第二章　随机向量

一个向量，若它的分量都是随机变量，则称之为**随机向量**。本章首先对一元（即随机变量）的情形作一简单的回顾，然后在此基础上将其许多概念和结果直接推广到多元（即随机向量）的情形。一元情形可以看成是多元情形的一个特例。

§2.1　一元分布

一、随机变量与概率分布函数

随机变量是随机事件的数量表现，我们用 x,y,z 等表示。为方便起见，在不致引起混淆的情况下，它们的取值也用 x,y,z 等表示。随机变量具有这样两个**特点**：(1) 取值的随机性，即事先不能确定 x 取哪个值；(2) 取值的统计规律性，即完全可以确定 x 取某个值或 x 在某一个区间内取值的概率。随机变量 x 的**概率分布函数**（简称**分布函数**）定义为

$$F(a)=P(x\leqslant a) \tag{2.1.1}$$

它全面地描述了随机变量 x 的统计规律性。分布函数 $F(x)$ 具有下述**性质**：

(i) $F(x)$ 是非降函数，即若 $x_1<x_2$，则 $F(x_1)\leqslant F(x_2)$；

(ii) $\lim\limits_{x\to-\infty}F(x)=0$，$\lim\limits_{x\to\infty}F(x)=1$；

(iii) $F(x)$ 是右连续函数，即 $F(x+0)=F(x)$。

二、概率分布的类型

常用的随机变量有**离散型**和**连续型**两种，相应的概率分布分别称为**离散型分布**和**连续型分布**。

1. 离散型分布

若随机变量 x 只取有限个或可列个值，则称 x 为**离散型随机变量**。设离散型随机变量 x 的可能取值为 $a_1, a_2, \cdots$，取这些值的概率分别为 $p_1, p_2, \cdots$，则称

$$P(x=a_k)=p_k, \quad k=1,2,\cdots \tag{2.1.2}$$

为 x 的**分布列**，它具有如下两个**性质**：

(i) $p_k \geqslant 0, \quad k=1,2,\cdots$；

(ii) $\sum_{k=1}^{\infty} p_k=1$。

因此，分布列表明全部概率 1 在各个可能值之间分配的规律，它全面描述了离散型随机变量的统计规律性。x 的分布函数可表示为

$$F(a)=\sum_{a_k \leqslant a} P(x=a_k) \tag{2.1.3}$$

2. 连续型分布

若随机变量 x 的分布函数可以表示成

$$F(a)=\int_{-\infty}^{a} f(x)\mathrm{d}x \tag{2.1.4}$$

对一切 $a \in R$ 成立，则称 x 为**连续型随机变量**，称 $f(x)$ 为 x 的**概率密度函数**，简称为**概率密度**或**密度函数**或**密度**。对 $f(x)$ 的连续点必有 $F'(x)=f(x)$，密度函数 $f(x)$ 具有如下两个**性质**：

(i) $f(x) \geqslant 0$；

(ii) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x=1$。

概率密度函数全面地描述了连续型随机变量的统计规律性。

三、随机变量的数学期望和方差

若 x 为离散型随机变量，其分布列为(2.1.2)式，则 x 的**数学期望**(或称**均值**)和**方差**定义为

$$\begin{aligned} \mu=E(x)&=\sum_{k=1}^{\infty} a_k p_k \\ \sigma^2=V(x)=E(x-\mu)^2&=\sum_{k=1}^{\infty} (a_k-\mu)^2 p_k \end{aligned} \tag{2.1.5}$$

若 x 为连续型随机变量，其密度函数为 $f(x)$，则 x 的数学期望和方差定义为

$$\begin{aligned}\mu &= E(x) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x \\ \sigma^2 &= V(x) = E(x-\mu)^2 = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)\mathrm{d}x\end{aligned} \tag{2.1.6}$$

方差的一个简便计算公式是

$$\sigma^2 = E(x^2) - \mu^2$$

数学期望反映了随机变量 x 取值的平均水平，方差反映了随机变量 x 的可能取值在其均值周围的分散程度。方差的正平方根 $\sigma = \sqrt{V(x)}$ 称为随机变量 x 的标准差。数学期望和方差是随机变量最重要的两个数字特征。

数学期望具有下述**性质**：

(1) 设 c 是常数，则 $E(c) = c$。

(2) 设 k 是常数，x 是随机变量，则

$$E(kx) = kE(x)$$

(3) 设 $x_1, x_2, \cdots, x_n$ 为 n 个随机变量，则

$$E(x_1 + x_2 + \cdots + x_n) = E(x_1) + E(x_2) + \cdots + E(x_n)$$

方差具有如下**性质**：

(1) 设 c 是常数，则 $V(c) = 0$。

(2) 设 k 是常数，x 是随机变量，则

$$V(kx) = k^2 V(x)$$

(3) 设 $x_1, x_2, \cdots, x_n$ 为 n 个相互独立的随机变量（即这 n 个随机变量中任一个随机变量的取值皆不受其余 $n-1$ 个随机变量取值的影响）[①]，则

$$V(x_1 + x_2 + \cdots + x_n) = V(x_1) + V(x_2) + \cdots + V(x_n)$$

四、一些重要的一元分布

1. 二项分布

---

① 可将独立性条件减弱为 $x_1, x_2, \cdots, x_n$ 两两不相关。

在统计学中，二项分布是一个极其重要的分布，其重要性仅次于正态分布，是离散型分布中最重要的一个分布。

若离散型随机变量 x 的分布列为

$$P(x=k)=\binom{n}{k}p^k q^{n-k},\quad k=0,1,\cdots,n \tag{2.1.7}$$

其中 $0<p<1$，$q=1-p$，n 为自然数，则称 x 服从**二项分布**，记作 $x\sim b(n,p)$。当 $n=1$ 时，相应的分布称为**二点分布**。

2. 超几何分布

若离散型随机变量 x 的分布列为

$$P(x=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,\cdots,\min(n,M) \tag{2.1.8}$$

则称 x 服从**超几何分布**，记作 $x\sim H(M,N,n)$。

当 N 很大，n 相对较小时，超几何分布近似于二项分布，即

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\approx\binom{n}{k}p^k q^{n-k} \tag{2.1.9}$$

其中 $p=M/N$，$q=1-p$，这种近似可使计算量大为减少，也是二项分布应用较广的一个重要原因。

3. 泊松分布

若离散型随机变量 x 的分布列为

$$P(x=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad k=0,1,2,\cdots \tag{2.1.10}$$

其中参数 $\lambda>0$，则称 x 服从**泊松分布**，记作 $x\sim P(\lambda)$。

在 $\lambda=np$ 恒定的条件下，当 n 趋向无穷，同时 p 趋于零时，二项分布趋向于泊松分布。因此，当 n 很大，p 很小时，有如下的近似公式：

$$\binom{n}{k}p^k q^{n-k}\approx\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \tag{2.1.11}$$

4. 正态分布

正态分布是统计学中最重要的一个分布，它的应用极为广泛，它之所以如此重要，原因有三个：(1) 许多随机现象近似服从正态分布；(2) 由于中心极限定理的作用，有不少统计量的极限分布为正态分布；(3) 正态分布的理论非常完善，有许多好的性质，便于数学上的处理。

若连续型随机变量 x 的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty \tag{2.1.12}$$

则称 x 服从**正态分布**，记作 $x\sim N(\mu,\sigma^2)$，其中参数 μ 是数学期望，σ 是标准差。正态分布族中最重要的一个成员是 $\mu=0,\sigma=1$ 的正态分布，称为**标准正态分布**。当 n 很大，p 和 q 都不太小时，二项分布可用正态分布近似计算。

正态分布具有下述**基本性质**：

(1) 设 $x\sim N(\mu,\sigma^2)$，$y=ax+b$，其中 $a(a\neq 0)$，b 为任意常数，则

$$y\sim N(a\mu+b,a^2\sigma^2) \tag{2.1.13}$$

(2) 设 $x_1,x_2,\cdots,x_n$ 相互独立，且 $x_i\sim N(\mu_i,\sigma_i^2)$，$i=1,2,\cdots,n$，则对任意 n 个常数 $k_1,k_2,\cdots,k_n$(不全为零)，有

$$\sum_{i=1}^{n}k_ix_i\sim N\Big(\sum_{i=1}^{n}k_i\mu_i,\ \sum_{i=1}^{n}k_i^2\sigma_i^2\Big) \tag{2.1.14}$$

5. 卡方分布

设随机变量 $x_1,x_2,\cdots,x_n$ 皆服从 $N(0,1)$，且相互独立，则随机变量 $x=\sum\limits_{i=1}^{n}x_i^2$ 所服从的分布称为**卡方分布**，记作 $x\sim\chi^2(n)$，其中参数 n 称为**自由度**，表示平方和 $\sum\limits_{i=1}^{n}x_i^2$ 中独立随机变量项的个数。卡方分布的密度曲线如图 2.1.1 所示。

6. t 分布

设随机变量 $x\sim N(0,1)$，$y\sim\chi^2(n)$，且 x 与 y 相互独立，则随机变量 $t=\dfrac{x}{\sqrt{y/n}}$ 的分布称为 **t 分布**，记作 $t\sim t(n)$，其中参数 n 称为自由度。

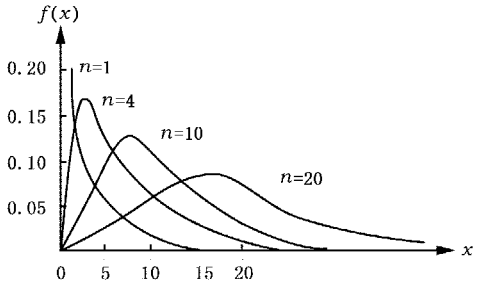

图 2.1.1 卡方分布的密度曲线

随着自由度 n 趋向于无穷大，t 分布以标准正态分布为极限。当 $n\geqslant 50$ 时，一般无法在 t 分布表中查出分位点，这时可用分布 $N(0,1)$ 替代分布 $t(n)$。t 分布的密度曲线如图 2.1.2 所示。

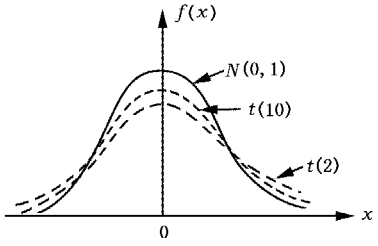

图 2.1.2 t 分布的密度曲线

7. F 分布

设随机变量 $x\sim\chi^2(n)$，$y\sim\chi^2(m)$，且 x 与 y 相互独立，则随机变量 $F=\dfrac{x/n}{y/m}$ 的分布称为自由度为 n 和 m 的 ***F* 分布**，记作 $F\sim F(n,m)$。F 分布的密度曲线如图 2.1.3 所示。

通常的 F 分布表只给出由右尾向左累加的概率，如图 2.1.4 所示。α 是一个较小的正数，给定 α，可查得临界值 $F_\alpha(n,m)$。而 $F_{1-\alpha}(n,m)$ 却不能直接查出，可利用 F 分布的一个**性质**：

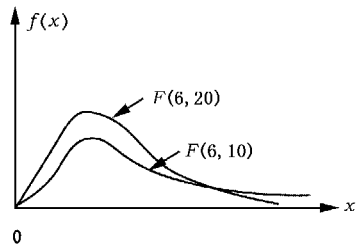

图 2.1.3　F 分布的密度曲线

$$F_{1-\alpha}(n,m)=\frac{1}{F_{\alpha}(m,n)} \tag{2.1.15}$$

查 F 分布表得出 $F_{\alpha}(m,n)$，再计算其倒数即可得到 $F_{1-\alpha}(n,m)$。

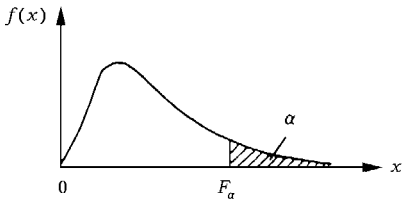

图 2.1.4　F 分布由右尾向左累加的概率

§2.2　多元分布

在许多随机现象中，我们需同时面对多个随机变量。例如，在体检时，要测量的指标有身高、体重、心跳、舒张压、收缩压等；在对居民家庭经济状况作调查时，调查指标有家庭收入、生活费支出、教育费支出、家庭人口等；医生在给病人诊断时，往往需根据病人的多项检查指标对其病症作出判断。这些指标都可以视为随机变量，那么，是否可以只是对这多个随机变量中的每一个单独地进行研究呢？如果这样的话，研究所得到的结论一般就仅是每一单个随机变量的结论，而不是多个随机