GONGYE SHENGCHAN ANQUAN JISHU SHOUC

工业生产安全技术手册

主　编　孙桂林　任　萍
副主编　赵家振　段绪华　高永新

中国劳动社会保障出版社

工业生产安全技术手册

主　编　孙桂林　任　萍
副主编　赵家振　段绪华　高永新

中国劳动社会保障出版社

图书在版编目(CIP)数据

工业生产安全技术手册/孙桂林,任萍主编 .—北京:中国劳动社会保障出版社,1999.4
ISBN 7-5045-2282-1

Ⅰ.工…
Ⅱ.①孙…　②任…
Ⅲ.工业生产-安全技术-手册
Ⅳ.X931-62

中国版本图书馆 CIP 数据核字(98)第 38110 号

中国劳动社会保障出版社出版发行
(北京市惠新东街 1 号　邮政编码:100029)
出 版 人 :唐云岐
*
中国青年出版社印刷厂印刷　新华书店经销
787×1092 毫米　16 开本　41.75 印张　1037 千字
2000 年 6 月第 1 版　　2000 年 6 月第 1 次印刷
印数:2000 册
定价:66.80 元

一、基 础 理 论

1

概率论与数理统计

1.1 概　率　论

1.1.1 必然与偶然

在一定的试验条件下，必然发生或者说必然出现的现象，称为必然现象。必然现象是人们可准确预测的现象。

在一定的试验条件下，可能出现也可能不出现的现象，称为偶然现象。偶然现象是人们不可准确预测的现象。

1.1.2 随机事件

称这样的试验为随机试验：试验会出现多种可能的结果，每一种结果都是一种偶然现象。

称随机试验所出现的结果为随机事件（以下简称为事件）。

1.1.3 事件　样本空间

随机试验所出现的每个基本的不能再细分割的结果，称为基本事件。将每一基本事件抽象为一个元素 ω，全部这种元素组成的集合

$$\Omega = \{\omega \mid \omega \text{ 是随机试验的基本事件}\}$$

称为随机试验的样本空间。

由基本事件组成的事件，即样本空间 Ω 的子集，称为复合事件。

由全体基本事件组成的事件，即样本空间 Ω，亦即随机试验必然出现的现象，称为必然事件；称随机试验不可能出现的现象，亦即空集 $\varnothing$，称为不可能事件。

1.2 随机变量及其概率分布

1.2.1 概率空间

设 Ω 是随机试验的样本空间，F 是这个随机试验的全体随机事件组成的集合，由于每个

随机事件都是样本空间 Ω 的一个子集,因此集合 F 中的元素是 Ω 的子集,并且要求 F 具有以下三个性质:

· $\Omega \in F$;

·若 $A \in F$,则 $\bar{A} \in F$;

·若 $A_i \in F, i = 1,2,\cdots,n,\cdots$,则 $\sum\limits_i A_i \in F$.

称具有上述三条性质的 F 为 σ - 代数。P 是随机事件的概率,具有以下三个性质(公理):

·对任意 $A \in F, 0 \leqslant P(A) \leqslant 1$;

· $P(\Omega) = 1$;

·若 $A_1, A_2, \cdots, A_n, \cdots$ 是一系列互不相容事件,则 $P(\sum\limits_i A_i) = \sum\limits_i P(A_i)$.

称上述的三位一体 (Ω, F, P) 为概率空间,它是概率论的基础。

1.2.2 随机变量及概率分布函数

观察每天 9:00 ~ 10:00 这一个小时内来到服务台前的顾客人数 X,则 X 是一个变量。它可能取的数值为 $0,1,2,\cdots,n,\cdots$ 取每个数值都是随机的,"$X = k$"表示"接受服务的顾客人数为 k"这个基本事件。

观察即将出生婴儿的性别,其样本空间为

$$\Omega = \{男婴、女婴\}$$

引入一个变量 X,定义

$$X(\omega) = \begin{cases} 1 & 当\ \omega\ 为男婴时 \\ 0 & 当\ \omega\ 为女婴时 \end{cases}$$

这样,"$X = 1$"便表示"出生的婴儿是男婴","$X = 0$"便表示"出生的婴儿是女婴"这两个基本事件,显得很方便。

所谓随机变量是这样一类变量 X,它不同于我们在微积分学中所熟悉的变量,它可能取一些离散的数值, $x_1, x_2, \cdots x_n, \cdots$;它也可能取实数上某一区间内的一切数值,取每个数值都是随机的,取每一数值都表示一个基本事件,称这样的变量 X 为随机变量。

以上是关于随机变量的描述性定义,其严格定义如下:

概率空间 (Ω, F, P) 上的随机变量 X 是定义在 Ω 上的一个实值函数 $X(\omega), \omega \in \Omega$,并且对任意的实数 x,"$X(\omega) \leqslant x$"表示一个随机事件,即

$$A = \{\omega \mid X(\omega) \leqslant x\} \in F.$$

有了关于随机变量的严格定义后,便有了关于随机变量的概率分布函数的严格定义:

设 X 是概率空间 (Ω, F, P) 上的一个随机变量,称定义在 $(-\infty, +\infty)$ 上的实值函数 $F(x)$:

$$F(x) = P(\text{“}X \leqslant x\text{”}), x \in (-\infty, +\infty)$$

为随机变量 X 的概率分布函数(简称分布函数)。

根据定义,随机变量的分布函数 $F(x)$ 具有以下性质:

· $0 \leqslant F(x) \leqslant 1$

·若 $x_1 < x_2$,则 $F(x_1) \leqslant F(x_2)$ (不减函数)

· $\lim\limits_{x \to x_0^+} F(x) = F(x_0)$ (右连续)

· $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$.

1.2.3 离散随机变量及其概率分布

称只取有限个或可数个实数值的随机变量为离散型随机变量。

设离散型随机变量 X 可能取的全部数值为 $x_1, x_2, \cdots x_n, \cdots$，记

$$p_i = P(“X = x_i”), i = 1,2,\cdots,n,\cdots$$

称二维数组 $\{(x_i, p_i); i = 1,2,\cdots n,\cdots\}$ 为离散随机变量 X 的概率分布列，并常以表格形式列出：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

根据分布函数的定义，离散随机变量 X 的分布函数 $F(x)$ 的特征是右连续的阶梯函数：

$$F(x) = \sum_{x_i \leqslant x} p_i, \qquad 0 \leqslant p_i \leqslant 1, \qquad \sum_i p_i = 1.$$

1.2.4 四种常见的离散型随机变量

(1)两点分布

用来描述只有两个基本事件或者只考虑两个相互对立事件的随机变量，例如，掷一枚硬币；观察即将出生婴儿的性别；从一批只有正品或次品之分的产品中任取一件进行质量检验；观察一射手的射击，只关心命中或未命中目标，等等。如同计算机中用 0,1 表示两种对立状态一样。这种随机变量 X，只取 0,1 两个数值，并记

$$P(“X = 1”) = p, P(“X = 0”) = q, (q = 1 - p)$$

称这样的离散随机变量 X 为服从两点分布的随机变量。

(2)二项分布

所谓服从二项分布的离散型随机变量，就是用来描述独立试验序列概型的随机变量，即随机变量 X 表示在 n 次重复试验中事件 A 出现的次数。记 $P(A) = p$，则有

$$p_k = P(“X = k”) = C_n^k p^k q^{n-k}, q = 1 - p, k = 0,1,2,\cdots n。$$

例如，从次品率为 p 的一批产品中，有放回地(即每次任取一件，检验后放回)抽检 n 次。用 X 表示抽得的次品件数，则 X 就是服从二项分布的随机变量，简记作 $X \sim B(n,p)$。

(3)超几何分布

N 件产品中有 M 件正品，$N - M$ 件次品，从中任取 n 件，用 X 表示取出产品中正品的数目，则有

$$P(“X = k”) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 0,1,2,\cdots \min(M,n)$$

称用来描述上述试验概型的随机变量 X 是服从超几何分布的随机变量。

(4)泊松分布

放射性物质在某一段时间内放射出的粒子数 X；在某一段时间内来到服务台前的顾客人数 X；在某一段时间内打入某单位的电话次数 X；一段布匹或一个铸件上出现的疵点数 X，等等。这类随机变量 X，其概率分布列为

$$P(“X=k”)=\frac{\lambda^k}{k!}e^{-\lambda},\lambda>0,k=0,1,2,\cdots n,\cdots$$

称之为服从泊松分布的随机变量,简记作 $X\sim P(\lambda)$。

以上四种常见的随机变量有以下关系:

①$B(1,p)$即两点分布

②服从超几何分布的随机变量 X,当 N 很大并且 n 相对于 N 又是很小时,则有

$$P(“X=k”)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}\approx C_n^kp^kq^{n-k},\quad p=\frac{M}{N}$$

即 X 近似服从二项分布 $B(n,\frac{M}{N})$.

③服从二项分布的随机变量 X,当 n 很大并且 p 很小时,有

$$P(“X=k”)=C_n^kp^kq^{n-k}\approx\frac{\lambda^k}{k!}e^{-\lambda},\quad \lambda=np$$

即有　$B(n,p)\approx P(np)$.

1.2.5 连续型随机变量　概率密度函数

对于取值在一个实数区间上的随机变量 X,如果存在一个非负实值函数 $p(x),x\in(-\infty,+\infty)$,使之随机变量 X 的分布函数 $F(x)$可以写成

$$F(x)=\int_{-\infty}^{x}p(t)dt,\qquad x\in(-\infty,+\infty)$$

则称随机变量 X 为连续型随机变量,称 $p(x)$为其概率密度函数,并且具有下述性质:

- $p(x)\geqslant 0,\quad x\in(-\infty,+\infty)$;
- $\int_{-\infty}^{+\infty}p(x)dx=1$.

连续型随机变量 X 的分布函数 $F(x)$与其概率密度函数 $p(x)$之间有如下关系:

(1)$F(x)=\int_{-\infty}^{x}p(t)dt$,$x\in(-\infty,+\infty)$,因此 $F(x)$是连续函数;

(2)在 $p(x)$的连续点处,有$\frac{d}{dx}F(x)=p(x)$,即

$$p(x)=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=\lim_{\Delta x\to 0^+}\frac{p(x<X\leqslant x+\Delta x)}{\Delta x}\cdots\text{密度特征}$$

(3)$P(a<X\leqslant b)=F(b)-F(a)=\int_a^bp(x)dx$

对于连续型随机变量 X,还有如下特征:

①$P(X=a)=0$;

②$P(a\leqslant X\leqslant b)=P(a<X<b)=P(a\leqslant X<b)=P(a<X\leqslant b)$.

1.2.6 常见的连续型随机变量

(1)均匀分布

以

$$p(x)=\begin{cases}\dfrac{1}{b-a} & a\leqslant x\leqslant b, a<b\\ 0 & \text{其他}\end{cases}$$

为概率密度函数的随机变量X称为服从$[a,b]$区间上的均匀分布，记作$X\sim U[a,b]$（图1-1）。

例如，已知某公交线路上每隔10 min发一趟车，那么人们来到公共汽车站候车的时间X便服从[0,10]区间上的均匀分布。

(2)负指数分布

以

$$p(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\leqslant 0\end{cases}$$

$$\lambda>0 \text{ 常数}$$

为概率密度函数的随机变量X称为服从参数为λ的负指数分布（图1-2）。

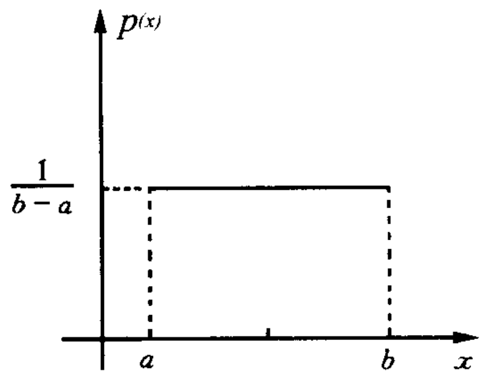

图1-1

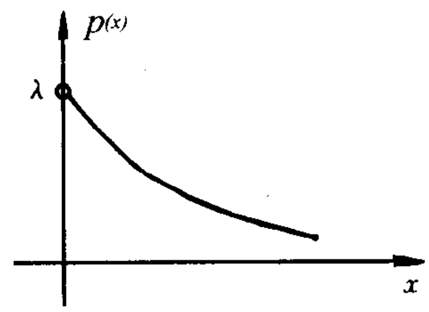

图1-2

例如，某些电子元件的使用寿命就近似地服从负指数分布。

(3)正态分布（高斯分布）

以

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

$$\sigma>0, -\infty<x<+\infty.$$

为概率密度函数的随机变量X，称为服从参数为μ和σ的正态分布，也称为服从高斯分布（图1-3），记作$X\sim N(\mu,\sigma^2)$。特别地，当$\mu=0,\sigma=1$时，称为标准正态分布，记作$X\sim N(0,1)$。

若$X\sim N(\mu,\sigma^2)$，则有

$$P(\mu-2\sigma<X<\mu+2\sigma)=0.954\ 4$$

$$P(\mu-3\sigma<X<\mu+3\sigma)=0.997\ 4$$

这表明X的取值基本上落在$(\mu-2\sigma,\mu+2\sigma)$区间内，几乎不在$(\mu-3\sigma,\mu+3\sigma)$区间外。

(4)柯西分布

以

$$p(x)=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2}$$

$$\lambda > 0,\ -\infty < x < +\infty$$

为概率密度函数的随机变量 X,称为服从参数为 μ 及 λ 的柯西分布,记作 $X \sim C(\mu,\lambda)$(图 1-4)。

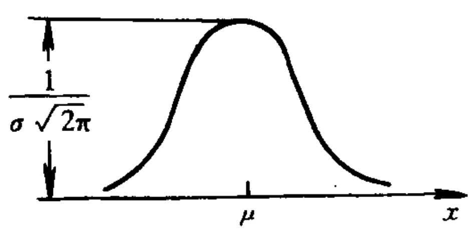

图 1-3

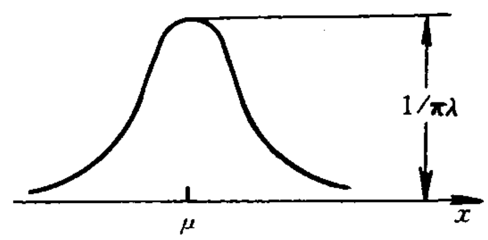

图 1-4

(5)对数正态分布

以

$$p(x)=\begin{cases} \dfrac{1}{\sigma x\ \sqrt{2\pi}}\exp\left\{-\dfrac{(\ln x-a)^2}{2\sigma^2}\right\} & (x>0), \\ 0 & (x\leqslant 0), \end{cases}$$

$$\sigma > 0, a \text{ 为实数}$$

为概率密度函数的随机变量 X 称为服从参数为 a,σ 的对数正态分布(图 1-5)。

(6)伽马分布

以

$$p(x)=\begin{cases} \dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} & (x>0), \\ 0 & (x\leqslant 0), \end{cases}$$

$$r>0, \lambda>0 \text{ 为常数}$$

为概率密度函数的随机变量 X 称为服从参数为 r,λ 的伽马(Γ)分布,记作 $X\sim\Gamma(r,\lambda)$(图 1-6)。

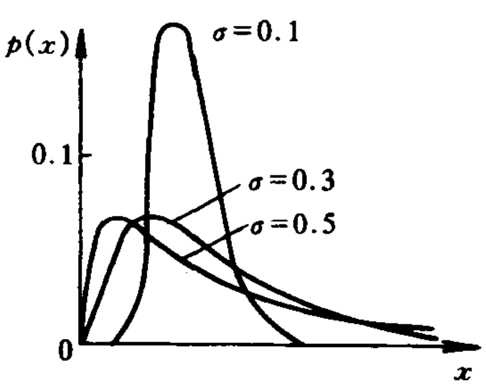

图 1-5

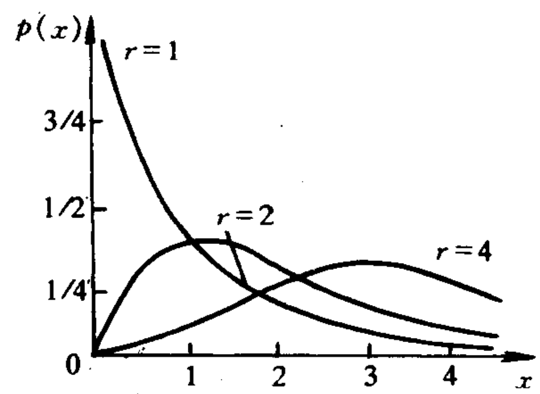

图 1-6

(7)倍塔分布

以

$$p(x)=\begin{cases}\dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1} & (0<x<1),\\ 0 & (\text{其他}),\end{cases}$$

$$p>0,q>0\ \text{为常数}$$

为概率密度函数的随机变量 X 称为服从参数为 p,q 的倍塔分布。这里 $\Gamma(p),\Gamma(q),\Gamma(p+q)$ 为伽马函数值(图 1-7)。

(8)威布尔分布

以

$$p(x)=\begin{cases}a\lambda x^{a-1}\exp(-\lambda x^{a}) & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$$

$$\lambda>0,a>0\ \text{为常数}$$

为概率密度函数的随机变量 X 称为服从参数为 a,λ 的威布尔分布,记作 $W(a,\lambda)$(图 1-8)。

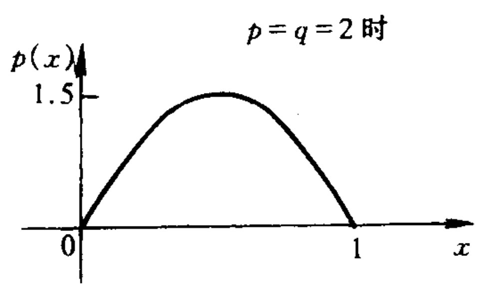

图 1-7

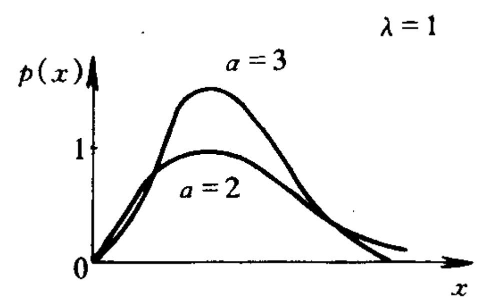

图 1-8

(9)拉普拉斯分布

以

$$p(x)=\frac{1}{2\lambda}\exp\left\{-\frac{|x-\mu|}{\lambda}\right\},x\ \text{为实数}$$

$$\lambda>0,\mu\ \text{为常数}$$

为概率密度函数的随机变量 X 称为服从参数为 μ,λ 的拉普拉斯分布,记作 $L(\mu,\lambda)$(图1-9)。

(10)χ^2 分布

以

$$p(x)=\begin{cases}\dfrac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$$

$$n\ \text{为正整数}$$

为概率密度函数的随机变量 X 称为服从自由度为 n 的 χ^2 分布,记作 $X\sim\chi^2(n)$(图 1-10)。

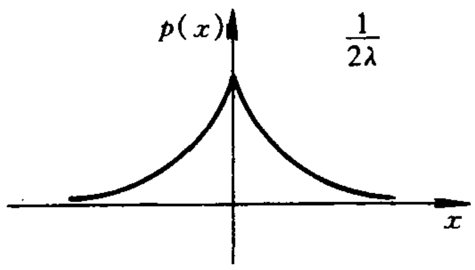

图 1－9

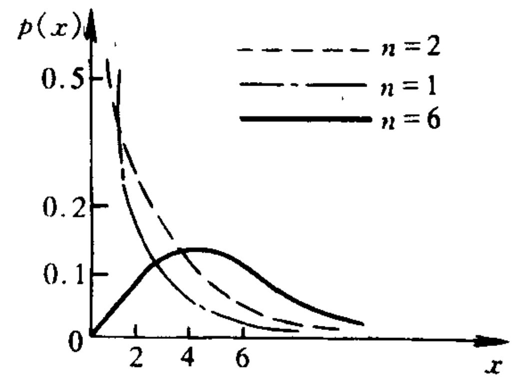

图 1－10

(11)学生 t 分布

以

$$p(x)=\frac{1}{\sqrt{n\pi}}\ \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

为概率密度函数的随机变量 X 称为服从自由度为 n 的学生 t 分布，记作 $X\sim t(n)$(图 1－11)。

(12)F 分布

以

$$p(x)=\begin{cases}\dfrac{x^{\frac{n_1}{2}-1}}{B(n_1,n_2)}n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}\times(n_2+n_1x)^{-\frac{n_1+n_2}{2}} & (x>0),\\ 0 & (x\leqslant 0),\end{cases}$$

n_1,n_2 为正整数

为概率密度函数的随机变量 X 称为服从自由度为 n_1,n_2 的 F 分布，记作 $X\sim F(n_1,n_2)$(图 1－12)。

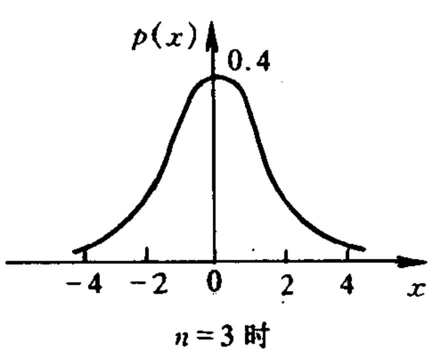

图 1－11

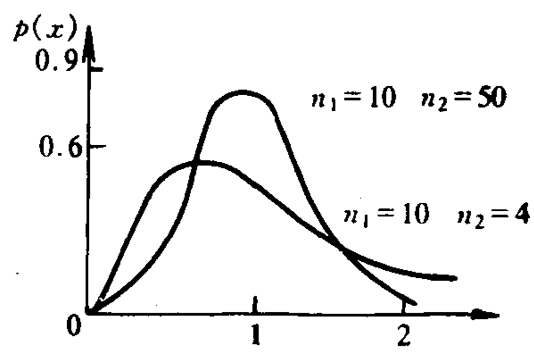

图 1－12

1.3 随机变量的数字特征

1.3.1 数学期望

设 X 是一个离散型随机变量,其概率分布列为 $P(“X=x_i”)=p_i, i=1,2,\cdots, g(x)$是一个单值函数,称$\sum\limits_i g(x_i)p_i$(当其为无穷级数时,要求它绝对收敛)为由随机变量 X 复合成的随机变量$g(X)$的数学期望,记作 $E[g(X)]$,即

$$E[g(X)]\equiv\sum_i g(x_i)p_i$$

设 X 是一个连续型随机变量,其概率密度函数为 $p(x)$,$g(x)$ 是一个单值函数,若 $\int_{-\infty}^{+\infty}|g(x)|p(x)\mathrm{d}x$收敛,则称$\int_{-\infty}^{+\infty}g(x)p(x)\mathrm{d}x$为由连续型随机变量$X$复合成的随机变量$g(X)$的数学期望,记作 $E[g(X)]$,即

$$E[g(X)]\equiv\int_{-\infty}^{+\infty}g(x)p(x)\mathrm{d}x$$

特别地,随机变量 X 本身的数学期望$E(X)$又称为随机变量 X 的均值。它是随机变量 X 的第一个重要数字特征,它表示了随机变量 X 取值的概率加权平均值。

数学期望的性质

- $E(c)=c$ (c 为常数)
- $E(kX)=kE(X)$ (k 为常数)
- $E(X+c)=E(X)+c$ (c 为常数)
- $E(kX+c)=kE(X)+c$ (k,c 为常数)
- $X_1,X_2,\cdots X_n$ 为 n 个随机变量,则有

  $E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$
- $X_1,X_2,\cdots X_n$ 为 n 个相互独立的随机变量,则有

  $E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$

1.3.2 方差与标准差

称随机变量$[X-E(X)]^2$ 的数学期望 $E[X-E(X)]^2$ 为随机变量 X 的方差,记作 $Var(X)$ 或记作 $D(X)$,即

$$Var(X)\text{或 } D(X)\equiv E[X-E(X)]^2$$

设 X 是一个离散型随机变量,其分布列为 $P(“X=x_i”)=p_i, i=1,2,\cdots$,数学期望 $E(X)$ 存在。则

$$D(X)=\sum_i[x_i-E(X)]^2p_i$$

设 X 是一个连续型随机变量,其概率密度为 $p(x)$,数学期望 $E(X)$存在,则

$$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2p(x)\mathrm{d}x$$

称$\sqrt{D(X)}$为随机变量 X 的标准差。

方差的性质：

- $D(X)\geqslant 0$
- $D(X)=E(X^2)-[E(X)]^2$
- $D(c)=0$ （c 为常数）
- $D(kX)=k^2D(X)$ （k 为常数）
- $D(kX+c)=k^2D(X)$ （k,c 为常数）
- $X_1,X_2,\cdots X_n$ 是相互独立的随机变量，则有

$$D(X_1+X_2+\cdots+X_n)=D(X_1)+D(X_2)+\cdots+D(X_n)$$

1.3.3 常见随机变量的均值与方差

(1) X 是服从两点分布的随机变量，则

$$E(X)=p,D(X)=pq$$

(2) X 是服从二项分布的随机变量，则

$$E(X)=np,D(X)=npq$$

(3) X 是服从超几何分布的随机变量，则

$$E(X)=\frac{nM}{N},D(X)=\frac{nM}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$$

(4) X 是服从泊松分布的随机变量，则

$$E(X)=\lambda,D(X)=\lambda$$

(5) X 是服从均匀分布的随机变量，则

$$E(X)=\frac{1}{2}(a+b),D(X)=\frac{1}{12}(b-a)^2$$

(6) X 是服从负指数分布的随机变量，则

$$E(X)=\frac{1}{\lambda},D(X)=\frac{1}{\lambda^2}$$

(7) X 是服从正态分布的随机变量，则

$$E(X)=\mu,D(X)=\sigma^2$$

(8) X 是服从柯西分布的随机变量，则

$E(X)$不存在，$D(X)$不存在

(9) X 是服从对数正态分布的随机变量，则

$$E(X)=\exp\left(a+\frac{\sigma^2}{2}\right),D(X)=\exp(2a+\sigma^2)(\exp\sigma^2-1)$$

(10) X 是服从伽马分布的随机变量，则

$$E(X)=\frac{\gamma}{\lambda},D(X)=\frac{\gamma}{\lambda^2}$$

(11) X 是服从倍塔分布的随机变量，则

$$E(X)=\frac{p}{p+q},D(X)=\frac{pq}{(p+q)^2(p+q+1)}$$

(12)X 是服从威布尔分布的随机变量,则

$$E(X)=\lambda^{-\frac{1}{2}}\Gamma\left(1+\frac{1}{a}\right),D(X)=\lambda^{-\frac{2}{a}}\left\{\Gamma\left(1+\frac{2}{a}\right)-\left[\Gamma\left(1+\frac{1}{a}\right)\right]^2\right\}$$

(13)X 是服从拉普拉斯分布的随机变量,则

$$E(X)=\mu,D(X)=2\lambda^2$$

(14)X 是服从 χ^2 分布的随机变量,则

$$E(X)=n,D(X)=2n$$

(15)X 是服从学生 t 分布的随机变量,则

$$E(X)=0\quad(n>1),D(X)=\frac{n}{n-2}\quad(n>2)$$

(16)X 是服从 F 分布的随机变量,则

$$E(X)=\frac{n_2}{n_2-2}\quad(n_2>2),D(X)=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}\quad(n_2>4)$$

1.4 数 理 统 计

1.4.1 总体与样本

在统计学中,把所研究对象的全体称为总体,构成总体的每个成员称为个体。

当总体中包含的个体数量有限时,称为有限总体;否则称为无限总体。

当个体数量很大时,或者调查时采用的手段对个体具有破坏性时,对总体的研究就要采取抽出相对于总体来说数量很小的一组个体作为样品(本),通过对样品的研究来推断总体的情况。

在数理统计学中,人们研究的是表征总体性能并且有随机性的一个或若干个数量指标。把表征总体性能的数量指标——随机变量或随机向量——称为总体,例如称呼总体 X。

从总体中抽取一个样品对其性能指标 X 的观(检)测,便是概率论中的一次随机试验,抽取 n 个样品进行观(检)测,便是做 n 次随机试验。用以描述这 n 次抽样结果的 n 个随机变量 $(X_1,X_2\cdots X_n)$,称为从总体中抽取的容量为 n 的样本(又称为子样)。

在数理统计学的术语中,将随机变量 X 的分布称为"总体 X 的分布"。总体 X 的分布,常常是未知的,或者只知其分布类型,其中一些分布参数是未知的。对样品进行观测便得到 n 个随机变量 $X_1,X_2\cdots X_n$ 的一组观测数据 $x_1,x_2\cdots x_n$,称之为样本值。根据样本值推断总体 X 的分布或者估计其中的未知参数,便是数理统计学要解决的主要问题。

从总体 X 中抽取容量为 n 的样本 $X_1,X_2\cdots X_n$,如果随机变量序列 $X_1,X_2\cdots X_n$ 是相互独立并且每个 X_i 的分布都与总体 X 的分布相同(以下简称 $X_1,X_2\cdots X_n$ 独立同分布于 X),则称这样的样本为简单随机样本。

对于有限总体,进行 n 次抽样,每次从总体中任取一个个体,观测后再放回去,称这种抽

样方式为有放回地随机抽样;观测后不再放回去,称这种抽样方式为无放回地随机抽样。

显然,采用有放回地随机抽样,得到的容量为 n 的样本便是简单随机样本。对于个体数量 N 很大的总体,当样本容量 n 相对于总体数量 N 很小时(即 $N \gg n$ 时),无放回抽样得到的样本可近似地看作为简单随机样本。

1.4.2 统计量

样本是总体的代表和反映。在抽取样本并获得样本值以后,不能直接从样本值推断出关于总体的内在性质,而需要对样本值进行一番加工和提炼,把样本中包含的信息集中起来。其做法是,针对不同问题构造出关于样本的某种函数,这种函数在数理统计学中称为统计量。下面给出统计量的严格定义。

设$(X_1, X_2 \cdots X_n)$是取自总体 X 的样本,若随机向量$(X_1, X_2 \cdots X_n)$的函数:

$$T = T(X_1, X_2 \cdots X_n)$$

也是一个随机变量,并且 T 的表达式中不包含总体分布中的任何未知参数,则称 T 为统计量。

常用的统计量及其概率分布如下:

(1)样本矩及样本矩的函数

设$(X_1, X_2 \cdots X_n)$是取自总体 X 的样本。称

$$A_k \equiv \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1,2,\cdots$$

为样本的 k 阶原点矩。特别地将一阶原点矩记作 $\bar{X}$,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

称为样本的均值。称:

$$B_k \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k, \quad k = 1,2,\cdots$$

为样本的 k 阶中心矩。特别地将$\frac{n}{n-1}\beta_2$ 记作 S^2,

$$S^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

称为样本的方差,以区别于样本的 2 阶中心矩。称 $S = \sqrt{S^2}$为样本的标准差;称 $S/\bar{X}$ 为样本的变异系数;称 $B_3/(\sqrt{B^2})^3$ 为样本偏度;称 $B_4/B_2^2 - 3$ 为样本峰度。

(2)取自正态总体的简单随机样本的 n 个常用统计量及其分布

设$(X_1, X_2 \cdots X_n)$是取自正态总体 $X \sim N(\mu, \sigma^2)$的简单随机样本。则有

①$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

②$\frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)\cdots$

③$\bar{X}$ 与$\sum_{i=1}^{n}(X_i - \bar{X})^2$ 相互独立

④$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$