# 目 录

# 第 1 章　数理统计的基本概念

数理统计学是以概率论为基础，研究如何合理有效地收集观测数据，并采用合理有效的数学方法对数据加以整理与分析，从而对随机变量的分布或数字特征做出合理估计和推断的一门学科. 经过多年的研究和发展，统计学已经深入许多学科中，可以说，凡是实际问题中涉及数据的地方，都应当用统计学的方法进行分析，统计学也就有了应用的舞台. 随着统计学的发展与完善，其研究内容越来越丰富，且已经形成多个学科分支，如抽样调查、试验设计、回归分析、多元统计分析、时间序列分析、非参数统计和贝叶斯统计等.

## 1.1　总体与样本

数理统计研究的对象是受随机性影响的数据资料，它所考虑的问题及采用的方法不同于一般的资料统计，它更侧重于运用随机现象本身的规律性来考虑数据资料的收集、整理和分析，从而得到相应随机变量的分布或数字特征. 大量重复试验，也就是进行足够多的观测或试验，随机现象的规律性就会清楚地显示出来. 但是，实际上所允许的观测或试验次数总是有限的，有时甚至是少量的，因此我们关心的问题是如何有效地收集有限的观测数据. 本节主要介绍以下内容：

(1) 总体与抽样：总体与样本的概念；抽样方法：简单随机抽样、系统抽样与分层抽样.

(2) 随机变量及其分布函数：随机变量的类型、分布函数及其性质以及分位数的概念与性质.

(3) 特征函数和数字特征：介绍特征函数与母函数的概念性质、数学期望、方差与协方差的相关性质.

### 1.1.1　总体与抽样

**定义 1.1.1**　在一个统计问题中，把研究对象的全体称为**总体**

(Population)，组成总体的每个单元或元素称为**个体**(Object).

例如，研究对象是一批灯泡，那么，这批灯泡就是一个总体，其中的每个灯泡就是个体. 如果研究某一天的气温，那么这一天的气温就是一个总体，而各个时刻的气温就是个体. 若一个总体只有有限个个体，就称该总体为**有限总体**，否则称为**无限总体**. 在上面两例中，前者是有限总体，而后者是无限总体. 当个体数相当多时，可以把有限总体近似看成无限总体.

数理统计上所研究的某个总体，是指总体的数量标志，即数量标志是研究的直接对象. 以"一批灯泡"这个总体为例，考察灯泡的质量，灯泡的使用寿命就是这样一个数量标志. 如果采取随机抽取的方法来观测灯泡的使用寿命，那么使用寿命可看成一个随机变量$\mathbb{X}$，$\mathbb{X}$ 的分布情况就反映了这批灯泡使用寿命的整体状况. 因此，在这种意义下，以后将直接用随机变(向)量$\mathbb{X}$ 来表示一个总体.

采用对个体逐个观测来获取总体$\mathbb{X}$ 的分布往往是不切实际的. 一方面，在一些问题中，观测$\mathbb{X}$ 取值的试验具有破坏性，灯泡寿命的观测即如此，一旦某灯泡的使用寿命被测得，该灯泡就报废了；另一方面，有些观测要耗费大量的时间、人力和物力，或由于技术条件的限制，不可能逐个观察，因此只能进行有限次观测. 以灯泡为例，随机地从总体中抽取个体，其使用寿命值在观测前不可预知. 因此，每次观测的使用寿命值也可看做是一个随机变量. 若观测 $n$ 次，则依次对应着 $n$ 个随机变量 $X_1$，$X_2$，…，$X_n$，称这组随机变量 $X_1$，$X_2$，…，$X_n$ 为来自总体$\mathbb{X}$ 的一组样本，记为 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$，称 $n$ 为**样本容量**. $X_i$ 的取值 $x_i$ 表示第 $i$ 次观测所得结果，称 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 为样本观测值，简称**样本值**. 由于抽样的随机性，在一个总体中，两次抽取相同容量的样本，所得的样本值 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 及 $\boldsymbol{x}'=(x_1', x_2', \cdots, x_n')^{\mathrm{T}}$ 不一定相同. 因此，样本 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 是一个 $n$ 维随机变量.

抽取样本的目的是根据样本的取值情况来推断总体的信息. 样本取值的随机性会使推断带有一定程度的不确定性. 因此，应尽可能使在有限次观测中所抽取的样本能反映总体的状况. 这就要求样本具有以下两个性质：

1)(**代表性**)样本中的每个随机变量 $X_i$ 与总体$\mathbb{X}$ 有相同的分布；

2)(**独立性**)$X_1$，$X_2$，…，$X_n$ 相互独立，即每次观测结果互不影响.

满足上述两个条件的样本 $X_1$，$X_2$，…，$X_n$ 称为**简单随机样本**，简称**样本**，由此可以看出 $X_1$，$X_2$，…，$X_n$ 独立同分布(independent and identically distributed，以下简写为 i. i. d. )，本书讨论的样本，如果没有特殊交待，是指上述简单随机抽样所得到的样本.

综上所述，给出如下定义：

**定义 1.1.2** 设随机变量$\mathbb{X}$为总体，由总体抽取的个体 $X_1$，$X_2$，…，$X_n$ 是一组相互独立且与总体$\mathbb{X}$具有相同分布的随机变量序列，称 $n$ 维随机变量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 为来自总体$\mathbb{X}$的一组简单随机样本，简称**样本**(Sample)，称 $n$ 为**样本容量**(Sample size)；$\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 的所有可能取值的全体称为**样本空间**；$X$ 的一个观测值 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 为样本空间的一个点.

由定义易知，如果总体$\mathbb{X}$的密度函数为 $f(x)$，分布函数为 $F(x)$，那么 $n$ 维随机变量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 的联合密度函数与分布函数就分别为

$$f(x_1,x_2,\cdots,x_n) = \prod_{i=1}^{n} f(x_i), F(x_1,x_2,\cdots,x_n) = \prod_{i=1}^{n} F(x_i).$$

样本来自总体，我们总是由样本来认识总体的，因此有必要对获得样本的方法即抽样调查方法进行研究，这属于抽样调查技术范畴，感兴趣的读者可参考关于抽样调查的专著进一步学习了解，这里介绍另外两种常用的基本抽样方法.

**定义 1.1.3(系统抽样)** 将总体中所有个体排成一行，随机确定一个起点作为第一个进入样本的个体，然后每隔相等的间隔抽取一个个体进入样本，称这种抽取样本的方法为**系统抽样**(Systematic Sampling).

直观上看，系统抽样类似于对传送带上的产品进行定时抽样，因此系统抽样又称为**等距抽样或机械抽样**. 系统抽样适用于大样本情形，其优点是实施方便. 如果我们对总体排序的标志有所认识，知道一些规律并加以应用，采用系统抽样能达到相当好的实际效果. 当排序完全随机时，系统抽样与简单随机抽样效果相当. 例如考察某产品在市场上的日销售量，设计一年内每周抽取一次样本，所得

到52天的日销售量组成的样本就是系统抽样样本.

**定义1.1.4(分层抽样)** 将总体中的个体按相似原则分成若干层(组),将每一层(组)视为一个子总体,对于这些子总体分别进行简单随机抽样或其他抽样,称这种抽取样本的方法为**分层抽样**(Stratified Sampling).

从上面叙述知,总体为随机变(向)量$\mathbb{X}$,其分布函数完全描述了它的概率结构. 但在实际问题中,分布函数通常是未知的. 数理统计的中心任务,就是通过样本观测值,对总体的分布函数以及由此产生的问题进行合理的推断.

统计问题的基本提法是:给定样本$\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$;大多情形为独立同分布(即i. i. d.)的样本,这时$X_1$服从未知分布$F(x)$或$F(x, \theta)$,$\theta$未知;如$X_1 \sim N(\mu, \sigma^2)$,$\boldsymbol{\theta}=(\mu, \sigma^2)^{\mathrm{T}}$,泊松分布$P(\lambda)$等,统计问题就是从样本$\boldsymbol{X}=(X_1, X_2, \cdots, X_n)^{\mathrm{T}}$出发,对未知分布$F(x)$或$F(x, \theta)$进行统计推断(诸如参数估计、假设检验、回归分析等).

因此,随机变量及其分布函数,既是统计推断的目的,也是统计推断的基础. 一切统计问题都离不开随机变量(random variable,以下简写为r. v.)及其分布,熟练掌握这方面的知识,对学好数理统计并应用于实践是非常必要和有益的,关于概率论的基础知识,读者可参见王梓坤(1979),李贤平(1997)以及严士健,刘秀芳(1994).

### 1.1.2 随机变量及其分布函数

随机变量$X$是关于样本点的实值函数,一元分布函数定义为

$$F(x)=P(X \leqslant x), \quad x \in \mathbf{R}.$$

由概率论的知识,分布函数$F(x)$的特征性质是:$F(x)$具有非降性、有界性与右连续性,即函数$F(x)$为分布函数的充分必要条件如下:

(i)(非降性)$F(x)$为$\mathbf{R}$上的非降函数,即$\forall x_1, x_2 \in \mathbf{R}$,$x_1<x_2 \Rightarrow F(x_1) \leqslant F(x_2)$.

(ii)(有界性)$0 \leqslant F(x) \leqslant 1$,$F(+\infty)=1$,$F(-\infty)=0$.

(iii)(右连续性)$F(x+0)=F(x)$.

通常分布函数可分为绝对连续型、离散型和奇异型.

**例 1.1.1** 设离散型随机变量 $X$ 的概率分布为

| $X$ | 0 | 1 | 2 |
|---|---|---|---|
| $P$ | 0.2 | 0.5 | 0.3 |

(1) 求 $X$ 的分布函数 $F(x)$，并画出 $F(x)$ 的图形；

(2) 求 $P(X\leqslant 3)$，$P(-1\leqslant X\leqslant 1)$.

**解** (1) 由于 $X$ 只可能取 0，1，2，故

当 $x<0$ 时，$\{X\leqslant x\}=\varnothing$，$F(x)=P(X\leqslant x)=0$；

当 $0\leqslant x<1$ 时，$\{X\leqslant x\}=\{X=0\}$，$F(x)=P(X\leqslant x)=0.2$；

当 $1\leqslant x<2$ 时，$\{X\leqslant x\}=\{X=0$ 或 $X=1\}=\{X=0\}\cup\{X=1\}$，$F(x)=P(X\leqslant x)=0.7$；

当 $x\geqslant 2$ 时，$\{X\leqslant x\}=\Omega$，$F(x)=P(X\leqslant x)=1$.

归纳上述结果得

$$F(x)=\begin{cases}0, & x<0,\\ 0.2, & 0\leqslant x<1,\\ 0.7, & 1\leqslant x<2,\\ 1, & x\geqslant 2.\end{cases}$$

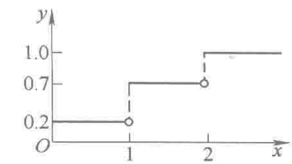


图 1.1.1 例 1.1.1 的分布函数的图形

若记**单位阶梯函数**为 $U(x)\triangleq\begin{cases}0, & x<0,\\ 1, & x\geqslant 0\end{cases}=I\{x\geqslant 0\}$，则上面的分布函数为

$$F(x)=0.2U(x)+0.5U(x-1)+0.3U(x-2).$$

其图形如图 1.1.1 所示.

(2) $P(X\leqslant 3)=F(3)=1$，$P(-1\leqslant X\leqslant 1)=F(1)-F(-1-0)=0.7$.

若随机变量 $X$ 取有限或可数个值，则其分布函数为离散型；其概率分布为

$$P(X=x_k)=p_k\triangleq f(x_k),k=0,\pm 1,\pm 2,\cdots,\sum_k p_k=1,$$

例如，泊松分布 $P(X=k)=\lambda^k\exp(-\lambda)/k!$，$k=0,1,2,\cdots$. 退化分布是一个重要特例，这时 $P(X=0)=1$ 或 $P(X=c)=1$，其分布函数为

$$F(x)\triangleq U(x)=\begin{cases}0, & x<0,\\ 1, & x\geqslant 0,\end{cases}$$

或 $F(x)=U(x-c)=I(x\geqslant c)$，其中 $I(x\geqslant c)$ 为示性函数. 对于一个

集合 $A$，示性函数 $I(A)$ 定义为 $I(A)=\begin{cases}0, & x\notin A,\\ 1, & x\in A\end{cases}\triangleq I_A(x)$.

因此，一般情形下离散型分布的分布函数可表示为

$$F(x)=\sum_k p_k U(x-x_k)=\sum_{x_k\leqslant x} f(x_k). \tag{1.1.1}$$

若连续型随机变量 $X$ 的密度函数 $f(x)$，使 $F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t$，$F'(x)=f(x)$，则称 $F(x)$ 为绝对连续型. 若 $F(x)$ 连续但不存在密度函数，则称 $F(x)$ 为奇异型；奇异型分布在实际问题中很少出现.

离散型分布函数亦可表示为积分形式，今简要介绍勒贝格-斯蒂尔切斯(Lebesgue-Stieltjes)积分如下(可参见 Cramer，1946，第 7 章和第 9 章；Rao，1973，第 2 章).

以下仅介绍 $\mathbf{R}^1$ 上的积分，$\mathbf{R}^n$ 上的积分完全类似.

给定 $\mathbf{R}^1$ 上的有界函数 $g(x)$ 和有限测度 $\mu(\cdot)$，将一个有穷或无穷区间 $[a,b]$ 分划为有限个互不相交的可测集 $\Delta x_i(i=1,\cdots,n)$ 之和. 对于任一"分划"，记 $S_m=\sum_{i=1}^{n} m_i\mu(\Delta x_i)$ 和 $S_M=\sum_{i=1}^{n} M_i\mu(\Delta x_i)$，其中 $\mu(\Delta x_i)$ 为 $\Delta x_i$ 的测度；$m_i$ 和 $M_i$ 分别为 $g(x)$ 在 $\Delta x_i$ 上的下确界与上确界. 若 $\max\limits_{1\leqslant i\leqslant n}\mu(\Delta x_i)\to 0$ 时，$S_m$ 和 $S_M$ 的极限存在且相等，则其极限称为函数 $g(x)$ 在区间 $[a,b]$ 上关于测度 $\mu(\cdot)$ 的勒贝格-斯蒂尔切斯积分，并记为 $\int_a^b g(x)\mathrm{d}\mu(x)$ 或 $\int_a^b g(x)\mu(\mathrm{d}x)$. 以下介绍与分布函数有关的两种情况.

(1) 若 $\mu(\cdot)$ 为勒贝格测度，$\Delta x_i$ 为区间，则 $\mu(\Delta x_i)$ 为区间长度；$\int_a^b g(x)\mathrm{d}\mu(x)=\int_a^b g(x)\mathrm{d}x$. 因此，若 $F(x)$ 为绝对连续型分布函数，存在密度函数 $f(x)$，则有

$$F(x)=\int_{-\infty}^{x} f(y)\mu(\mathrm{d}y). \tag{1.1.2}$$

(2) 若 $\mu(\cdot)$ 为计数测度. 设点列 $x_k$ 有可数个值 $x_1, x_2, \cdots, x_k, \cdots$，计数测度 $\mu(A)$ 表示 $A$ 所包含点列 $x_k$ 中点的个数. 因此对区间 $\max\limits_{1\leqslant i\leqslant n}\mu(\Delta x_i)\to 0$ 有：若 $\Delta x_i$ 中包含一个点列 $x_k$ 中的点，则 $\mu(\Delta x_i)=1$；若 $\Delta x_i$ 不包含 $x_k$ 中的点，则 $\mu(\Delta x_i)=0$；对于定义在点

列 $x_k$ 上的函数 $g(x)$，由勒贝格-斯蒂尔切斯积分的定义可知

$$\int_a^b g(x)\,\mathrm{d}\mu(x) = \sum_{x_k \in [a,b]} g(x_k). \tag{1.1.3}$$

特别地，对于离散型分布的分布函数 $F(x)$，由式(1.1.1)可知，若在点列 $x_k$ 上定义计数测度以及 $f(x_k)=p_k=P(X=x_k)$，则有

$$F(x) = \sum_{x_k \leqslant x} f(x_k) = \int_{-\infty}^{x} f(y)\,\mathrm{d}\mu(y). \tag{1.1.4}$$

比较式(1.1.2)和式(1.1.4)，通常把 $P(X=x_k)=f(x_k)=p_k$ 称为离散型分布的密度函数.

因此，对于服从绝对连续型和离散型分布的随机变量，其函数 $\varphi(X)$ 的数学期望可表示为

$$\begin{aligned} E[\varphi(X)] &= \int_{\mathbf{R}} \varphi(x) f(x)\,\mathrm{d}\mu(x) \\ &= \int_{\mathbf{R}} \varphi(x)\,\mathrm{d}F(x);\ \mathrm{d}F(x) = f(x)\,\mathrm{d}\mu(x). \end{aligned}$$

特别地，概率 $P(A)$ 可表示为 $P(A) = \int_{\mathbf{R}} I_A(x) f(x)\,\mathrm{d}\mu(x) = \int_A f(x)\,\mathrm{d}\mu(x) = \int_A \mathrm{d}F(x).$

另外，勒贝格-斯蒂尔切斯积分可推广到 $n$ 维空间，可参见 Cramer(1946，第 9 章).

在理论与实际问题中，经常要知道分布函数 $F(x)$ 反函数的值，即 $F^{-1}(p)$ 的值，其中 $0<p<1$(如 $p=0.05$，$0.95$ 等)；通常称为分位数或分位点. 由于 $F(x)$ 非降，不一定连续，因此 $F^{-1}(p)$ 的值可能不存在或不唯一. 以下定义保证了分位数的存在性和唯一性.

**定义 1.1.5** 分布函数 $F(x)$ 的 $p$ 分位数或 $p$ 分位点，定义为

$$x_p \triangleq \inf\{x: F(x) \geqslant p\}.$$

由下确界的定义知 $x_p$ 存在且唯一. 也容易证明，分位数 $x_p$ 满足

i) 若 $x'<x_p$ 则 $F(x')<p$；若 $x' \geqslant x_p$，则 $F(x') \geqslant p$；

ii) $F(x_p-0) \leqslant p \leqslant F(x_p)$；若 $x_p$ 为 $F(x)$ 的连续点，则 $F(x_p)=p$.

iii) 定义 $y=F(x)$ 的反函数为 $x=F^{-1}(y) \triangleq \inf\{x:\ F(x)>y\}$，若 $Y \sim U(0,\ 1)$，则有 $X=F^{-1}(Y)$ 的分布函数为 $F(x)$.

常用的分位点有 $x_{0.05}$，$x_{0.1}$，$x_{0.90}$，$x_{0.95}$，$x_{0.5}$等. $x_{0.5}$通常称为中位数，对于密度函数为$f(x)$的连续型随机变量，其 $p$ 分位数如图 1.1.2 所示.

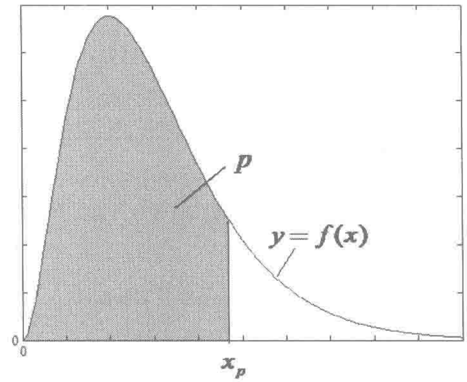


图 1.1.2　$p$ 分位数示意图

**定理 1.1.1**　函数 $g(c)=E|X-c|$在中位数 $c=x_{0.5}$时达到最小值.

**证明**　设 $X$ 的分布函数为 $F(x)$，若 $c<x_{0.5}$，则

$$
\begin{aligned}
g(c) &= \int_{-\infty}^{+\infty}|x-c|\mathrm{d}F(x) = \int_{-\infty}^{c}(c-x)\mathrm{d}F(x) + \\
&\quad \left(\int_{c}^{x_{0.5}}+\int_{x_{0.5}}^{+\infty}\right)(x-c)\mathrm{d}F(x) \\
&= \int_{-\infty}^{c}[(x_{0.5}-x)-(x_{0.5}-c)]\mathrm{d}F(x) + \\
&\quad \int_{c}^{x_{0.5}}[(x_{0.5}-x)+2(x-c)-(x_{0.5}-c)]\mathrm{d}F(x) + \\
&\quad \int_{x_{0.5}}^{+\infty}[(x-x_{0.5})+(x_{0.5}-c)]\mathrm{d}F(x) \\
&= \int_{-\infty}^{+\infty}|x-x_{0.5}|\mathrm{d}F(x)+2\int_{c}^{x_{0.5}}(x-c)\mathrm{d}F(x) + \\
&\quad (x_{0.5}-c)[P(X\geqslant x_{0.5})-P(X<x_{0.5})].
\end{aligned}
$$

由 $c<x_{0.5}$ 及 $x_{0.5}$ 的定义知 $P(X\geqslant x_{0.5})\geqslant 1/2$ 和 $P(X<x_{0.5})=F(x_{0.5}-0)\leqslant 1/2$，从而 $g(c)\geqslant E|X-x_{0.5}|$；反之，若 $c>x_{0.5}$，则

$$
\begin{aligned}
g(c) &= \int_{-\infty}^{+\infty} |x-c| \mathrm{d}F(x) = \left(\int_{-\infty}^{x_{0.5}} + \int_{x_{0.5}}^{c}\right)(c-x)\mathrm{d}F(x) + \\
&\quad \int_{c}^{+\infty}(x-c)\mathrm{d}F(x) \\
&= \int_{-\infty}^{x_{0.5}}[(x_{0.5}-x)+(c-x_{0.5})]\mathrm{d}F(x) + \int_{x_{0.5}}^{c}[(x-x_{0.5}) + \\
&\quad 2(c-x)-(c-x_{0.5})]\mathrm{d}F(x) + \\
&\quad \int_{c}^{+\infty}[(x-x_{0.5})-(c-x_{0.5})]\mathrm{d}F(x) \\
&= \int_{-\infty}^{+\infty}|x-x_{0.5}|\mathrm{d}F(x) + 2\int_{x_{0.5}}^{c}(c-x)\mathrm{d}F(x) + (c-x_{0.5}) \\
&\quad [P(X \leqslant x_{0.5}) - P(X > x_{0.5})].
\end{aligned}
$$

由 $c > x_{0.5}$ 及 $x_{0.5}$ 的定义知 $P(X \leqslant x_{0.5}) \geqslant 1/2$ 和 $P(X > x_{0.5}) \leqslant 1/2$，从而 $g(c) \geqslant E|X - x_{0.5}|$. 总之，对于任何 $c$，$g(c) \geqslant g(x_{0.5})$，且当 $c = x_{0.5}$ 时等号成立.

### 1.1.3 特征函数和数字特征

分布函数虽然完全描述了一个随机变量的概率结构，但也存在一些不能令人满意的地方. 比如，分布函数具有单调非降、有界、右连续性，不能保证它是一致连续和绝对连续的. 因此分布函数的分析性质并不完美. 通常引进其他工具作为分布函数的有益补充.

随机变量 $X$ 的**数学期望**或**均值**为 $\mu = EX$；方差为 $\sigma^2 = \mathrm{Var}(X)$，而 $\sigma^2$ 的算术平方根 $\sigma$ 通常称为**标准差**.

随机变量 $X$ 的**矩母函数**(Moment Generating Function)和**特征函数**分别定义为

$$M(\boldsymbol{t}) \triangleq E(\mathrm{e}^{tX}) \quad 和 \quad \varphi(\boldsymbol{t}) \triangleq E(\mathrm{e}^{\mathrm{i}tX}) = M(\mathrm{i}\boldsymbol{t}).$$

随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 的**矩母函数**和**特征函数**分别定义为

$$M(t) \triangleq E(\mathrm{e}^{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}}) \quad 和 \quad \varphi(t) \triangleq E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{X}}) = M(\mathrm{i}\boldsymbol{t}),$$

式中，$\boldsymbol{t} = (t_1, t_2, \cdots, t_n)^{\mathrm{T}}$.

特征函数与矩母函数和随机变量的分布具有一一对应的关系，通常用特征函数作为工具来确定统计分布.

常用的性质有

i) $M^{(k)}(0) = E(X^k) \triangleq a_k$，$\varphi^{(k)}(0) = \mathrm{i}^k a_k$，$E(X^k) = \mathrm{i}^{-k}\varphi^{(k)}(0)$；

ii) $\varphi(t) = 1 + \sum_{k=1}^{\infty} a_k \frac{(\mathrm{i}t)^k}{k!}, M(t) = 1 + \sum_{k=1}^{\infty} a_k \frac{t^k}{k!}$;

iii) 给定随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$，其分量 $X_1, X_2, \cdots, X_n$ 相互独立的充要条件为

$$\varphi_X(t_1, \cdots, t_n) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2)\cdots\varphi_{X_n}(t_n).$$

由性质 ii) 可知，特征函数和矩母函数的泰勒展开式的系数为随机变量 $X$ 的各阶矩. 近年来，其对数的泰勒展开式的系数也经常用到，今介绍如下.

**定义 1.1.6** (**累积量**，Cumulant)　$\log\varphi(t)$ 的泰勒展开式的系数称为累积量，若

$$\log\varphi(t) = \sum_{r=1}^{\infty} \kappa_r \frac{(\mathrm{i}t)^r}{r!},$$

则系数 $\kappa_r$ 称为 $r$ **阶累积量**.

随机变量 $X$ 的累积量 $\kappa_r$ 与其各阶矩 $a_k$ 有密切的关系. 其常用性质有

i) 累积量与各阶矩可互相表示. 其前 3 阶矩的关系为

$$\kappa_1 = a_1,\ \kappa_2 = m_2 = E(X - EX)^2,\ \kappa_3 = a_3 - 3a_1a_2 + 2a_1^3;$$

$$a_1 = \kappa_1,\ a_2 = \kappa_2 + \kappa_1^2,\ a_3 = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3;$$

ii) 若 $X, Y$ 独立，则有 $\kappa_r(X+Y) = \kappa_r(X) + \kappa_r(Y)$;

iii) $\kappa_r(X+C) = \kappa_r(X)$, $r > 1$.

另外，若 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$，$\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_m)^{\mathrm{T}}$ 为随机向量，则 $\boldsymbol{X}$ 的数学期望与方差定义为

$$E(\boldsymbol{X}) \triangleq (EX_1, \cdots, EX_n)^{\mathrm{T}},\ \mathrm{Var}(\boldsymbol{X}) \triangleq \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X})_{n\times n}.$$

$\boldsymbol{X}, \boldsymbol{Y}$ 的协方差定义为

$$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) \triangleq (\sigma_{ij})_{n\times m},\ \sigma_{ij} = \mathrm{Cov}(X_i, Y_j).$$

随机向量的期望与方差的常用公式有

i) $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) = E[(\boldsymbol{X} - E\boldsymbol{X})(\boldsymbol{Y} - E\boldsymbol{Y})^{\mathrm{T}}] = E(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}) - (E\boldsymbol{X})(E\boldsymbol{Y})^{\mathrm{T}}$;

$\mathrm{Var}(\boldsymbol{X}) = E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}) - (E\boldsymbol{X})(E\boldsymbol{X})^{\mathrm{T}}$;

ii) $\mathrm{Cov}(\boldsymbol{A}\boldsymbol{X}, \boldsymbol{B}\boldsymbol{Y}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{B}^{\mathrm{T}}$，其中 $\boldsymbol{A}, \boldsymbol{B}$ 为非随机的常数矩阵.

iii) $E\|\boldsymbol{X}\|^2 = \|E\boldsymbol{X}\|^2 + \mathrm{tr}[\mathrm{Var}(\boldsymbol{X})]$；其中 tr 表示矩阵的迹(即矩阵对角线元素之和).

iv) 设 $\boldsymbol{X}=\begin{pmatrix}X_1\\X_2\end{pmatrix}$，则 $\mathrm{Var}(\boldsymbol{X})=\begin{pmatrix}\mathrm{Var}(X_1) & \mathrm{Cov}(X_1,\ X_2)\\ \mathrm{Cov}(X_2,\ X_1) & \mathrm{Var}(X_2)\end{pmatrix}$.

以下与条件期望、方差有关的公式非常重要，今后将多次用到

$$E(X)=E_T\{E(X\mid T)\}. \tag{1.1.5}$$

$$\mathrm{Var}(X)=E_T\{\mathrm{Var}(X\mid T)\}+\mathrm{Var}_T\{E(X\mid T)\}. \tag{1.1.6}$$

下面通过两个实例说明上述两个重要公式.

**例 1.1.2** 设随机变量 $(X,\ Y)$ 的联合概率密度为

| $Y \backslash X$ | 1 | 2 |
|---|---|---|
| 0 | 1/3 | 0 |
| 1 | 1/3 | 1/3 |

求：(1) 随机变量 $X$，$Y$ 的分布函数；(2) $X\mid Y=1$ 的条件分布；(3) $E(X\mid Y=1)$，$E(X\mid Y)$ 与 $\mathrm{Var}(X\mid Y)$.

**解** (1) $X$，$Y$ 的边缘分布为

| $X$ | 0 | 1 |
|---|---|---|
| $P$ | 1/3 | 2/3 |

| $Y$ | 1 | 2 |
|---|---|---|
| $P$ | 2/3 | 1/3 |

由式(1.1.1)知随机变量 $X$，$Y$ 的分布函数分别为

$$F_X(x)=\frac{1}{3}U(x)+\frac{2}{3}U(x-1),\quad F_Y(y)=\frac{2}{3}U(y-1)+\frac{1}{3}U(y-2).$$

(2) 由 $P\{X=x_i\mid Y=1\}=\dfrac{P(X=x_i,\ Y=1)}{P\{Y=1\}}$，得

| $X\mid Y=1$ | 0 | 1 |
|---|---|---|
| $P$ | 1/2 | 1/2 |

(3) $E(X\mid Y=1)=1/2$，$\mathrm{Var}(X\mid Y=1)=1/4$，同理，$E(X\mid Y=2)=1$，$\mathrm{Var}(X\mid Y=1)=0$，从而有

| $E(X\mid Y)$ | 1/2 | 1 |
|---|---|---|
| $P$ | 2/3 | 1/3 |

| $\mathrm{Var}(X\mid Y)$ | 1/4 | 0 |
|---|---|---|
| $P$ | 2/3 | 1/3 |

由上述分布可得到：$E\{\mathrm{Var}(X\mid Y)\}=1/6$；$\mathrm{Var}\{E(X\mid Y)\}=1/18$. 有兴趣的读者可以验证式(1.1.5)、式(1.1.6)对本例的正确性.

**例 1.1.3** 设随机变量 $(X,\ Y)\sim N(\mu_1,\ \sigma_1^2,\ \mu_2,\ \sigma_2^2,\ \rho)$，求

$E(X\mid Y)$，$\mathrm{Var}(X\mid Y)$.

**解**　$X$，$Y$ 的边缘分布为 $X\sim N(\mu_1,\ \sigma_1^2)$，$Y\sim N(\mu_2,\ \sigma_2^2)$，$X\mid Y=y$的条件密度函数为

$$f_{X\mid Y=y}(x\mid y)=f(x,\ y)/f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_1\cdot\sqrt{1-\rho^2}}\cdot$$

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

$X\mid Y=y\sim N\left[\mu_1+\rho\dfrac{\sigma_1}{\sigma_2}(y-\mu_2),\ \sigma_1^2(1-\rho^2)\right]$，所以

$$E(X\mid Y=y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\ \mathrm{Var}(X\mid Y=y)=\sigma_1^2(1-\rho^2),$$

$$E(X\mid Y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(Y-\mu_2),\ \mathrm{Var}(X\mid Y)=\sigma_1^2(1-\rho^2).$$

读者可以验证式(1.1.5)和式(1.1.6)对本例的正确性.

## 习题1.1

1. 某地电视台想了解某电视栏目(如每晚9:00—9:30的节目)在该地区的收视率，于是委托一家市场咨询公司进行一次电话调查，问:

(1) 该项目研究的总体是什么?

(2) 该项目研究的样本是什么?

2. 为了了解统计学专业本科毕业生的就业情况，某研究机构调查了某地区30名2009年毕业的统计学专业本科生实习期满后的月薪，问:

(1) 该项目研究的总体是什么?

(2) 该项目研究的样本是什么?

(3) 样本容量是多少?

3. $X_1$，$X_2$，…，$X_n$ 为来自总体$X\sim B(1,\ p)$分布的简单随机样本，$p$ 为未知参数，

(1) 写出 $\boldsymbol{X}=(X_1,\ X_2,\ \cdots,\ X_n)^{\mathrm{T}}$ 的样本空间和概率分布.

(2) 样本均值 $\overline{X}\triangleq\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$，样本方差 $S^2\triangleq\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$，

求 $E\overline{X}$, $\mathrm{Var}(\overline{X})$, $ES^2$.

4.* 设 $x_p$ 为 $F(x)$ 的 $p$ 分位数，证明：

(1) $F(x')<p$ 的充要条件为 $x'<x_p$；$F(x')\geqslant p$ 的充要条件为 $x'\geqslant x_p$；是否有 $F(x')>p$ 的充要条件为 $x'>x_p$？

(2) $F(x_p-0)\leqslant p\leqslant F(x_p)$；若 $x_p$ 为 $F(x)$ 的连续点，则 $F(x_p)=p$.

(3) 若 $F(x'-0)>p$，则 $x'>x_p$.

5.* 设 $\alpha$ 分位函数定义为

$$\begin{aligned}\rho_\alpha(t)&=(\alpha-I\{t<0\})t\\&=|t|\left[\alpha I\{t>0\}+(1-\alpha)I\{t<0\}\right],\ 0<\alpha<1,\ t\in\mathbf{R}.\end{aligned}$$

(1) 设随机变量 $X$ 的概率密度函数为 $f(x;\ \alpha,\ \theta)=\alpha(1-\alpha)\exp[-\rho_\alpha(x-\theta)]$，证明：$X$ 的 $\alpha$ 分位数为 $\theta$；

(2) 设 $Y$ 为连续型随机变量，$g(\mu)=E[\rho_\alpha(Y-\mu)]$ 关于一切 $\mu$ 存在，则当 $\mu=y_\alpha$ 时 $g(\mu)$ 达到最小值，其中 $y_\alpha$ 为 $Y$ 的 $\alpha$ 分位数.

6. 设 $T(X)$ 为随机变量 $X$ 的可测函数，$E[T(X)-\theta]^2$ 在 $\theta\in[a,\ b]$ 上存在，

$$S(X)=\begin{cases}a, & T(x)<a,\\ T(X), & a\leqslant T(x)\leqslant b,\\ b, & T(x)>b.\end{cases}$$

证明：$E[S(X)-\theta]^2\leqslant E[T(X)-\theta]^2$.

7. 证明：$\mathrm{Var}(X)=E_T[\mathrm{Var}(X|T)]+\mathrm{Var}_T[E(X|T)]$.

8. 随机变量 $X$ 的密度函数 $f(x)$ 满足 $f(\xi_0+x)=f(\xi_0-x)$，$\xi_0$ 为某一常数，则称 $X$ 的分布关于 $\xi_0$ 对称，证明：

(1) $X$ 的分布关于原点对称的充要条件为 $X$ 与 $-X$ 同分布；

(2) $X$ 的分布关于 $\xi_0$ 对称的充要条件为 $X-\xi_0$ 的分布关于原点对称；

(3) 若 $X$ 的分布关于 $\xi_0$ 对称，则 $EX=\xi_0$，$E(X-\xi_0)^{2k-1}=0$，其中 $k$ 为正整数；

(4) 设 $X_1$, …, $X_n$ 为 i. i. d. 的样本，$X_1$ 的分布关于 $\xi_0$ 对称；则其样本均值 $\overline{X}$ 与样本方差 $S^2$ 不相关，即 $\mathrm{Cov}(\overline{X},\ S^2)=0$.

9. 设随机变量 $X$ 的 $r$ 阶累积量为 $\kappa_r$，证明：

(1) 若 $X$, $Y$ 独立，则有 $\kappa_r(X+Y)=\kappa_r(X)+\kappa_r(Y)$；

(2) $\kappa_r(X+C)=\kappa_r(X)$, $(r>1)$.

10. 设 $X$ 为连续型正值随机变量，其分布函数为 $F(t)$，$F'(t)=f(t)$，记 $h(t)=f(t)/[1-F(t)]$，通常称 $h(t)$ 为危险率函数. 证明：

(1) $h(t)$ 表示 $X$ 大于 $t$，但不超过 $t+\Delta t$ 的相对概率，或危险率，则

$$h(t)=\lim_{\Delta t\to 0}\frac{P(t<X\leqslant t+\Delta t \mid X>t)}{\Delta t}.$$

(2) 危险率函数 $h(t)$ 与密度函数 $f(t)$ 有以下一一对应的关系：

$$f(t)=h(t)\exp[-H(t)];\ H(t)=\int_0^t h(x)\,\mathrm{d}x.$$

11. 设随机变量 $X$ 服从泊松分布 $P(\lambda)$，并已知事件 $\{X=0\}$ 不可能发生，求此时 $X$ 的分布(截尾泊松分布)及其期望和方差.

## 1.2 经验分布函数与统计图示

### 1.2.1 经验分布函数

根据样本观测值推断总体的分布函数是数理统计要解决的基本问题之一. 为此引入经验分布函数的概念.

**定义 1.2.1** 设总体 $X$ 的分布函数为 $F(x)$，样本 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 的观测值为 $x_1,x_2,\cdots,x_n$，将 $x_1,x_2,\cdots,x_n$ 从小到大排序为 $x_{(1)},x_{(2)},\cdots,x_{(n)}$，令

$$F_n(x)=\begin{cases}0, & x<x_{(1)},\\ 1/n, & x_{(1)}\leqslant x<x_{(2)},\\ 2/n, & x_{(2)}\leqslant x<x_{(3)},\\ & \vdots\\ k/n, & x_{(k)}\leqslant x<x_{(k+1)},\\ & \vdots\\ 1, & x\geqslant x_{(n)},\end{cases}$$

则称 $F_n(x)$ 为**经验分布函数**(empirical cumulative distribution function, ecdf).

**例 1.2.1** 设总体 $X$ 的一个样本 $(X_1,X_2,\cdots,X_8)$ 的一组观测值为 3，−1，2，2.5，3，0，4，2.5. 将它们从小到大排列为

$$-1<0<2<2.5=2.5<3=3<4$$