	λ
几何分布	$P(X=k)=pq^{k-1},\ 0<p<1,\ p+q=1,\ k=1,2,\cdots$	$1/p$	q/p^2
负二项分布	$P(X=j)=C_{j-1}^{k-1}p^k q^{j-k},\ 0<p<1,\ p+q=1,\ j\geqslant k$	k/p	kq/p^2
离散均匀分布	$P\left(X=a+i\dfrac{b-a}{n}\right)=\dfrac{1}{n+1},\quad i=0,\ 1,\ \cdots,\ n$	$\dfrac{a+b}{2}$	$\dfrac{(n+2)(b-a)^2}{12n}$

表 1.2 常见的连续型随机变量

分 布	概率密度函数	期 望	方 差
指数分布	$p(x)=\lambda\mathrm{e}^{-\lambda x},\quad \lambda>0,x>0$	$1/\lambda$	$1/\lambda^2$
正态分布	$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-a)^2/2\sigma^2}$	a	σ^2
瑞利分布	$p(x)=\dfrac{x}{\sigma^2}\mathrm{e}^{-\frac{x^2}{2\sigma^2}},\quad \sigma>0,x>0$	$\sqrt{\dfrac{\pi}{2}}\sigma$	$\left(2-\dfrac{\pi}{2}\right)\sigma^2$
均匀分布	$p(x)=\begin{cases}1/(b-a), & a<x<b\\ 0, & 其他\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
χ^2 分布	$p(x)=\dfrac{x^{(n/2)-1}}{2^{N/2}\Gamma(n/2)}\mathrm{e}^{-\frac{x}{2}},\ n>0,x>0$	n	$2n$
Γ 分布	$p(x)=\dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x},\quad \alpha,\lambda>0,x>0$	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$
β 分布	$p(x)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, & 0<x<1\\ 0, & 其他\end{cases},\quad \alpha,\beta>0$	$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

1.3 数 字 特 征

随机变量的分布函数反映了随机变量的所有特征，但是在实际问题中，确定随机变量的分布一般是相当麻烦的．有时只需要知道随机变量的某些特征就够了．

定义 1.5 设随机变量 X 的分布函数为 $F(x)$，若 $\int_{-\infty}^{\infty}|x|\,\mathrm{d}F(x)<\infty$，则称

$$E(X)=\int_{-\infty}^{\infty}x\mathrm{d}F(x)=\begin{cases}\int_{-\infty}^{\infty}xp(x)\mathrm{d}x, & X\text{为连续型},\\ \sum_{k=1}^{\infty}x_kP\{X=x_k\}, & X\text{为离散型}.\end{cases} \tag{1.5}$$

为 X 的**数学期望**（Expectation）或**均值**.

注意：定义 1.5 中的积分为黎曼–斯蒂杰斯（Riemann-Stieltjes）积分，其性质与普通的黎曼（Riemann）积分类似.

定义 1.6 设 X 是随机变量，若 $E(X^2)<\infty$，则称

$$\mathrm{Var}(X)=E[X-E(X)]^2 \tag{1.6}$$

为 X 的**方差**（Variance），或记为 $D(X)$．称 $\sigma=\sqrt{\mathrm{Var}(X)}$ 为 X 的**标准差**（或均方差）.

表 1.1 和表 1.2 中给出了一些常见分布的期望和方差.

定义 1.7 设 X、Y 是随机变量，若 $E(X^2)<\infty$，$E(Y^2)<\infty$，则称

$$\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\} \tag{1.7}$$

为 X、Y 的**协方差**（Covariance）．称

$$\rho_{\mathrm{XY}}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \tag{1.8}$$

为 X、Y 的**相关系数**（Correlation coefficient）.

若 $\rho_{\mathrm{XY}}=0$ 或 $\mathrm{Cov}(X,Y)=0$，则称 X、Y 不相关.

下面给出一个常用的概率不等式.

定理 1.1 **（柯西–施瓦兹不等式 Cauchy-Schwarz inequality）**

如果 $E(X^2)$和 $E(Y^2)$存在，则

$$|E(XY)|\leqslant[E(X^2)]^{1/2}[E(Y^2)]^{1/2}\,. \tag{1.9}$$

定义 1.8 设 X、Y 是随机变量，若下列期望值存在，则

（1）称 $E(X^k)$为 X 的 k 阶（原点）**矩**（Moment）；

（2）称 $E\{[X-E(X)]^k\}$ 为 X 的 k 阶**中心矩**；

（3）称 $E(X^kY^l)$为 X 和 Y 的 $k+l$ 阶**混合矩**；

（4）称 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$为 X 和 Y 的 $k+l$ 阶**混合中心矩**.

数学期望 $E(X)$为 X 的一阶矩，方差 $\mathrm{Var}(X)$为 X 的二阶中心矩，协方差 $\mathrm{Cov}(X,Y)$为 X 和 Y 的二阶混合中心矩.

数学期望的一个重要性质是随机变量和的期望等于各个随机变量的期望的和，即

$$E(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}E(X_i)\,. \tag{1.10}$$

相应的方差的性质为

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j). \tag{1.11}$$

利用期望和方差的性质来求某些问题的期望和方差,有时会比用分布直接来求更加简单.

【例 1.2】（例 1.1 续）在一次集会上，n 个人把他们的帽子放到房间的中央混合在一起，而后每人随机选取一顶，求拿到自己帽子的人数 X 的期望和方差.

解：在例 1.1 中，已经算出拿到自己帽子的人数 X 的概率分布，由其概率分布求解期望和方差显然是比较烦琐的．为了求 X 的期望和方差，设

$$X = X_1 + X_2 + \cdots + X_n,$$

其中

$$X_i = \begin{cases} 1, & \text{如果第 } i \text{ 个人拿到自己的帽子}, \\ 0, & \text{其他}. \end{cases}$$

因第 i 个人是等可能地选 n 个帽子中的任何一顶，所以 $P(X_i = 1) = \dfrac{1}{n}$，这时 X_i 服从 0-1 分布，$E(X_i) = \dfrac{1}{n}$，$\mathrm{Var}(X_i) = \dfrac{1}{n}\left(1 - \dfrac{1}{n}\right) = \dfrac{n-1}{n^2}$．从而 X 的期望为

$$E(X) = \sum_{i-1}^{n} \frac{1}{n} = 1.$$

为了求出 X 的方差，先求协方差 $\mathrm{Cov}(X_i, X_j)$：

$$\mathrm{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j).$$

由于 $X_i X_j$ 也服从 0-1 分布，且

$$P(X_i X_j = 1) = P(X_i = 1, X_j = 1) = P(X_i = 1)P(X_j = 1 \mid X_i = 1) = \frac{1}{n} \cdot \frac{1}{n-1},$$

因此

$$E(X_i X_j) = \frac{1}{n(n-1)},$$

协方差

$$\mathrm{Cov}(X_i, X_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}.$$

X 的方差为

$$\mathrm{Var}(X) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \frac{n-1}{n^2} + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \frac{1}{n^2(n-1)}$$

$$= \frac{n-1}{n} + 2C_n^2 \frac{1}{n^2(n-1)} = \frac{n-1}{n} + \frac{1}{n} = 1.$$

从例 1.1 可以看出，当 $n \to \infty$ 时，X 的分布为

$$P\{X=k\}=\frac{\mathrm{e}^{-1}}{k!},\ k=0,1,2,\ \cdots.$$

这正好是$\lambda=1$的泊松（Poisson）分布.

【例 1.3】 为了使某种商品能够满足下个月的销量需求，商店需给出这种商品的订购量，假定需求量服从参数为λ的指数分布. 如果商店以每千克c元的价格购入该商品，以$s(s>c)$元的价格卖出.

（1）假定月底有存货，存货一文不值，缺货也不受处罚，商店应订购多少商品才能使商店的期望利润最大？

（2）假定月底没有卖出的存货以每千克$r(r<c)$元的价格退回，缺货会处以每千克p元的罚款，这时商店应订购多少商品才能使商店的期望利润最大？

解：（1）设X为需求量，商品订购量为x，利润为Z，则有

$$Z=s\min\{X,x\}-cx.$$

其中

$$\min\{X,x\}=X-(X-x)^{+},$$

$$(X-x)^{+}=\begin{cases}X-x, & X>x\\ 0, & X\leqslant x\end{cases}.$$

由条件期望公式

$$\begin{aligned}E[(X-x)^{+}]&=E[(X-x)^{+}\mid X>x]P\{X>x\}+E[(X-x)^{+}\mid X\leqslant x]P\{X\leqslant x\}\\&=E[(X-x)^{+}\mid X>x]\int_{x}^{+\infty}\lambda\mathrm{e}^{-\lambda t}\mathrm{d}t=E[X-x\mid X>x]\mathrm{e}^{-\lambda x}=\frac{1}{\lambda}\mathrm{e}^{-\lambda x},\end{aligned}$$

其中，最后一个等式成立，由指数分布的无记忆性可知，在$X>x$条件下，$Y=X-x$服从参数为λ的指数分布，因此$E[X-x\mid X>x]=\frac{1}{\lambda}$，期望利润

$$\begin{aligned}E(Z)&=sE[\min\{X,x\}]-cx=s[E(X)-E[(X-x)^{+}]-cx\\&=\frac{s}{\lambda}(1-\mathrm{e}^{-\lambda x})-cx.\end{aligned}$$

为使得商店的期望利润最大，可以通过对x求导，并令导数为 0，得

$$x=\frac{1}{\lambda}\ln\frac{s}{c}$$

时期望利润最大，这时最大的期望利润为

$$\max E(Z)=\frac{s}{\lambda}\left(1-\frac{c}{s}\right)-\frac{c}{\lambda}\ln\frac{s}{c}=\frac{s-c}{\lambda}-\frac{c}{\lambda}\ln\frac{s}{c}.$$

（2）如果月底没有卖出的存货以每千克r（$r<c$）元的价格退回，缺货会处以每千克p元的罚款，则

$$Z=s\min\{X,x\}-cx+r(x-X)^{+}-p(X-x)^{+},$$

其中

$$E[(x-X)^+]=x-E[\min\{X,x\}]=x-\frac{1}{\lambda}(1-\mathrm{e}^{-\lambda x}),$$

故

$$E(Z)=\frac{s}{\lambda}(1-\mathrm{e}^{-\lambda x})-cx+rx-\frac{r}{\lambda}(1-\mathrm{e}^{-\lambda x})-\frac{p}{\lambda}\mathrm{e}^{-\lambda x}$$

$$=\frac{s-r}{\lambda}-\frac{s+p-r}{\lambda}\mathrm{e}^{-\lambda x}-(c-r)x,$$

对 x 求导，并令导数为 0，得

$$x=\frac{1}{\lambda}\ln\frac{s+p-r}{c-r}$$

时期望利润最大，这时最大的期望利润为

$$\max E(Z)=\frac{s-r}{\lambda}-\frac{c-r}{\lambda}-\frac{c-r}{\lambda}\ln\frac{s+p-r}{c-r}=\frac{s-c}{\lambda}-\frac{c-r}{\lambda}\ln\frac{s+p-r}{c-r}.$$

注意到最佳订购量 x 是关于 s、p 和 r 的增函数，是关于 c 和 λ 的减函数.

1.4 矩母函数与特征函数

矩母函数和特征函数都是研究随机变量的分布的重要工具，它们均具有良好的分析性质.

定义 1.9 设随机变量 X 的分布函数为 $F(x)$，若 $E(\mathrm{e}^{tX})$ 存在，则称

$$G(t)=E(\mathrm{e}^{tX})=\int_{-\infty}^{\infty}\mathrm{e}^{tx}\,\mathrm{d}F(x),\quad -\infty<t<+\infty \tag{1.12}$$

为 X 的**矩母函数**（Moment generating function）.

若 X 为离散型随机变量，概率分布为 $P\{X=x_k\}=p_k, k=1,2,\cdots$，则

$$G(t)=\sum_{k=1}^{\infty}p_k\mathrm{e}^{tx_k}.$$

若 X 为连续型随机变量，概率密度函数为 $p(x)$，则

$$G(t)=\int_{-\infty}^{\infty}p(x)\mathrm{e}^{tx}\mathrm{d}x.$$

对 $G(t)$逐次求导，并计算在 $t=0$ 时的值，能得到 X 的各阶矩，即

$$G'(t)=E(X\mathrm{e}^{tX}),$$

$$G''(t)=E(X^2\mathrm{e}^{tX}),$$

$$\cdots$$

$$G^{(n)}(t)=E(X^n\mathrm{e}^{tX}).$$

令 $t=0$，得

$$E(X^n)=G^{(n)}(0). \tag{1.13}$$

【例 1.4】 求标准正态分布 $N(0, 1)$的矩母函数.

解:

$$G(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{tx-\frac{x^2}{2}}\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x-t)^2-t^2}{2}}\mathrm{d}x$$

$$=\mathrm{e}^{\frac{t^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{(x-t)^2}{2}}\mathrm{d}x=\mathrm{e}^{\frac{t^2}{2}},$$

故标准正态分布的矩母函数为

$$G(t)=\mathrm{e}^{\frac{t^2}{2}}.$$

当矩母函数存在时，随机变量的分布函数由矩母函数唯一地确定．但有时随机变量的矩母函数不一定存在．这时，使用特征函数更方便.

定义 1.10　设随机变量 X 的分布函数为 $F(x)$，则称

$$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}F(x),\quad -\infty<t<+\infty \tag{1.14}$$

为随机变量 X 的**特征函数**（Characteristic function），其中 $\mathrm{i}=\sqrt{-1}$.

若 X 为离散型随机变量，概率分布为 $P\{X=x_k\}=p_k, k=1,2,\cdots$，则

$$\varphi(t)=\sum_{k=1}^{\infty}p_k\mathrm{e}^{\mathrm{i}tx_k}.$$

若 X 为连续型随机变量，概率密度为 $p(x)$，则

$$\varphi(t)=\int_{-\infty}^{\infty}p(x)\mathrm{e}^{\mathrm{i}tx}\,\mathrm{d}x.$$

注意到特征函数是一个自变量为实数的复值函数，由于$|\mathrm{e}^{\mathrm{i}t}|=1$，所以随机变量的特征函数一定存在. 当 X 为连续型随机变量时，其特征函数是概率密度在 $t=-t$ 时的傅里叶（Fourier）变换.

【例 1.5】 求参数为 λ 的泊松分布的特征函数.

解: 设 X 服从参数为 λ 的泊松分布，则

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

其特征函数

$$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})=\sum_{k=0}^{\infty}\mathrm{e}^{\mathrm{i}tk}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda\mathrm{e}^{\mathrm{i}t})^k}{k!}=\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\mathrm{e}^{\mathrm{i}t}}=\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)},$$

故泊松分布的特征函数为

$$\varphi(t)=\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}.$$

对特征函数 $\varphi(t)$ 逐次求导并计算在 $t=0$ 时的值，同样能得到 X 的各阶矩，即

$$\varphi'(t)=\mathrm{i}E(X\mathrm{e}^{\mathrm{i}tX}),$$

$$\varphi''(t)=\mathrm{i}^2E(X^2\mathrm{e}^{\mathrm{i}tX}),$$

$$\cdots$$

$$\varphi^{(n)}(t) = \mathrm{i}^n E(X^n \mathrm{e}^{\mathrm{i}tX}) .$$

令 $t=0$，得

$$\varphi^{(n)}(0) = \mathrm{i}^n E(X^n) ,$$

故有

$$E(X^n) = \frac{\varphi^{(n)}(0)}{\mathrm{i}^n} . \tag{1.15}$$

特征函数有下列性质：

（1）有界性：$|\varphi(t)| \leqslant \varphi(0) = 1$；

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

（3）一致连续性：对任意 $\varepsilon > 0$，存在 $\delta > 0$，当 $|h| < \delta$ 时，对 t 一致地有

$$|\varphi(t+h) - \varphi(t)| < \varepsilon ;$$

（4）非负定性：对任意正整数 n，实数 $t_1, t_2, \cdots, t_n$ 及复数 $\lambda_1, \lambda_2, \cdots, \lambda_n$，有

$$\sum_{k=1}^{n}\sum_{j=1}^{n} \varphi(t_k - t_j)\lambda_k \overline{\lambda}_j \geqslant 0 ;$$

（5）若 $Y=aX+b$，则 Y 的特征函数为 $\varphi_Y(t) = \mathrm{e}^{\mathrm{i}tb}\varphi_X(at)$.

定理 1.2 （波赫纳尔-辛钦（Bochner-Khintchine）定理）

设 $\varphi(t)$ 满足 $\varphi(0)=1$，且在 $-\infty < t < \infty$ 上 $\varphi(t)$ 是连续的复值函数，则 $\varphi(t)$ 是特征函数的充要条件是它是非负定的.

表 1.3 和表 1.4 给出了部分常见分布的矩母函数和特征函数.

表 1.3 常见离散型随机变量的矩母函数和特征函数

分　　布	概 率 分 布	矩 母 函 数	特 征 函 数
二项分布	$P(X=k) = C_n^k p^k q^{n-k}$, $0<p<1$, $p+q=1$, $k=0,1,\cdots,n$	$(p\mathrm{e}^t + q)^n$	$(p\mathrm{e}^{\mathrm{i}t} + q)^n$
泊松分布	$P(X=k) = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$, $\lambda>0$, $k=0,1,\cdots$	$\mathrm{e}^{\lambda(\mathrm{e}^t-1)}$	$\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}$
几何分布	$P(X=k) = pq^{k-1}$, $0<p<1$, $p+q=1$, $k=1,2,\cdots$	$\dfrac{p\mathrm{e}^t}{1-q\mathrm{e}^t}$	$\dfrac{p\mathrm{e}^{\mathrm{i}t}}{1-q\mathrm{e}^{\mathrm{i}t}}$
负二项分布	$P(X=j) = C_{j-1}^{k-1} p^k q^{j-k}$, $0<p<1$, $p+q=1$, $j \geqslant k$	$\left(\dfrac{p\mathrm{e}^t}{1-q\mathrm{e}^t}\right)^k$	$\left(\dfrac{p\mathrm{e}^{\mathrm{i}t}}{1-q\mathrm{e}^{\mathrm{i}t}}\right)^k$

表 1.4 常见连续型随机变量的矩母函数和特征函数

分　　布	概率密度函数	矩 母 函 数	特 征 函 数
均匀分布	$p(x) = \begin{cases} 1/(b-a), & a<x<b \\ 0, & \text{其他} \end{cases}$	$\dfrac{\mathrm{e}^{tb}-\mathrm{e}^{ta}}{t(b-a)}$	$\dfrac{\mathrm{e}^{\mathrm{i}tb}-\mathrm{e}^{\mathrm{i}ta}}{\mathrm{i}t(b-a)}$
正态分布	$p(x) = \dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-a)^2/2\sigma^2}$	$\mathrm{e}^{at+\frac{1}{2}\sigma^2 t^2}$	$\mathrm{e}^{\mathrm{i}at-\frac{1}{2}\sigma^2 t^2}$
指数分布	$p(x) = \lambda\mathrm{e}^{-\lambda x}$, $x>0, \lambda>0$	$\left(1-\dfrac{t}{\lambda}\right)^{-1}$	$\left(1-\dfrac{\mathrm{i}t}{\lambda}\right)^{-1}$
Γ 分布	$p(x) = \dfrac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}$, $x>0, \alpha,\lambda>0$	$\left(1-\dfrac{t}{\lambda}\right)^{-\alpha}$	$\left(1-\dfrac{\mathrm{i}t}{\lambda}\right)^{-\alpha}$

对 n 维随机变量也可以定义特征函数.

定义 1.11 设 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 是 n 维随机变量，$\boldsymbol{t}=(t_1, t_2, \cdots, t_n)^{\mathrm{T}}\in\mathbf{R}^n$，则称

$$\varphi(\boldsymbol{t})=\varphi(t_1,t_2,\cdots,t_n)=E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}^T\boldsymbol{X}})=E(\mathrm{e}^{\mathrm{i}\sum\limits_{k=1}^{n}t_kX_k}) \tag{1.16}$$

为 n 维随机变量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 的**特征函数**.

可以证明，n 维随机变量的特征函数唯一地确定其联合分布函数.

当随机变量相互独立时，用其矩母函数或特征函数表示其分布，要比用分布函数表示更为方便.

例如，设随机变量 X_1 与 X_2 相互独立，$F_1(x)$ 和 $F_2(x)$ 分别是它们的分布函数，则由独立性知 X_1+X_2 的分布函数为

$$F(x)=P\{X_1+X_2\leqslant x\}=\int_{-\infty}^{+\infty}P\{X_1+X_2\leqslant x\mid X_1=t\}\mathrm{d}F_1(t)=\int_{-\infty}^{+\infty}F_2(x-t)\mathrm{d}F_1(t)\,.$$

称之为两个相互独立随机变量的和的分布函数等于这两个分布函数的**卷积**（Convolution），记为 $F_1*F_2(x)$，即

$$F_1*F_2(x)=\int_{-\infty}^{+\infty}F_2(x-t)\mathrm{d}F_1(t)\,. \tag{1.17}$$

由定义可看出，卷积具有对称性，即 $F_1*F_2(x)=F_2*F_1(x)$．卷积还有下列性质：

（1） $F_1*F_2(x)+F_1*F_3(x)=F_1*(F_2+F_3)(x)$；

（2） $F_1*(F_2*F_3)(x)=(F_1*F_2)*F_3(x)$.

如果知道随机变量 X_1 与 X_2 的特征函数 $\varphi_1(t)$ 和 $\varphi_2(t)$，由独立性可知 X_1+X_2 的特征函数是

$$\varphi(t)=E[\mathrm{e}^{\mathrm{i}t(X_1+X_2)}]=E(\mathrm{e}^{\mathrm{i}tX_1})\cdot E(\mathrm{e}^{\mathrm{i}tX_2})=\varphi_1(t)\cdot\varphi_2(t)\,.$$

即两个相互独立随机变量的和的特征函数等于这两个随机变量特征函数的**乘积**.

同样，n 个相互独立随机变量 $X_1, X_2, \cdots, X_n$ 的和 $X_1+X_2+\cdots+X_n$ 的分布函数等于这 n 个分布函数的 n 重卷积，记为

$$F_1*F_2*\cdots*F_n(x)\,.$$

n 个相互独立随机变量的和的特征函数等于这 n 个特征函数的乘积．即随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，它们的特征函数分别是 $\varphi_1(t),\varphi_2(t),\cdots,\varphi_n(t)$，则 $X_1+X_2+\cdots+X_n$ 的特征函数是

$$\varphi(t)=\varphi_1(t)\cdot\varphi_2(t)\cdots\varphi_n(t)\,. \tag{1.18}$$

矩母函数也有类似的结果.

【例 1.6】 设 n 个相互独立随机变量 $X_1, X_2, \cdots, X_n$ 服从均值为 $\dfrac{1}{\lambda}$ 指数分布，求其矩母函数及 $X_1+X_2+\cdots+X_n$ 的矩母函数.

解： X_i 的密度为 $p(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0\\ 0, & x\leqslant 0\end{cases}$，则矩母函数为

$$G(t)=E(\mathrm{e}^{tX_i})=\int_0^{+\infty}\lambda\mathrm{e}^{(t-\lambda)x}\mathrm{d}x=\frac{\lambda}{t-\lambda}\mathrm{e}^{(t-\lambda)x}\Big|_0^{+\infty}=\frac{\lambda}{t-\lambda}\mathrm{e}^{(t-\lambda)x}\Big|_0^{+\infty}=\frac{\lambda}{\lambda-t}，\ |t|<\lambda，$$

或写成 $G(t)=\left(1-\dfrac{t}{\lambda}\right)^{-1}$.

由于 $X_1+X_2+\cdots+X_n$ 的矩母函数 $G_n(t)=E(\mathrm{e}^{t\sum_{i=1}^{n}X_i})=E(\prod_{i=1}^{n}\mathrm{e}^{tX_i})=[G(t)]^n$，故有

$$G_n(t)=\left(1-\frac{t}{\lambda}\right)^{-n}.$$

$G_n(t)$ 是伽马分布 $\Gamma(n,\lambda)$ 的矩母函数.

1.5 条件期望

如果 X、Y 是离散型随机变量，则若 $P\{Y=y\}>0$，定义在 $Y=y$ 时，X 的条件概率为

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

给定 $Y=y$ 时，X 的条件分布函数为

$$F(x|y)=P\{X\leqslant x|Y=y\}=\sum_{x_i\leqslant x}\frac{P\{X=x_i,Y=y\}}{P\{Y=y\}},$$

而给定 $Y=y$ 时，X 的**条件期望**（Conditional expectation）定义为

$$E[X|Y=y]=\int x\mathrm{d}F(x|y)=\sum_{x_i}x_iP\{X=x_i|Y=y\}. \tag{1.19}$$

如果 X、Y 是连续型随机变量，其联合密度函数为 $p(x,y)$，若在 $Y=y$ 时 Y 的边缘密度函数 $p_Y(y)>0$，则在 $Y=y$ 时，X 的条件密度函数（Conditional probability density function）为

$$p(x|y)=\frac{p(x,y)}{p_Y(y)},$$

给定 $Y=y$ 时，X 的条件分布函数（Conditional distribution function）为

$$F(x|y)=P\{X\leqslant x|Y=y\}=\int_{-\infty}^{x}p(x|y)\mathrm{d}x,$$

而给定 $Y=y$ 时，X 的**条件期望**定义为

$$E[X|Y=y]=\int x\mathrm{d}F(x|y)=\int xp(x|y)\mathrm{d}x. \tag{1.20}$$

所以除了概率都是关于 $Y=y$ 时的条件概率外，其定义与无条件的情形完全一样.

$E[X|Y=y]$ 是 y 的函数，从而 $E[X|Y]$ 是随机变量 Y 的函数，也是随机变量.

下面给出条件期望的一个极其有用的性质.

定理 1.3 当随机变量 X 与 Y 的期望存在时，有

$$E(X)=E[E(X|Y)]=\int E(X\mid Y=y)\mathrm{d}F_Y(y). \tag{1.21}$$

如果 Y 是离散型，则定理 1.3 中的公式变为

$$E(X)=\sum_y E(X|Y=y)P\{Y=y\}.$$

如果 Y 是连续型，且有密度 $p_Y(y)$，则定理 1.3 中的公式为

$$E(X)=\int E(X|Y=y)\,p_Y(y)\mathrm{d}y.$$

对随机变量 X 与 Y 同时是离散型或连续型的情形，读者可以根据定义自行证明.

设 A 为任意事件，定义示性函数

$$I_A(\omega)=\begin{cases}1, & \omega\in A\\ 0, & \omega\notin A\end{cases},\tag{1.22}$$

是一个二值随机变量，显然

$$E(I_A)=P(A),$$

$$E(I_A\,|\,Y=y)=P(A\,|\,Y=y),$$

由定理 1.3 可知，对任意的随机变量，有

$$P(A)=\int P(A\,|\,Y=y)\mathrm{d}F_Y(y).\tag{1.23}$$

当 Y 是离散型随机变量时，有

$$P(A)=\sum_y P(A\,|\,Y=y)P\{Y=y\}\tag{1.24}$$

这正是全概率公式.

推论 1.1 （**条件方差公式**）

$$\mathrm{Var}(X)=E[\mathrm{Var}(X\,|\,Y)]+\mathrm{Var}[E(X\,|\,Y)].$$

证明：

$$\begin{aligned}E[\mathrm{Var}(X\,|\,Y)]&=E[E(X^2\,|\,Y)-(E(X\,|\,Y))^2]\\&=E[E(X^2\,|\,Y)]-E[(E(X\,|\,Y))^2]\\&=E(X^2)-E[(E(X\,|\,Y))^2],\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[E(X\,|\,Y)]&=E[(E(X\,|\,Y))^2]-[E[E(X\,|\,Y)]]^2\\&=E[(E(X\,|\,Y))^2]-[E(X)]^2,\end{aligned}$$

故

$$E[\mathrm{Var}(X\,|\,Y)]+\mathrm{Var}[E(X\,|\,Y)]=E(X^2)-[E(X)]^2=\mathrm{Var}(X).$$

【例 1.7】（**复合随机变量的期望和方差**）设 $X_1,X_2,\cdots$ 是独立同分布的随机变量，均值为 μ，方差为 σ^2. N 为取非负整数值的随机变量，方差存在，并与 $X_1,X_2,\cdots$ 相互独立. 记 $S=\sum_{i=1}^{N}X_i$，称 S 为复合随机变量. 求 S 的期望和方差.

解： $E(S)=E[E(S\,|\,N)]=\sum_{n=0}^{\infty}E(\sum_{i=1}^{N}X_i\,|\,N=n)P\{N=n\}$，

其中

$$E(\sum_{i=1}^{N} X_i \mid N=n) = E(\sum_{i=1}^{n} X_i \mid N=n) = E(\sum_{i=1}^{n} X_i) = nE(X_1) = n\mu ,$$

因此

$$E(S) = \mu \sum_{n=0}^{\infty} nP\{N=n\} = \mu E(N) .$$

由于 $\mathrm{Var}(S) = E[\mathrm{Var}(S \mid N)] + \mathrm{Var}[E(S \mid N)]$，而

$$\mathrm{Var}(S \mid N=n) = \mathrm{Var}(\sum_{i=1}^{N} X_i | N=n) = \mathrm{Var}(\sum_{i=1}^{n} X_i | N=n) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = n\sigma^2 ,$$

$$E(S \mid N=n) = n\mu ,$$

所以

$$\mathrm{Var}(S) = E(N\sigma^2) + \mathrm{Var}(N\mu) = \sigma^2 E(N) + \mu^2 \mathrm{Var}(N) .$$

1.6 *n* 维高斯分布

高斯分布在概率论中扮演着极为重要的角色．在概率论中曾经讨论过一维和二维正态随机变量，其概率密度函数分别为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} ,$$

$$p(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2{}^2}\right]\right\} .$$

如果记 $\boldsymbol{\mu} = (\mu_1,\mu_2)^{\mathrm{T}}$，$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1{}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2{}^2 \end{pmatrix}$，$\boldsymbol{x} = (x_1,x_2)^{\mathrm{T}}$，则二维正态随机变量的概率密度函数写成矩阵形式

$$p(\boldsymbol{x}) = p(x_1,x_2) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} .$$

这里$|\boldsymbol{\Sigma}|$表示矩阵 $\boldsymbol{\Sigma}$ 的行列式，$\boldsymbol{\Sigma}^{-1}$ 表示矩阵 $\boldsymbol{\Sigma}$ 的逆矩阵．

下面定义 n 维情形．

定义 1.12 若 n 维随机变量 $\boldsymbol{X} = (X_1, X_2 \cdots, X_n)^{\mathrm{T}}$ 的联合概率密度为

$$p(\boldsymbol{x}) = p(x_1,x_2,\cdots,x_n) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} , \qquad (1.25)$$

式中，$\boldsymbol{\mu} = (\mu_1, \mu_2 \cdots, \mu_n)^{\mathrm{T}}$ 是 $\boldsymbol{X}$ 的均值向量，$\boldsymbol{\Sigma} = (\sigma_{ij})_{n\times n}$ 是 $\boldsymbol{X}$ 的协方差矩阵，则称 $\boldsymbol{X}$ 为 ***n* 维正态随机变量**（Normal random variable）或 $\boldsymbol{X}$ 服从 ***n* 维高斯分布**，记为 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

可以证明，若 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则 $\boldsymbol{X}$ 的特征函数为

$$\varphi(\boldsymbol{t})=\varphi(t_1,t_2,\cdots,t_n)=\exp\left\{\mathrm{i}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{t}\right\},\tag{1.26}$$

这里 $\boldsymbol{t}=(t_1,t_2\cdots,t_n)^{\mathrm{T}}\in\mathbf{R}^n$ 为 n 维向量.

下面的高斯分布的性质可由特征函数方便地证明出.

定理 1.4　若 $\boldsymbol{X}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$，则 $\boldsymbol{Y}=\boldsymbol{AX}+\boldsymbol{b}\sim N(\boldsymbol{A\mu}+\boldsymbol{b},\boldsymbol{A\Sigma A}^{\mathrm{T}})$.

证明过程作为习题，请读者作答.

1.7　收　敛　性

本节给出几种不同的收敛性．先给出几个概率不等式.

引理 1.1　（Markov 不等式）

若 X 是非负值的随机变量，则对于任意 $a>0$，均有

$$P\{X\geqslant a\}\leqslant\frac{EX}{a}.$$

定理 1.5　（Chernoff 界）

令 X 的矩母函数为 $G(t)$，则对于任意 $a>0$，均有

$$P\{X\geqslant a\}\leqslant \mathrm{e}^{-ta}G(t),\ t>0,$$

$$P\{X\leqslant a\}\leqslant \mathrm{e}^{-ta}G(t),\ t<0.$$

定理 1.6　（Jensen 不等式）

若 $f(x)$ 是凸函数，如果期望存在，则有

$$E[f(X)]\geqslant f(EX).$$

注意，这里 $f(x)$ 是凸函数是指对于任意 $x_1<x_2$，有

$$\frac{f(x_1)+f(x_2)}{2}\geqslant f\left(\frac{x_1+x_2}{2}\right).$$

如果二阶导数存在，则 $f''(x)>0$．如 $f(x)=x^2$ 是凸函数，典型的凸函数还有 $f(x)=\mathrm{e}^x$、$f(x)=-\ln x$ 等.

1.7.1　依概率收敛

定义 1.13　设随机变量序列 $X_1,X_2,X_3,\cdots$，若存在某随机变量 X，使得 $\forall\varepsilon>0$，均有

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

则称随机变量序列 $\{X_n\}$**依概率收敛**（Convergence in probability）于 X，记为 $X_n\xrightarrow{P}X$，或 $\lim\limits_{n\to\infty}X_n=X,(P)$.

定义 1.14　设 $\{X_n\}$ 为一随机变量序列，$E(X_n)$存在，令 $\bar{X}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$，若

$$\bar{X}_n - E(\bar{X}_n) \xrightarrow{P} 0,$$

则称随机变量序列$\{X_n\}$服从（弱）**大数定律**.

1.7.2 概率 1 收敛

定义 1.15 设随机变量序列 $X_1, X_2, X_3, \cdots$，若存在某随机变量 X，使得

$$P\{\lim_{n\to\infty} X_n = X\} = 1,$$

则称随机变量序列$\{X_n\}$**概率 1 收敛**（Convergence with probability 1）于 X 或**几乎处处收敛**或**几乎必然收敛**（Convergence almost surely）于 X，记为 $X_n \xrightarrow{a.s.} X$ 或 $\lim\limits_{n\to\infty} X_n = X, (a.s.)$.

定义 1.16 设$\{X_n\}$为一随机变量序列，$E(X_n)$存在，令 $\bar{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$，若

$$\bar{X}_n - E(\bar{X}_n) \xrightarrow{a.s.} 0,$$

则称随机变量序列$\{X_n\}$服从**强大数定律**（Strong law of large number）.

定理 1.7 （Kolmogorov 强大数定律）

设 X_1, X_2, X_3, $\cdots$是相互独立的随机变量序列，且 $\sum\limits_{k=1}^{\infty}\frac{\mathrm{Var}(X_k)}{k^2} < \infty$，则$\{X_n\}$服从强大数定律．即

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

定理 1.7 中，如果 $X_1, X_2, X_3, \cdots$是同分布的，且期望 $E(X_n) = \mu$，令 $\bar{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$，则

$$P\left\{\lim_{n\to\infty}\bar{X}_n = \mu\right\} = 1.$$

1.7.3 均方收敛

定义 1.17 设随机变量序列 $X_1, X_2, X_3, \cdots$的 r 阶矩存在，若存在某一具有 r 阶矩的随机变量 X，使得

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

则称随机变量序列$\{X_n\}$***r* 次平均收敛**于 X 或 ***r* 阶矩收敛**于 X．若 $r=2$，称随机变量序列$\{X_n\}$**均方收敛**（Convergence in mean square）于 X，记为 $X_n \xrightarrow{m.s.} X$ 或 $\lim\limits_{n\to\infty} X_n = X, (m.s.)$.

1.7.4 依分布收敛

定义 1.18 设随机变量序列 $X_1, X_2, X_3, \cdots$，$F_1(x), F_2(x), F_3(x), \cdots$ 是对应的分布函数列，若存在某随机变量 X，分布函数为 $F(x)$，使得在 $F(x)$的所有连续点 x 上有

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

则称随机变量序列$\{X_n\}$**依分布收敛**（Convergence with distribution）于 X，记为 $X_n \xrightarrow{d} X$．这时称分布函数列$\{F_n(x)\}$**弱收敛**于分布函数 $F(x)$.

4 种收敛的关系如图 1.1 所示.

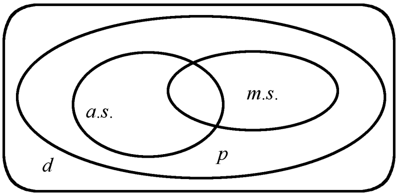

图 1.1　4 种收敛的关系

习　题　一

1.1　证明概率的下连续性：若 $A_1\subset A_2\subset\cdots$，且 $\bigcup_{n=1}^{\infty} A_n = A$，则有

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

1.2　设 $P(A)=0.8$， $P(B)=0.9$，证明 $P(AB)\geqslant 0.7$．更一般地，有

$$P(AB)\geqslant P(A)+P(B)-1.$$

1.3　设随机变量 X 的概率密度为

$$p(x)=\begin{cases} c(1-x^2), & |x|<1 \\ 0, & \text{其他} \end{cases}$$

求 c 的值，并求出 X 的分布函数.

1.4　设非负随机变量 X 的分布函数为 $F(x)$，证明

$$E(X)=\int_0^{+\infty}[1-F(x)]\mathrm{d}x=\int_0^{+\infty}P\{X>x\}\mathrm{d}x.$$

特别是当 X 取值为非负整数的离散型随机变量时，有

$$E(X)=\sum_{n=0}^{\infty}P\{X>n\}.$$

1.5　设随机变量 X 服从几何分布，即 $P(X=k)=pq^{k-1}$，$k=1,2,\cdots$，求随机变量 X 的期望、方差和特征函数.

1.6　设随机变量 X 服从参数为 α、λ的 Γ 分布，即密度函数为

$$p(x)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\lambda x),\quad x>0,$$

求随机变量 X 的期望、方差和矩母函数.

1.7 设 X、Y 的联合密度函数为

$$p(x,y)=\begin{cases}\dfrac{1}{y}\mathrm{e}^{-\left(y+\frac{x}{y}\right)}, & x>0, y>0\\ 0, & \text{其他}\end{cases},$$

（1）求 Y 的边缘密度；

（2）求 X 与 Y 的期望和方差；

（3）求 X 与 Y 的协方差.

1.8 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立同分布，且都服从参数为λ的指数分布，即密度函数为 $p(x)=\lambda\mathrm{e}^{-\lambda x}$，$x>0$，证明 $X_1+X_2+\cdots+X_n$ 是服从参数为 n、λ的 Γ 分布（提示：用特征函数的性质证）.

1.9 用特征函数证明定理 1.4.

1.10 设 X 为某动物生蛋的个数，服从参数为 λ 的泊松分布，即 $X\sim P(\lambda)$，其中 $\lambda>0$，如果每个蛋发育成小动物的概率是 p，Y 为此动物的后代的个数，证明 $Y\sim P(\lambda p)$.

1.11 设 X 和 Y 相互独立，分别服从参数 λ_1 和 λ_2 的泊松分布，求给定 $X+Y=n$ 的条件下 X 的条件期望.

1.12 一矿工被困在矿井中，要达到安全地带，有三个通道可选择. 他从第一个通道出去要走 3 小时可到达安全地带，从第二个通道出去要走 5 小时又返回原处，从第三个通道出去要走 7 小时又返回原处. 这名矿工任何时候做选择都等可能地选中三个通道中的一个，求他能到达安全地带的平均时间.

1.13 证明：$X_n\xrightarrow{P}\mu\Leftrightarrow X_n\xrightarrow{L}\mu$，其中 μ 为常数.