元).故由题意知

$$R=\begin{cases}30X & \text{若 } X\leqslant Y\\ 30Y+10(X-Y) & \text{若 } X>Y\end{cases}$$

于是当 $20\leqslant y\leqslant 30$ 时,有

$$E(R\mid Y=y)=\int_{10}^{20}30x\cdot\frac{1}{10}\mathrm{d}x=450(\text{万元})$$

当 $10\leqslant y<20$ 时,有

$$\begin{aligned}E(R\mid Y=y)&=\int_{10}^{y}30x\cdot\frac{1}{10}\mathrm{d}x+\int_{y}^{20}[30y+10(x-y)]\frac{1}{10}\mathrm{d}x\\&=50+40y-y^{2}\end{aligned}$$

从(5.6)知

$$\begin{aligned}E(R)&=\int_{10}^{30}E(R\mid Y=y)p_{Y}(y)\mathrm{d}y\\&=\int_{10}^{20}(50+40y-y^{2})\frac{1}{20}\mathrm{d}y+\int_{20}^{30}450\cdot\frac{1}{20}\mathrm{d}y\\&\approx 433(\text{万元})\end{aligned}$$

即该工厂每月的平均利润约为 433 万元.

条件分布和条件期望的概念可以推广到两个随机向量的情形.设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_m)$和 $\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_n)$分别是 m 维和 n 维的随机向量,我们也可讨论 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)$的条件下 $\boldsymbol{X}$ 的条件分布函数.

设对任何 $\varepsilon>0$,$P(y_1-\varepsilon<Y_1\leqslant y_1+\varepsilon,y_2-\varepsilon<Y_2\leqslant y_2+\varepsilon,\cdots,y_n-\varepsilon<Y_n\leqslant y_n+\varepsilon)>0$.如果下列极限

$$\lim_{\varepsilon\to 0}P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_m\leqslant x_m\mid y_1-\varepsilon<Y_1\leqslant y_1+\varepsilon,y_2-\varepsilon<Y_2\leqslant y_2+\varepsilon,\cdots,y_n-\varepsilon<Y_n\leqslant y_n+\varepsilon)$$

存在,则称这个极限为 $Y=(y_1,y_2,\cdots,y_n)$的条件下 $\boldsymbol{X}$ 的条件分布函数,记作 $F_{X|Y}(x_1,x_2,\cdots,x_n\mid y_1,y_2,\cdots,y_n)$.可以证明,在相当广泛的条件下,若$(\boldsymbol{X},\boldsymbol{Y})$有联合密度 $p(x_1,x_2,\cdots,x_m,y_1,y_2,\cdots,y_n)$,则

$$F_{X|Y}(x_1, x_2, \cdots, x_m \mid y_1, y_2, \cdots, y_n) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots$$

$$\int_{-\infty}^{x_m}\frac{p(u_1,u_2,\cdots,u_m,y_1,y_2,\cdots,y_n)}{p_Y(y_1,y_2,\cdots,y_n)}\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_m$$

很自然称这里的被积函数为 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)$的条件下 $\boldsymbol{X}$ 的条件分布密度.固定 $y=(y_1,y_2,\cdots,y_n)$.不难推知在 $\boldsymbol{Y}=y$ 的条件下 X_i 的条件分布密度为

$$p_i(u_i\mid y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{p(u_1,u_2,\cdots,u_m,y)}{p_Y(y)}\mathrm{d}u_1\mathrm{d}u_2\cdots\mathrm{d}u_{i-1}$$

$\mathrm{d}u_{i+1}\cdots\mathrm{d}u_n$(当 $p_Y(y)>0$ 时)

于是在 $\boldsymbol{Y}=y$ 的条件下 X_i 的条件期望为

$$E(X_i\mid\boldsymbol{Y}=y)=\int_{-\infty}^{+\infty}up_i(u\mid y)\mathrm{d}u\quad(i=1,2,\cdots,m).$$

自然定义 $E(\boldsymbol{X}\mid\boldsymbol{Y}=y)$为向量$(E(X_1\mid\boldsymbol{Y}=y),E(X_2\mid\boldsymbol{Y}=y),\cdots,E(X_m\mid\boldsymbol{Y}=y))$.

对离散型随机向量可进行类似的讨论,从略.

最佳预测与条件期望

设 X 和 Y 是两个随机变量,一个重要问题是如何根据 X 的观测值去预测 Y 的值(例如根据成年人的足长(脚趾到脚跟的长度)推测该人的身高,这在刑侦工作中相当重要).换句话说,如何寻找函数 $\psi(x)$使得 $\psi(X)$的值最接近 Y.一个重要提法是:如何找 $\psi(x)$使 $E[Y-\psi(X)]^2$达到最小(均方误差最小).

定理 5.1 设(X,Y)有联合密度 $p(x,y)$,$E(Y^2)$存在,令

$$\varphi(x)=\begin{cases}E(Y\mid X=x) & \text{当 } p_X(x)>0\\ 0 & \text{当 } p_X(x)=0\end{cases}$$

这里 $p_X(x)$是 X 的分布密度,则

$$E[Y-\varphi(X)]^2=\min_{\psi}E[Y-\psi(X)]^2 \tag{5.12}$$

换句话说.用 $\varphi(X)$预测 Y 时均方误差最小(当(X,Y)是离散型时,有类似的结论.从略).

证 不妨设 $E[\psi(X)]^2$存在.易知

$$E[Y-\psi(X)]^2=E[Y-\varphi(X)+\varphi(X)-\psi(X)]^2$$
$$=E[Y-\varphi(X)]^2+E[\varphi(X)-\psi(X)]^2+2E[Y-\varphi(X)][\psi(X)-\psi(X)]$$

我们指出,上式等号右边第三项等于0.为此只须证明

$$E|Y[\varphi(X)-\psi(X)]|=E|\varphi(X)[\varphi(X)-\psi(X)]| \qquad (5.13)$$

若 $p_X(x)=0$,则 $\int_{-\infty}^{+\infty}p(x,y)\mathrm{d}y=0$.从而不难推知 $\int_{-\infty}^{+\infty}yp(x,y)\mathrm{d}y=0$.于是利用(5.10)知

$$\int_{-\infty}^{+\infty}yp(x,y)\mathrm{d}y=\varphi(x)p_X(x) \qquad (\text{一切 } x)$$

利用均值公式(4.5)知

$$E|Y[\varphi(X)-\psi(X)]|=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y[\varphi(x)-\psi(x)]p(x,y)\mathrm{d}x\mathrm{d}y$$
$$=\int_{-\infty}^{+\infty}[\varphi(x)-\psi(x)]\left[\int_{-\infty}^{+\infty}yp(x,y)\mathrm{d}y\right]\mathrm{d}x$$
$$=\int_{-\infty}^{+\infty}[\varphi(x)-\psi(x)]\varphi(x)p_X(x)\mathrm{d}x=E|\varphi(X)[\varphi(X)-\psi(X)]|$$

故(5.13)成立.于是 $E[Y-\psi(X)]^2\geqslant E[Y-\varphi(X)]^2$.从而(5.12)成立.证毕.

这个定理告诉我们,用条件期望进行预测,均方误差最小.

用 X 表示我国成年人的身高,Y 表示成年人的足长,经过我国公安部门研究,有下列公式

$$E(X|Y=y)=6.876y$$

一案犯在保险柜前面留下足迹,测得足长 25.3 cm,代入上式算出此案犯的身高大约在 174 cm 左右.这一信息对于刻画案犯外形有着重要的作用.

例 5.5 若(X,Y)服从二维正态分布,其密度函数见本章(1.10).易知条件分布密度

$$p_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2}\exp\left\{-\frac{(y-m)^2}{2(1-\rho^2)\sigma_2^2}\right\},$$

其中 $m=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$ (参看例 5.2).

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

这表明用 $\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)$去预测 Y,均方误差最小.

若 $X_1, X_2, \cdots, X_m, Y$ 是 $m+1$ 个随机变量，如何根据 $X_1, X_2, \cdots, X_m$ 的值去预测 Y 的值呢？即如何找出 $\psi(x_1, x_2, \cdots, x_m)$ 使得用 $\psi = \psi(X_1, X_2, \cdots, X_m)$ 去预测 Y，均方误差最小. 可以证明下列一般性结论：

设 $E(Y^2)$ 存在，$X = (X_1, X_2, \cdots X_m)$

$$\varphi(x_1, x_2, \cdots, x_m) = E[Y \mid X = (x_1, x_2, \cdots, x_m)]$$

(这里用到条件期望的一般性定义，由于涉及较深的数学理论，我们不细说了.)则 $E[\varphi(X_1, X_2, \cdots, X_m) - Y]^2 = \min_{\psi} E[\psi(X_1, X_2, \cdots, X_m) - Y]^2$. 这表明用条件期望去预测，均方误差最小.

习 题 十 五

1. 设 X 与 Y 相互独立. X 服从泊松分布，$E(X) = \lambda_1$，Y 也服从泊松分布，$E(Y) = \lambda_2$. 试在 $X + Y = n$ 的条件下求出 X 的条件分布.

2. 一只小猫不幸陷进一个有三扇门洞的大山洞中. 第一个门洞通到一条通道，沿此通道走 2 h 后可到达地面. 第二个门洞通到另一个通道，沿它走 3 h 后又回到原处. 第三个门洞通到第三个通道，沿它走 5 h 后也回到原处. 假定这只小猫总是等可能地在三个门洞中任意选择一个. 试计算这只小猫到达地面的时间的期望.

3. 设 X 和 Y 都是离散型随机变量，$E(Y^2)$ 存在.

$$\varphi(x) = \begin{cases} E(Y \mid X = x) & \text{当 } P(X = x) > 0 \\ 0 & \text{否则} \end{cases}$$

试证明：对任何非负函数 $\psi(x)$，只要 $E[\psi(X)]^2$ 存在，必成立：

$$E[\varphi(X) - Y]^2 \leqslant E[\psi(X) - Y]^2$$

4. 设一天走进某百货商店的顾客数是均值为 1 200 的随机变量，又设这些顾客所花的钱数是相互独立的，均值为 50 元的随机变量. 又设任一顾客所花的钱数和进入该商店的总人数相互独立. 试问该商店一天的平均营业额是多少？

§6 大数定律和中心极限定理

作为本章的末尾，我们要简略地介绍一下概率论中基本的极限定理——著名的大数定律与中心极限定理. 我们只考虑最基本

的情况.

定义 6.1 称随机变量列 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的，如果对任何 $n \geqslant 1$，$X_1, X_2, \cdots, X_n$ 是相互独立的，此时，若所有的 X_i 又有相同的分布函数，则说 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量列.

定理 6.1 (大数定律)设 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量列，且 $E(X_1), D(X_1)$ 存在，则对任何 $\varepsilon > 0$，有

$$\lim_{n \to \infty} P\left\{\left|\frac{S_n}{n} - E(X_1)\right| \geqslant \varepsilon\right\} = 0 \tag{6.1}$$

其中 $S_n = X_1 + X_2 + \cdots + X_n$. 换句话说，只要 n 充分大，算术平均值 $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ 以很大的概率取值接近于期望.

证 利用切比雪夫不等式知

$$P\left\{\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2} D\left(\frac{S_n}{n}\right)$$

但 $E(S_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = nE(X_1)$，$D(S_n) = D(X_1) + D(X_2) + \cdots + D(X_n) = nD(X_1)$. 故

$$P\left\{\left|\frac{S_n}{n} - E(X_1)\right| \geqslant \varepsilon\right\} \leqslant \frac{D(X_1)}{n\varepsilon^2}$$

所以 $\lim_{n \to \infty} P\left\{\left|\frac{S_n}{n} - E(X_1)\right| \geqslant \varepsilon\right\} = 0$，这就证明了定理.

经过细致的数学研究知道，只要 $E(X_1)$ 存在，不管 $D(X_1)$ 是否存在，(6.1) 式仍然成立，而且可以证明比(6.1) 更强的结论：

$$P\left\{\lim_{n \to \infty} \frac{S_n}{n} = E(X_1)\right\} = 1 \tag{6.2}$$

通常把适合(6.1) 式的服从同一分布的随机变量列 $X_1, X_2, \cdots, X_n, \cdots$ 叫做服从大数定律(或弱大数定律)；把适合(6.2) 式的服从同一分布的随机变量列 $X_1, X_2, \cdots, X_n, \cdots$ 叫做服从强大数定律. 综上所述，具有数学期望的独立同分布的随机变量列是服从

大数定律和强大数定律的.

例6.1 设条件 S 下事件 A 的概率是 p. 将条件 S 独立地重复 n 次. 设 A 出现的次数是 μ. 令

$$X_i = \begin{cases} 1 & \text{当第 } i \text{ 次重复条件 } S \text{ 时 } A \text{ 出现} \\ 0 & \text{当第 } i \text{ 次重复条件 } S \text{ 时 } A \text{ 不出现} \end{cases}$$

显然 $X_1 + X_2 + \cdots + X_n = \mu, E(X_1) = P\{X_1 = 1\} = p$. 据(6.1)知 $\lim\limits_{n\to\infty} P\left\{\left|\frac{\mu}{n} - p\right| \geqslant \varepsilon\right\} = 0$, 即 A 发生的频率与概率 p 可任意接近. 从概率的定义来看, 这是很自然的.

定理 6.2 (中心极限定理)设 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量列, 而且 $E(X_1), D(X_1)$ 存在, $D(X_1) \neq 0$, 则对一切实数 $a < b$, 有

$$\lim_{n\to\infty} P\left\{a < \frac{S_n - nE(X_1)}{\sqrt{nD(X_1)}} < b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (6.3)$$

这里 $S_n = X_1 + X_2 + \cdots + X_n$.

由于这个定理的证明很长, 用到较多的数学知识, 我们就不证了, 读者可参阅参考书目[1].

记 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, (5.3)也可写成:

$$\lim_{n\to\infty} P\left\{a < \frac{\overline{X} - E(X_1)}{\sqrt{D(X_1)/n}} < b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

这表明, 只要 n 充分大, 随机变量 $\frac{\overline{X} - E(X_1)}{\sqrt{D(X_1)/n}}$ 就近似地服从标准正态分布, 从而 $\overline{X}$ 近似地服从正态分布. 故中心极限定理表达了正态分布在概率论中的特殊地位: 尽管 X_1 的概率分布是任意的, 但只要 n 充分大, 算术平均值 $\overline{X}$ 的分布却是近似正态的. 正态分布在理论上和应用上都具有极大的重要性, 在本讲义的后面几章里将多次看到这一点.

一般情形下的大数定律和中心极限定理

对于不服从同一分布甚至不相互独立的随机变量列，也可以讨论相应的“大数定律”、“强大数定律”及“中心极限定理”是否还成立的问题. 设 $X_1, X_2, \cdots$ 是随机变量列，$S_n = \sum_{i=1}^{n} X_i (n \geqslant 1)$. 条件(6.1),(6.2),(6.3)分别改为

$$\lim_{n \to \infty} P\left(\left| \frac{S_n - E(S_n)}{n} \right| \geqslant \varepsilon \right) = 0 \quad (\text{一切 } \varepsilon > 0) \tag{6.4}$$

$$P\left(\lim_{n \to \infty} \frac{S_n - E(S_n)}{n} = 0 \right) = 1 \tag{6.5}$$

$$\lim_{n \to \infty} P\left(a < \frac{S_n - E(S_n)}{\sqrt{D(S_n)}} < b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \tag{6.6}$$

(对一切 $a < b$)

若(6.4)成立，则称序列 $\{X_n, n \geqslant 1\}$ 服从大数定律；若(6.5)成立，则称 $\{X_n, n \geqslant 1\}$ 服从强大数定律；若(6.6)成立，则称中心极限定理对序列 $\{X_n, n \geqslant 1\}$ 成立.

设 $X_1, X_2, \cdots$ 是相互独立的随机变量列，方差 $D(X_n)$ 对 n 有界，利用切比雪夫不等式不难推知(6.4)成立，即大数定律成立.

定理 6.3 (Kolmogorov A N)设 $X_1, X_2, \cdots$ 是相互独立的随机变量列，若 $\sum_{n=1}^{\infty} \frac{D(X_n)}{n^2}$ 收敛，则该序列服从强大数定律.

定理 6.4 (Liapunov, A. M). 设 $X_1, X_2, \cdots$ 是相互独立的随机变量列，$\sigma_i^2 = D(X_i)$ 存在 $(i \geqslant 1)$，$B_n = \left(\sum_{i=1}^{n} \sigma_i^2 \right)^{\frac{1}{2}}$ $(n \geqslant 1)$ 且满足条件：

$$\lim_{n \to \infty} \frac{1}{B_n^3} \sum_{i=1}^{n} E|X_i - E(X_i)|^3 = 0 \tag{6.7}$$

则对一切 $a < b$ 有

$$\lim_{n \to \infty} P\left(a < \frac{S_n - E(S_n)}{B_n} < b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

即中心极限定理成立.

定理 6.3 和定理 6.4 的证明都用到较深的数学知识，从略.

从定理 6.4 知，只要 n 相当大，$\frac{S_n - E(S_n)}{B_n}$ 近似服从标准正态分布，从而 S_n 近似服从正态分布 $N(E(S_n), B_n^2)$. 这也表示正态分布的极大重要性：虽然

各 X_i 的分布是相当任意的($1\leqslant i\leqslant n$),但总和 $\sum_{i=1}^{n} X_i$ 却近似服从正态分布.

例 6.2 有一条河经过某城市.该河上有一座桥,该桥的强度服从正态分布 $N(300,40)$(强度的单位是吨(t).有很多车要经过此桥.如果各车的平均重量是 5 t,方差是 2,试问:为了保证此桥不出问题的概率(安全度)不小于 0.999 97,最多允许在桥上同时出现多少辆车?

解 用 Y 表示该桥的强度,若有 M 辆车在桥上,第 i 辆的重量是 X_i($i=1,2,\cdots,M$).则 M 辆车的总重量为 $S_M=\sum_{i=1}^{M} X_i$.我们可以认为 $Y,X_1,X_2,\cdots,X_M$ 是相互独立的,$E(X_i)=5$,$D(X_i)=2$.该桥不出问题的概率 R 为

$P(M$ 辆车的总重量不超过桥的强度).显然 $R=P(S_M\leqslant Y)=P(S_M-Y\leqslant 0)$.我们要找满足不等式 $R\geqslant 0.999\ 97$ 的最大的 M,不难想到,这个最大的 M 一定相当大,根据中心极限定理,S_M 近似服从正态分布 $N(M\mu_1,M\sigma_1^2)$,这里 $\mu_1=E(X_i)=5$,$\sigma_1^2=D(X_i)=2$.又 $Y\sim N(300,40)$.知 S_M-Y 近似服从 $N(M\mu_1-300,M\sigma_1^2+40)$.于是 $R\approx\Phi\left\{\dfrac{0-(M\mu_1-300)}{\sqrt{M\sigma_1^2+40}}\right\}$.由于 $\Phi(4)=0.999\ 97$,故为了 $R\geqslant 0.999\ 97$ 必须且只需 M 满足

$$\frac{0-(M\mu_1-300)}{\sqrt{M\sigma_1^2+40}}\geqslant 4\quad(\mu_1=5,\sigma_1^2=2)$$

令 $x=\sqrt{2M+40}$,则 $M=\dfrac{x^2-40}{2}$,上述不等式化为

$$\frac{5}{2}x^2+4x-400\leqslant 0$$

由此知 $x\leqslant 11.87$.故 $M\leqslant\dfrac{(11.87)^2-40}{2}=50.5$.由此知,最多允许 50 辆车同时在桥上.

第五章　统计估值

§1　总体与样本

前面四章我们初步研究了事件的概率和随机变量，很多实际问题(特别是自然现象和技术过程)中的随机现象可以用随机变量来描述．而要弄清一个随机变量，就必须知道它的概率分布，至少也要知道它的数字特征(期望、方差等)．怎样才能知道或大体知道一个随机变量的概率分布或数字特征呢？特别是，当我们对所要研究的随机变量知道不多或知之甚少的时候，用什么办法才能确定出这个随机变量的概率分布或数字特征呢？

这确实是应用中很重要的问题．请看两个简单例子．

例 1.1　某钢铁厂某一天生产 10 000 根 16 Mn 型钢筋，按规定强度小于 52 kg/mm^2 的算作次品，怎样求这批钢筋的次品率 p (也就是任取一根钢筋，它是次品的概率)呢？

例 1.2　灯泡厂生产灯泡，由于种种随机因素的影响，生产出来的灯泡的寿命是不同的．为了断定所生产灯泡的质量，怎样去估计某天所生产的灯泡的平均寿命以及使用时数长短的相差程度？

怎样解决这类问题呢？一个很重要的方法就是随机抽样法(或称抽样法)．这个方法的基本思想是，从要研究的对象的全体中抽取一小部分来进行观察和研究，从而对整体进行推断．

这种方法的重要性是很明显的，因为在工业生产和科学研究等领域里，有时普查方法是行不通的：不仅耗费的人力物力太多，时间上不允许；而且遇到检验产品质量是破坏性试验时，根本就不

能逐个检验，并且检验的数量还要适当地少．

例如要研究钢筋的强度，就从 10 000 根中抽出几根作为代表，比方说抽 50 根，对这 50 根进行检验，看看有多少根是次品．我们自然把这 50 根中的次品率当作 10 000 根的次品率的近似估计．对于灯泡寿命问题也可类似考虑．现在要讨论，为什么这样做是科学的？

这种随机抽样法（抽样法）是一种从局部推断整体的方法．因为局部是整体的一部分，所以局部的特性在某种程度上能反映整体的特性，另一方面又不能完全精确无误地反映整体的特性．作为研究整体与局部间辩证的数量关系的随机抽样法，包含两个组成部分，一是研究如何抽样，抽多少，怎样抽，这是抽样方法问题；另一是研究如何对抽查的结果（一批数据）进行合理的分析，作出科学的推断，这就是数据处理问题，即所谓统计推断的问题．数理统计学着重研究这两方面的问题．这两个部分又有着特别紧密的联系，研究抽样方法时必须要考虑到对抽查得到的数据能进行分析，抽查量太大是浪费，抽查量太小得不到可靠的结论，抽样的方法不合理（如得到的数据无代表性）根本就不能进行数据处理．所以要评价一个抽样方法，不仅要看它是否简便易行，更重要的是要看它的后果如何，即对抽查得到的一批原始数据能否用比较有效的方法进行数据处理，引出科学的结论．就是说，人们必须根据数据处理的要求，才能设计出好的抽样方法（抽样方案）．

由此可见，如何处理数据是一个更为基本的问题．

我们的分析和判断都是根据原始数据（抽查的结果）进行的，是一种统计推断．我们把所研究的对象的全体（包括有形的和潜在的）称为“总体”（例如，例 1.1 中的 10 000 根钢筋是一个总体，例 1.2 中某天所生产的灯泡的全体是一个总体）；把总体中每一个基本单位称为“个体”（例如，每一根钢筋都是一个个体）．

我们主要关心的是每个个体的某一特性值（即数量指标，例如钢筋的强度、灯泡的寿命）及其在总体中的分布情况（例如强度在

50 kg/mm^2 到 60 kg/mm^2 间的钢筋在 10 000 根中所占的比例，灯泡寿命在1 000小时至2 000小时的占全天生产的灯泡中的百分比）. 要考察总体中个体特性值的分布规律，可以采用这样的观点：将个体特性值看成一个随机变量，亦即从总体中随机抽取一个个体，所得个体的特性值 X 的大小是不能预先确定的，它依赖于被抽到的个体 . 很明显，这个随机变量 X 的概率分布正好体现总体中个体特性值的分布规律 . 由于我们只研究总体中个体特性值的分布规律，干脆把每一个总体用特性值随机变量 X 代表 . 这一段话的目的无非是提醒读者：要善于把你所研究的对象（某个特性值）看成一个“随机变量”.

在一个总体（例如 10 000 根钢筋，考虑其强度）X 中，抽取 n 个个体 $X_1, X_2, \cdots, X_n$（实际上 $X_1, X_2, \cdots, X_n$ 是所取的 n 根钢筋的强度），这 n 个个体 $X_1, X_2, \cdots, X_n$ 称为总体 X 的一个容量为 n 的“样本”（或叫**子样**），也称 n 为样本量 .

由于 $X_1, X_2, \cdots, X_n$ 是从总体 X 中随机抽取出来的可能结果，可以看成是 n 个随机变量；但是，在一次抽取之后，它们都是具体的数值，记作 $x_1, x_2, \cdots, x_n$，称作**样本值** . 今后，以 $X_1, X_2, \cdots, X_n$ 表示 n 个随机变量，以 $x_1, x_2, \cdots, x_n$ 表示样本值 . 在一次具体的抽取之后，$x_1, x_2, \cdots, x_n$ 都是具体的数值，但在两次抽取（每次取 n 个）中得到的两批数据一般是不同的 . 在不会引起混乱的情况下也用 $x_1, x_2, \cdots, x_n$ 表示 n 个随机变量 . 这样，记号 $x_1, x_2, \cdots, x_n$ 有双重意义：有时指的是某次具体抽取后的样本值，有时泛指任一次抽取后的结果（即看成 n 个随机变量）. 这在初学时会感到有些不习惯 .

我们的任务就是根据样本值 $x_1, x_2, \cdots, x_n$ 的性质，来对总体 X 的某些特性进行估计、推断 . 正因为如此，我们要求样本值尽可能地有代表性 . 这就对样本如何选取提出了一些要求，最有实用价值也比较自然的是要求样本 $X_1, X_2, \cdots, X_n$ 是相互独立的而

且与 X 有相同的概率分布．这种样本叫做“简单随机样本”①．由于本书主要讨论“简单随机样本”，所以，以后如果不特别声明，凡提到样本，都是指简单随机样本．怎样才能得到“简单随机样本”呢？有两种基本方法．

a．“有放回地逐次随机抽取法”．总体中的每个个体都有同样的机会被抽入样本，且每次抽出的个体，在记下其值后，还要放回到总体中去，以保证在下次抽取时每个个体仍有与第一次抽取时相等的机会被抽入样本．随机性表现在：样本中包含哪些个体，是出自机会，而不是在抽样前预定的．“有放回地抽取”有时很不方便，当总体所含个体的个数很大时可用“无放回地抽取”代替．例如前面所提到的那 10 000 根钢筋这个总体 X，从其中随机地选取 n 根 $X_1, X_2, \cdots, X_n$．只要 n 相对于 10 000 来说很小，那么 $X_1, X_2, \cdots, X_n$ 就可近似地看作一个简单随机样本．

b．对总体 X 进行多次独立的重复观测，这时观测到的值可以看成是总体的所有可能值（无形地存在着）的一部分．例如用仪器对某一物体的长度进行精密测量，我们把测量结果看成随机变量（总体可想像为一切可能值的集合，例如全体正数或更大的集合），把 n 次重复测量的结果记为 $X_1, X_2, \cdots, X_n$，则得到简单随机样本．

从数学上说，所谓总体就是一个随机变量 X，所谓样本就是 n 个相互独立且与 X 有相同概率分布的随机变量 $X_1, X_2, \cdots, X_n$（可看成是一个随机向量）．我们每一次具体的抽样，所得的数据

---

① 在总体只含有限个个体的情形（如有 N 个个体），“随机抽样法”往往指：从总体中随机地抽取 n 个．这里“随机”的含义是：从 N 个个体中任意抽取 n 个，共有 C_N^n 个可能的结果，这些结果有相等的概率，都是 $(C_N^n)^{-1}$．这是无放回的抽取法，得到的样本 $(X_1, X_2, \cdots, X_n)$ 不是这里定义的“简单随机样本”，我们称之为“单纯随机样本”，以示区别．参看参考书[10]的第八章．不难看出，当 N 很大时“单纯随机样本”就和“简单随机样本”性质相近了．

就是这 n 个随机变量的值(样本值),用 $x_1, x_2, \cdots, x_n$ 来表示．容易看出,如果 X 有分布密度 $p(x)$,则样本 $X_1, X_2, \cdots, X_n$ 有联合分布密度 $p(x_1)p(x_2)\cdots p(x_n)$．这个事实以后要多次用到．

最后,我们把上面关于总体与样本的讨论,用定义和定理的形式小结一下:

定义 1.1 称随机变量 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的容量是 n 的(简单随机)样本,如果 $X_1, X_2, \cdots, X_n$ 相互独立,而且每个 X_i 与 X 有相同的概率分布．(单个 X_i 叫做来自总体 X 的样品．)这时,若 X 有分布密度 $p(x)$,则常简称 $X_1, X_2, \cdots, X_n$ 是来自总体 $p(x)$ 的样本．

定理 1.1[①] 若 $X_1, X_2, \cdots, X_n$ 是来自总体 $p(x)$ 的样本,则 $(X_1, X_2, \cdots, X_n)$ 有联合密度 $p(x_1)p(x_2)\cdots p(x_n)$．

请读者就 $p(x)$ 为正态分布、指数分布的情形,分别写出样本 $(X_1, X_2, \cdots, X_n)$ 的联合密度．

§2 分布函数与分布密度的估计

设 X 是一个随机变量,怎样根据样本值 $x_1, x_2, \cdots, x_n$ 估计 X 的分布函数 $F(x)$ 的值呢? 给定 x 后,记 v_n 为 $x_1, x_2, \cdots, x_n$ 中不超过 x 的个数,自然用频率 $F_n(x) \triangleq \frac{v_n}{n}$ 去估计概率 $F(x) = P(X \leqslant x)$．

定义 2.1 称 x 的函数 $F_n(x)$ 为 X 的经验分布函数．

---

① 若 X 是离散型的随机变量,其可能值是 $a_1, a_2, \cdots$,概率分布为 $p(a_k) \triangleq P\{X = a_k\}(k = 1, 2, \cdots)$,则 X 的样本 $X_1, X_2, \cdots, X_n$ 有联合概率分布

$$p(x_1, x_2, \cdots, x_n) = P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = p(x_1)p(x_2)\cdots p(x_n)$$

$$(x_i = a_{j_i}, i = 1, 2, \cdots, n, j_i \geqslant 1)$$

将样本值 $x_1, x_2, \cdots, x_n$ 从小到大重排，得 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$（$x_{(1)}$ 是最小的，$x_{(2)}$ 是次小的，$\cdots$，$x_{(n)}$ 是最大的）. 这里 $x_{(i)}$ 叫做第 i 个次序统计量（$i=1,2,\cdots,n$）. 不难看出

$$F_n(x)=\begin{cases}0 & x<x_{(1)}\\ \dfrac{k}{n} & x_{(k)}\leqslant x<x_{(k+1)} \quad (k=1,2,\cdots,n-1)\\ 1 & x\geqslant x_{(n)}\end{cases}$$

从大数定律知，对固定的 x，只要 n 相当大，$F_n(x)$ 与 $F(x)$ 很接近.[①]

在实际工作中有时需要估计随机变量的分位数. 设 x_p 是 X 的 p 分位数（$0<p<1$），即有 $P(X<x_p)\leqslant p\leqslant P(X\leqslant x_p)$.

若 $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 是样本 $x_1, x_2, \cdots, x_n$ 的次序统计量. 令 $r=[pn]+1$，我们可用第 r 个次序统计量 $x_{(r)}$ 作为 x_p 的估计（数学上可以证明，当方程“$F(x)=p$”至多有一个根时，则 $P(\lim\limits_n x_{(r)}=x_p)=1$，这里 $F(x)$ 是 X 的分布函数. 证明较复杂，从略.）

例 2.1 某食品厂用自动装罐机生产额定净重为 345 克的午餐肉罐头，由于随机性，每个罐头的净重都有差别. 现在从生产线上随机抽取 10 个罐头，秤其净重，得下列结果：

344, 336, 345, 342, 340, 338, 344, 343, 344, 343 （单位：克）

试求该生产线上生产出的罐头的净重的分布函数，并估计其中位数.

解 我们用经验分布函数 $F_{10}(x)$ 作为分布函数 $F(x)$ 的估

---

① 数学上可以证明更强的结论：

定理（Glivenko－Cantelli）设

$$D_n=\sup_x|F_n(x)-F(x)|$$

则 $P(\lim\limits_n D_n=0)=1$ （证明见[10]的第二章）.

计值.将样本值从小到大排列,得

336,338,340,342,343,343,344,344,344,345

于是可得经验分布函数 $F_{10}(x)$ 如下:

$$F_{10}(x)=\begin{cases}0 & x<336\\ \frac{1}{10} & 336\leqslant x<338\\ \frac{2}{10} & 338\leqslant x<340\\ \frac{3}{10} & 340\leqslant x<342\\ \frac{4}{10} & 342\leqslant x<343\\ \frac{6}{10} & 343\leqslant x<344\\ \frac{9}{10} & 344\leqslant x<345\\ 1 & x\geqslant 345\end{cases}$$

这就是分布函数的近似值.注意 $p=\frac{1}{2}$ 时,$[pn]+1=6$.故可用 $x_{(6)}$ 作为中位数 $x_{\frac{1}{2}}$ 的估计.即罐头净重的中位数约为 343 g.

如果随机变量 X 有分布密度 $p(x)$,则应研究分布密度 $p(x)$ 如何估计,因为密度函数更能直观地刻画出概率分布的特性(如对称性、峰值等等).特别是对于多维随机向量,分布函数的实际用处较少,又不便于处理,因此估计密度函数的意义更大,在图像识别及多元判决中要经常用到.这里仅讨论一维随机变量的密度估计问题.方法有很多种,这里首先介绍历史悠久,现在仍在广泛使用的直方图法,然后简略介绍较晚发展起来的核估计法和最近邻估计法.

直方图法

设 $x_1,x_2,\cdots,x_n$ 是来自密度为 $p(x)$ 的总体的样本,用

$R_n(a,b)$表示样本中落入区间$(a,b]$的个数.若区间$(a,b]$之长度相当小,则对任何 $x\in(a,b]$,可用$\frac{1}{n(b-a)}R_n(a,b)$作为 $p(x)$的估计值.实际上,可用频率$\frac{1}{n}R_n(a,b)$估计概率 $P(a<X\leqslant b)$.这个概率$=\int_a^b p(x)\,\mathrm{d}x$,利用中值定理知,有 $x_0\in(a,b)$ 使 $p(x_0)=\frac{1}{b-a}\int_a^b p(x)\mathrm{d}x$.当 $p(x)$连续且 $b-a$ 很小时,$p(x)\approx p(x_0)$.可见用$\frac{1}{n(b-a)}R_n(a,b)$ 去估计 $p(x)(x\in(a,b])$是合理的.

基于上述思想,可用下法给出密度函数 $p(x)$的估计.设 $t_0<t_1<\cdots<t_m$ 是$m+1$个实数.通常假定 $t_i-t_{i-1}\equiv h>0(i=1,2,\cdots,m)$令

$$p_n(x)=\begin{cases}\frac{1}{nh}R_n(t_{i-1},t_i) & \text{当}x\in(t_{i-1},t_i],(i=1,2,\cdots,m)\\ 0 & \text{当}x\leqslant t_0\text{或}x>t_m\end{cases}$$

用 $p_n(x)$作为 $p(x)$的估计.这就是直方图估计法.

实际使用此法时,有三个步骤.叙述如下.

(1) 对样本值 $x_1,x_2,\cdots,x_n$ 进行分组.

首先找出 $x_1,x_2,\cdots,x_n$ 中的最小值$x_{(1)}$ 和最大值 $x_{(n)}$.取 a 为比$x_{(1)}$略小的数,b 为比$x_{(n)}$略大的数.将区间$(a,b]$$m$ 等分,分点为

$$t_i=a+i\frac{b-a}{m}\qquad(i=0,1,\cdots,m)$$

(m 的大小没有硬性规定,当样本量 n 小时m 也应小些,应使得大

多数小区间$(t_{i-1},t_i]$里包含有样本中的值①.另外,为方便起见,一般使t_i比样本值多一位小数.)

然后用唱票的办法,数出样本值落在区间$(t_{i-1},t_i]$中的个数,记为$\nu_i(i=1,2,\cdots,m)$

(2) 计算样本值落入各组的频率

$$f_i=\frac{\nu_i}{n}\qquad(i=1,2,\cdots,m)$$

(3) 作直方图

对每个$i(i=1,2,\cdots,n)$,在数轴上作以区间$[t_{i-1},t_i]$为底,以f_i/h为高的长方形(这里$h=t_i-t_{i-1}=(b-a)/m$).这一列长方形叫做直方图.这个图的好处在于,它大致地描述了X的概率分布情形.因为每个竖着的长方形的面积,刚好近似地代表了X取值落入"底边"的概率.

注意,图5.1中$(t_i,t_{i+1}]$上的长方形(阴影部分)的面积为:

$$\frac{f_{i+1}}{t_{i+1}-t_i}\cdot(t_{i+1}-t_i)=f_{i+1}$$

$$\approx P\{t_i<X\leqslant t_{i+1}\}$$

再回忆随机变量X的分布密度曲线的直观意义("曲边梯形"的面积代表X取值落入底边的概率),我们可以说,上面竖着的长方形面积近似地等于有同样底边的"曲边梯形"的面积.

大致经过每个竖着的长方形的"上边".换句话说,直方图提

---

① 有人建议采用下列公式:

$$m\approx 1+3.322\lg n$$

也可按下表选择m

n	m
<50	$5\sim6$
$50\sim100$	$6\sim10$
$100\sim250$	$7\sim12$
>250	$10\sim20$

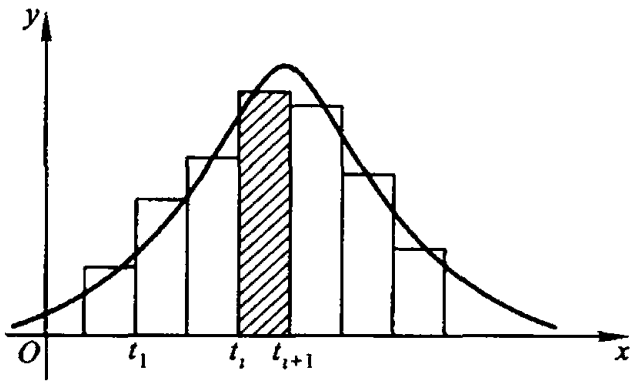

图　5.1

供了分布密度的大致样子．容易看出，如果样本容量越大(即 n 越大)，分组越细(即 m 越大)，则直方图就越接近分布密度曲线下的“曲边梯形”，因而提供了分布密度更加准确的样子．

对以下例 2.2 的详细叙述，相信可以帮助读者更好地掌握直方图法．

例 2.2　某炼钢厂生产了一种钢种叫 25 MnSi，由于各种偶然因素的影响，各炉钢的含 Si 量是有些差异的，因而应该把含 Si 量 X 看成一个随机变量，现在看看它的概率分布函数是怎样的？

为了确定分布密度，记录了 120 炉正常生产的 25MnSi 钢的含 Si 量的数据(百分数) 如下：

0.86	0.83	0.77	0.81	0.81	0.80
0.79	0.82	0.82	0.81	0.81	0.87
0.82	0.78	0.80	0.81	0.87	0.81
0.77	0.78	0.77	0.78	0.77	0.77
0.77	0.71	0.95	0.78	0.81	0.79
0.80	0.77	0.76	0.82	0.80	0.82
0.84	0.79	0.90	0.82	0.79	0.82
0.79	0.86	0.76	0.78	0.83	0.75
0.82	0.78	0.73	0.83	0.81	0.81
0.83	0.89	0.81	0.86	0.82	0.82

0.78	0.84	0.84	0.84	0.81	0.81
0.74	0.78	0.78	0.80	0.74	0.78
0.75	0.79	0.85	0.75	0.74	0.71
0.88	0.82	0.76	0.85	0.73	0.78
0.81	0.79	0.77	0.78	0.81	0.87
0.83	0.65	0.64	0.78	0.75	0.82
0.80	0.80	0.77	0.81	0.75	0.83
0.90	0.80	0.85	0.81	0.77	0.78
0.82	0.84	0.85	0.84	0.82	0.85
0.84	0.82	0.85	0.84	0.78	0.78

下面对这 120 个数据进行分组：

(1) 找出它们的最小值为 0.64，最大值为 0.95，其差为 0.31.

(2) 取起点 $a = 0.635$，终点 $b = 0.955$. 共分 16 组，组距 $= 0.02$.

(3) 分组及频数如下：

分组	频数 ν_i
0.635 ~ 0.655	2
0.655 ~ 0.675	0
0.675 ~ 0.695	0
0.695 ~ 0.715	2
0.715 ~ 0.735	2
0.735 ~ 0.755	8
0.755 ~ 0.775	13
0.775 ~ 0.795	23
0.795 ~ 0.815	24
0.815 ~ 0.835	21
0.835 ~ 0.855	14
0.855 ~ 0.875	6
0.875 ~ 0.895	2
0.895 ~ 0.915	2

0.915 ~ 0.935	0
0.935 ~ 0.955	1

以上用实例介绍了如何分组.下面根据分组情况及其频数来作直方图:

注意

$$\frac{t_i}{h} = \frac{\nu_i}{nh} \qquad (i = 1,2,\cdots,m)$$

在 x 轴上的每个区间$[t_{i-1}, t_i]$上作高为 ν_i/nh 的长方形($nh = 120 \times 0.02 = 2.4$).这一列长方形便是我们所要的直方图.为了方便起见,取纵坐标的单位长为$\frac{1}{nh} = \frac{1}{2.4}$,则直方图中第 i 个长方形的高度正好是 ν_i 个单位(见图 5.2).

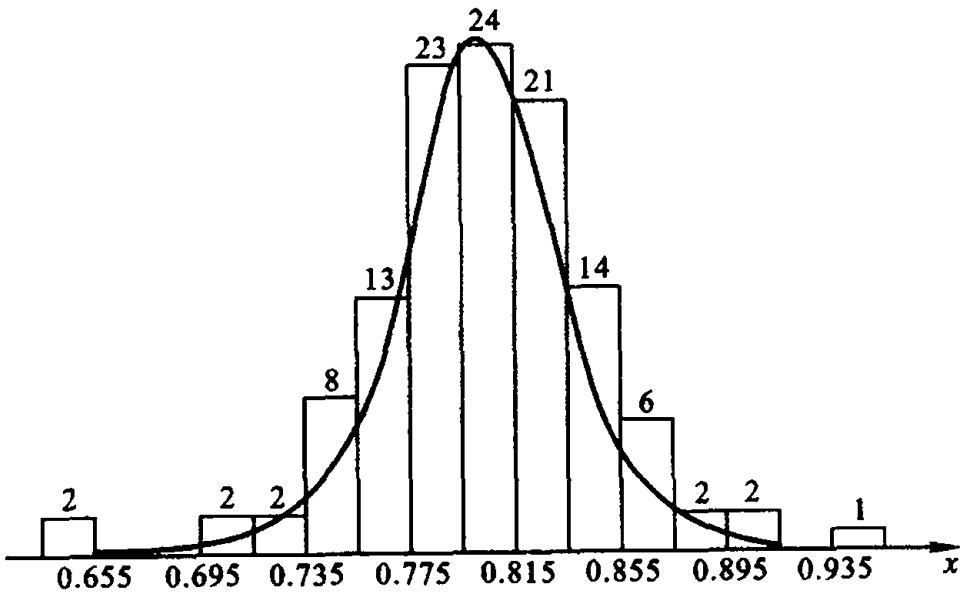

图 5.2

有了直方图,可以看出,X 的分布密度大体是图中曲线的位置.从图上看,这条曲线很像是正态分布密度的曲线,怎样根据数据判断 X 是否服从正态分布呢?解决这个问题的办法是有的.请看下一章最后一节.

对于直方图法，不难看出，当 n 无限增大且分点的间距 $h = h_n$ 无限减少时，估计量 $p_n(x)$ 与真正的密度 $p(x)$ 任意接近①.

核估计和最近邻估计

在引进一般的核估计之前，先讲一个特殊情形，以便读者理解核估计的思想.设随机变量 X 有分布函数 $F(x)$ 和密度函数 $p(x)$，若 $p(x)$ 连续，则 h 很小时有

$$\frac{F(x+h)-F(x-h)}{2h} \approx p(x)$$

而 $F(x)$ 可用经验分布函数 $F_n(x)$ 来估计，从而可用

$$\hat{p}_n(x) = \frac{1}{2h}[F_n(x+h) - F_n(x-h)]$$

来估计 $p(x)$，这叫做 $p(x)$ 的 Rosenblatt 估计，是 M. Rosenblatt 于 1956 年首先提出来的.不难看出，

$$\hat{p}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K_0\left(\frac{x-x_i}{h}\right)$$

这里 $x_1, x_2, \cdots, x_n$ 是样本，

$$K_0(x) = \begin{cases} \dfrac{1}{2} & -1 \leqslant x < 1 \\ 0 & \text{其他情形} \end{cases}$$

故 $\hat{p}_n(x)$ 可以通过一个“核函数”$K_0(x)$ 表达出来.

定义 2.2 设 $K(x)$ 是非负函数且 $\int_{-\infty}^{+\infty} K(x)\mathrm{d}x = 1$，则称 $K(x)$ 是核函数.此时称

$$\hat{p}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)$$

为 $p(x)$ 的核估计.

核函数有很大的选择自由，如

---

① 经过数学上的深入研究，可以证明（见[16]）：若密度函数 $p(x)$ 在 $(-\infty, +\infty)$ 上一致连续，对某个 $\delta > 0$，$\int_{-\infty}^{+\infty} |x|^{\delta} p(x)\mathrm{d}x$ 收敛，又 $\lim\limits_{n\to\infty} h_n = 0$，$h_n \geqslant \dfrac{1}{n}(\ln n)^2$，则 $P\left(\lim\limits_{n\to\infty} \sup\limits_{x} |p_n(x) - p(x)| = 0\right) = 1$.

$$K_0(x)=\begin{cases}\frac{1}{2} & -1\leqslant x<1\\ 0 & \text{其他}\end{cases}$$

$$K_1(x)=\begin{cases}1 & |x|\leqslant\frac{1}{2}\\ 0 & \text{其他}\end{cases}$$

$$K_2(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$$

$$K_3(x)=\frac{1}{\pi(1+x^2)}$$

$$K_4(x)=\frac{1}{2\pi}\left(\frac{\sin\frac{x}{2}}{\frac{x}{2}}\right)^2$$

可以证明,在一定条件下,当 n 无限增大且 $h=h_n$ 无限减小时,核估计 $\tilde{p}_n(x)$ 与 $p(x)$ 任意接近.①

只要核函数选得适当,核估计往往比直方图估计有较好的精度.例如,当 n 很大时,前面介绍的 Rosenblatt 估计(注意,这是一种特殊的核估计!)往往比直方图估计更接近真正的密度函数 $p(x)$.

对于密度函数 $p(x)$,还有一种估计法,就是所谓的最近邻估计.这是1965年提出来的.方法是:选定自然数 $K(n)$(n 是样本量).令

$$a_n(x)=\min\{t:t>0,R_n(x-t,x+t)\geqslant K(n)\}$$

$$p_n^*(x)=\frac{K(n)}{2na_n(x)}$$

其中 $R_n(x-t,x+t)$ 是样本 $x_1,x_2,\cdots,x_n$ 中落入区间 $(x-t,x+t]$ 的 x_i 的个数.

定义 2.3 称 $p_n^*(x)$ 为 $p(x)$ 的最近邻估计.

可以证明,在一定条件下,只要 n 充分大,最近邻估计 $p_n^*(x)$ 与 $p(x)$ 任

---

① 经过数学上的深入研究,可以证明下列结论(参看[16]):若密度函数 $p(x)$ 在 $(-\infty,+\infty)$ 上一致连续,且 $\lim\limits_{n\to\infty}h_n=0$,$\sum\limits_{n=1}^{\infty}\exp\{-rnh_n^2\}$ 收敛(对一切 $r>0$),又核函数是有界变差的,则 $P(\lim\limits_{n\to\infty}\sup\limits_{x}|\tilde{p}_n(x)-p(x)|=0)=1$

意接近.①

虽然核估计和最近邻估计在理论上有许多优点,在实际工作中用得最多的还是直方图估计.

§3 最大似然估计法

上节介绍了分布函数和分布密度的估计方法.这些方法要求样本量很大,即原始数据要很多,一般至少要 50 个以上,这在一些实际工作中是较难做到的.不过,许多实际工作中碰到的随机变量,其类型我们往往是知道的,只是不知道参数的值,因而写不出确切的密度函数或概率函数.例如,产品的某些指标很多是服从正态分布的,即密度函数是如下类型:

$$p(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

但 μ,σ^2 的值不知道.为了写出确切的密度函数,可以根据样本值 $x_1,x_2,\cdots,x_n$ 来估计出 μ,σ^2 的值.

又如产品的寿命,大量实践经验表明,它常常服从韦布尔分布或对数正态分布,服从其他分布的比较少见.问题就在于要根据样本值来估计出韦布尔分布或对数正态分布中的未知参数.

问题的一般提法是,设 X 的密度函数或概率函数②是 $p(x;\theta_1,\theta_2,\cdots,\theta_m)$,其中 $\theta_1,\theta_2,\cdots,\theta_m$ 是未知参数(只知它们属于一定范围 G 内,但具体数值不知).若 X 的样本值是 $x_1,x_2,\cdots,x_n$,问:如何估计出参数 $\theta_1,\theta_2,\cdots,\theta_m$ 的值?

---

① 经过数学上的深入研究,可以证明下列结论(参看[16]):若密度函数 $p(x)$ 在 $(-\infty,+\infty)$ 上一致连续,且

$$\lim_{n\to\infty}\frac{K(n)}{n}=0,\quad \lim_{n\to\infty}\frac{K(n)}{\ln n}=\infty$$

则

$$P\left(\lim_{n\to\infty}\sup_x|p_n^*(x)-p(x)|=0\right)=1.$$

② 设 X 是离散型随机变量,$y_1,y_2,\cdots$ 是其可能值,则 y_i 的函数 $P(X=y_i)$ 叫做 X 的概率函数.这个 $P(X=y_i)$ 常写作 $p(y_i)$.

这就是数理统计学中的参数估计问题．现代的估计法有很多种，最重要的是矩估计法与最大似然估计法．本节介绍理论上比较优良、适用范围较广的最大似然估计法．至于矩估计法，也很常用，我们将在§4中讲到它．

给定样本值 $x_1, x_2, \cdots, x_n$ 之后，令

$$L_n(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m) = \prod_{i=1}^{n} p(x_i; \theta_1, \theta_2, \cdots, \theta_m)$$

（这里符号 $\prod_{i=1}^{n}$ 代表连乘，例如 $\prod_{i=1}^{n} a_i \equiv a_1 a_2 \cdots a_n$．）

这 $L_n(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m)$ 叫做样本 $x_1, x_2, \cdots, x_n$ 的似然函数（注意，作为 $\theta_1, \theta_2, \cdots, \theta_m$ 的函数！）．

定义 3.1 如果 $L_n(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_m)$ 在 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ 达到最大值，则称 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ 分别是 $\theta_1, \theta_2, \cdots, \theta_m$ 的最大似然估计．

要注意的是，最大似然估计 $\hat{\theta}_i (i = 1, 2, \cdots, m)$ 与样本 $x_1, x_2, \cdots, x_n$ 有关，它是样本的函数，即 $\hat{\theta}_i = \hat{\theta}_i(x_1, x_2, \cdots, x_n)$，$(i = 1, 2, \cdots, m)$．

为了介绍最大似然估计的基本思想，我们考虑一个非常简单的估计问题．假定一个盒子里有许多黑球和白球，且假定已知它们的数目之比是 3:1，但不知白球多还是黑球多．也就是说抽出一个黑球的概率或者是 $\frac{1}{4}$ 或者是 $\frac{3}{4}$．如果有放回地从盒子里抽 3 个球，那么黑球数目 X 服从二项分布：

$$P\{X = x\} = C_3^x p^x (1-p)^{3-x}$$

$$x = 0, 1, 2, 3; p = \frac{1}{4}, \frac{3}{4}$$

其中，p 是抽到黑球的概率．

现在根据样本中的黑球数，来估计未知参数 p．在这种情况下估计问题实际上是很简单的，因为我们只要在两个数字 $\frac{1}{4}$ 和

$\frac{3}{4}$之间作一选择．抽样后，共有四种可能结果，它们的概率如下：

X	0	1	2	3
$p=\frac{1}{4}$时 $P\{X=x\}$的值	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$p=\frac{3}{4}$时 $P\{X=x\}$的值	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

如果样本中黑球数为0，那么，就应当估计 p 为$\frac{1}{4}$，而不估计为$\frac{3}{4}$，因为概率$\frac{27}{64}$比$\frac{1}{64}$大．就是说，具有 $X=0$ 的样本来自 $p=\frac{1}{4}$ 的总体的可能性比来自 $p=\frac{3}{4}$的总体的可能性要大．一般来说，当 $X=0,1$ 时，我们应当用$\frac{1}{4}$来估计 p；而当$X=2,3$时，应当用$\frac{3}{4}$来估计 p．估计量 $\hat{p}$ 是：

$$\hat{p}(x)=\begin{cases}\frac{1}{4} & \text{当 } x=0,1\\ \frac{3}{4} & \text{当 } x=2,3\end{cases}$$

也就是说，根据样本的具体情况来选择 $\hat{p}$，使得该样本发生的可能性最大．

怎样求最大似然估计呢？为方便起见，以下把似然函数 $L_n(x_1,x_2,\cdots,x_n;\theta_1,\cdots,\theta_m)$简记为 L_n．因为 L_n 与 $\ln L_n$ 同时达到最大值，有时只须求 $\ln L_n$ 的最大值点，这在计算上常常带来方便．

根据微积分的知识，当 $\ln L_n$ 的一阶偏微商存在时，则 $\ln L_n$ 在最大值点的一阶偏微商等于0．即最大似然估计 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m$ 满足方程组(称为似然方程组)：

$$\begin{cases} \dfrac{\partial \ln L_n}{\partial \theta_1} = 0 \\ \dfrac{\partial \ln L_n}{\partial \theta_2} = 0 \\ \cdots\cdots\cdots\cdots \\ \dfrac{\partial \ln L_n}{\partial \theta_m} = 0 \end{cases} \tag{3.1}$$

数学上可以严格证明，一定条件下（这些条件在大多数实际工作中常得到满足），只要样本量 n 足够大，最大似然估计和未知参数的真值可相差任意小．而且在一定意义上没有比最大似然估计更好的估计．

下面我们对几类常见的分布，来找出它们的参数的最大似然估计．

(1) 指数分布

$$p(x,\lambda) = \lambda e^{-\lambda x}, x>0, \lambda>0$$

样本 $x_1, x_2, \cdots, x_n$ 的似然函数：

$$L_n(x_1, x_2, \cdots, x_n; \lambda) = \lambda^n \prod_{i=1}^{n} e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_1^n x_i}$$

于是 $\ln L_n = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$，

$$\frac{\partial \ln L_n}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

故似然方程 $\dfrac{\partial \ln L_n}{\partial \lambda} = 0$ 的根 $\hat{\lambda} = \dfrac{n}{\sum\limits_{i=1}^{n} x_i} = \dfrac{1}{\bar{x}}$，这 $\hat{\lambda}$ 就是 λ 的最大似然估计（数学上容易验证，$\ln L_n$ 在 $\lambda = \hat{\lambda}$ 处确实达到最大值）．

例 3.1 已知某种电子设备的使用寿命（从开始使用到出现失效为止）服从指数分布，分布密度是 $p(x;\lambda) =$

$\lambda e^{-\lambda x}(x>0,\lambda>0)$,今随机抽取十八台,测得寿命数据如下(单位:小时):

16,29,50,68,100,130,140

270,280,340,410,450,520,620

190,210,800,1 100

问:如何估计出 λ?

解 采用最大似然估计法,利用公式 $\hat{\lambda}=\frac{1}{\bar{x}}$. 现在 $n=18,\bar{x}=318$. 知 $\hat{\lambda}=\frac{1}{318}$,这就是 λ 的估计值 .

(2) 正态分布

分布密度 $p(x;\mu,\delta)=\frac{1}{\sqrt{2\pi\delta}}e^{-\frac{1}{2\delta}(x-\mu)^2}$,其中 $\delta=\sigma^2>0$.

样本 $x_1,x_2,\cdots,x_n$ 的似然函数

$$L_n(x_1,x_2,\cdots,x_n;\mu,\delta)=\left(\frac{1}{\sqrt{2\pi\delta}}\right)^n\prod_{i=1}^{n}e^{-\frac{1}{2\delta}(x_i-\mu)^2}$$

$$=(2\pi)^{-\frac{n}{2}}\delta^{-\frac{n}{2}}e^{-\frac{1}{2\delta}\sum_{i=1}^{n}(x_i-\mu)^2}$$

于是,

$$\ln L_n=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\delta-\frac{1}{2\delta}\sum_{i=1}^{n}(x_i-\mu)^2$$

因此,似然方程组:

$$\begin{cases}\dfrac{\partial\ln L_n}{\partial\mu}=\dfrac{1}{\delta}\sum\limits_{i=1}^{n}(x_i-\mu)=0\\[2ex]\dfrac{\partial\ln L_n}{\partial\delta}=-\dfrac{n}{2\delta}+\dfrac{1}{2\delta^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$$

其根

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$

$$\hat{\delta}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

这就是 μ,δ 的最大似然估计(数学上可以验证,L_n 确实在 $\hat{\mu},\hat{\delta}$ 处达到最大值).

(3) 韦布尔分布

$$p(x;m,\eta)=\frac{m}{\eta^m}x^{m-1}\mathrm{e}^{-\left(\frac{x}{\eta}\right)^m}$$

$$(x>0;m>0,\eta>0)$$

这时,样本 $x_1,x_2,\cdots,x_n$ 的似然函数

$$L_n(x_1,x_2,\cdots,x_n;m,\eta)$$

$$=\left(\frac{1}{\eta^m}\right)^n\cdot m^n\cdot\prod_{i=1}^{n}x_i^{m-1}\mathrm{e}^{-\frac{1}{\eta^m}\sum_{i=1}^{n}x_i^m}$$

于是,

$$\ln L_n=-nm\ln\eta+n\ln m+(m-1)\sum_{i=1}^{n}\ln x_i-\frac{1}{\eta^m}\sum_{i=1}^{n}x_i^m$$

故似然方程组为:

$$\begin{cases}\dfrac{\partial\ln L_n}{\partial m}=-n\ln\eta+\dfrac{n}{m}+\sum_{i=1}^{n}\ln x_i\\ \qquad\qquad -\dfrac{1}{\eta^m}\sum_{i=1}^{n}x_i^m\ln x_i+\dfrac{\ln\eta}{\eta^m}\sum_{i=1}^{n}x_i^m=0 & (3.2a)\\ \dfrac{\partial\ln L_n}{\partial\eta}=-\dfrac{nm}{\eta}+\dfrac{m}{\eta^{m+1}}\sum_{i=1}^{n}x_i^m=0 & (3.2b)\end{cases}$$

从(3.2b)得

$$\eta=\left(\frac{1}{n}\sum_{i=1}^{n}x_i^m\right)^{\frac{1}{m}} \tag{3.3}$$

再代入(3.2a)得

$$\frac{1}{m}+\frac{1}{n}\sum_{i=1}^{n}\ln x_i-\frac{\sum_{i=1}^{n}x_i^m\ln x_i}{\sum_{i=1}^{n}x_i^m}=0 \tag{3.4}$$

可以证明,(当 $n\geqslant 2,x_1,x_2,\cdots,x_n$ 不全相等时)方程(3.4)恰

有一个根 $\hat{m}$. 再代入(3.3),得

$$\hat{\eta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\hat{m}}\right)^{\frac{1}{\hat{m}}} \tag{3.5}$$

这 $\hat{m}$, $\hat{\eta}$ 便是韦布尔分布中参数 m, η 的最大似然估计($\hat{m}$, $\hat{\eta}$ 不仅是似然方程组的解;经过数学上的细致研究知道,似然函数 L_n 在 $m=\hat{m}$, $\eta=\hat{\eta}$ 处达到最大值). 与指数分布、正态分布的情形不同,这里的 $\hat{m}$ 没有明显的数学表达式. 要找它,就需解超越方程(3.4). 应该指出,方程(3.4)中等号左边是 m 的严格减函数(对 m 微商后再利用 Schwarz 不等式即可推知),因而利用二分法极易求出方程的根.

例 3.2 轴承的寿命一般服从韦布尔分布. 我国某工厂对所生产的某型轴承进行质量检查. 随机抽取了 20 件进行寿命试验,测得寿命数据如下(单位:h):

153,223,313,373,378,385,424,232,452

452,547,561,634,699,759,859,1000,1132

1152,1466

试估计该韦布尔分布所含的形状参数 m 和刻度参数 η.

解 利用最大似然估计法. 现在样本量 $n=20$. 解方程(3.4)可求得 m 的最大似然估计 $\hat{m}=1.9$. 再利用(3.5)可求得 η 的最大似然估计 $\hat{\eta}=685$.

应注意的是,在寻找最大似然估计时,碰到似然函数不可微,则要直接研究似然函数的极值.

(4) 均匀分布

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

这是 $[a,b]$ 上均匀分布的密度函数,a 和 b 是未知参数,$a<b$.

这时,样本 $x_1, x_2, \cdots, x_n$ 的似然函数

$$L_n(x_1,x_2,\cdots,x_n;a,b)=\prod_{i=1}^{n}p(x_i;a,b)$$

令

$x_{(1)}=\min(x_1,x_2,\cdots,x_n)$，$x_{(n)}=\max(x_1,x_2,\cdots,x_n)$不难知道

$$L_n(x_1,x_2,\cdots,x_n;a,b)$$

$$=\begin{cases}\dfrac{1}{(b-a)^n} & \text{当}a\leqslant x_{(1)}\text{且}\ x_{(n)}\leqslant b\ \text{时}\\ 0 & \text{其他}\end{cases}$$

这个函数不是 a,b 的连续函数，不能对 a,b 求偏导数．但容易看出，要使 $L_n(x_1,x_2,\cdots,x_n;a,b)$最大，必须且只需 $b-a$ 最小．因而 $a=x_{(1)}$，$b=x_{(n)}$时，似然函数达到最大值．故 a 的最大似然估计 $\hat{a}=x_{(1)}$，b 的最大似然估计 $\hat{b}=x_{(n)}$．

§4 期望与方差的点估计

从前面的讨论看出，为了求出分布密度函数，直方图法要求数据很多，最大似然估计法又要求解一个有时并不好解的似然方程组，这些都非易事．好在有许多实际问题只要求对随机变量的一些数字特征（主要是期望和方差）有个恰当的估计值就够了，并不需要求出分布密度来．（当然，当分布密度函数中的未知参数刚好是期望、方差时（例如正态分布的情形），估出了期望和方差，也就估出了整个分布密度函数．）例如§1 例 1.2 中所举的灯泡质量检验问题，常常只需要估计寿命的期望（平均寿命）和寿命的方差（各灯泡寿命长短的相差程度）就可以了．对离散型随机变量可进行类似的讨论．

1．期望的点估计

§1 例 1.1 中的钢筋次品率问题，实际上也可以看作是如何估计一个随机变量的期望．我们可以用一个随机变量 X 来描述任抽一根钢筋的检查结果，如果抽到的钢筋是次品（即强度小于 52 kg/mm^2），则

令 $X=1$,如果抽到的钢筋不是次品,则令 $X=0$. 显然这样定义的 X 是一个离散型的随机变量.我们要求的次品率 p 就是 $P\{X=1\}$. 但是 $E(X)=1\cdot P\{X=1\}+0\cdot P\{X=0\}=P\{X=1\}=p$,这样求次品率 p 的问题就化成了估计期望 $E(X)$的值.

现在来研究一般性问题:如何去估计一个随机变量 X 的期望.

只要想起期望 $E(X)$是代表随机变量取值的"平均水平",就不难知道,可以把样本值的平均值 $\frac{x_1+x_2+\cdots+x_n}{n}$ 当作 $E(X)$的估计量.

人类的长期实践证明,这种用样本平均值去估计总体平均值(期望)的办法是很好的,而且样本容量 n 越大,估计得就越准.

我们今后也是用这个办法来估计 $E(X)$.

估计期望的办法是最简单不过的,无须多说.值得研究的是,这个办法为什么好?好在哪里?这个问题的回答就不那么简单了.问题在于 $E(X)$本身等于多少你不知道,看不见,摸不着,你能看见的只是样本值 $x_1,x_2,\cdots,x_n$. 但样本的具体数值却可以随机而变.(以刚才的灯泡问题为例,检验十只,得到了那样的数据是带有偶然性的,再检查另外十只得到的数据一般就变了样,换一人检查得到的十个数也会和这十个数不一样.总之,样本本身是随机向量.)

既然样本 $X_1,X_2,\cdots,X_n$ 是随机向量,则样本平均值 $\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}$ 是随机变量.我们说,$\bar{X}$ 取值虽然有偶然性,有时比 $E(X)$大,有时比 $E(X)$小,但是有下列定理:

定理 4.1 设 $E(X)$存在,则

$$E(\bar{X})=E(X)$$

证 实际上,

$$E(\bar{X})=E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{1}{n}[E(X_1)+E(X_2)+\cdots+E(X_n)]$$

$$=E(X)$$

最后一个等号是因为 $X_1,X_2,\cdots,X_n$ 与 X 有相同的概率分布,从

而期望也相等．证完．

这个定理告诉我们，用 $\bar{X}$ 估计 $E(X)$ 没有“系统偏差”．

记 $\bar{X}_n=\frac{1}{n}(X_1+X_2+\cdots+X_n)$，$\bar{X}_1$ 与 $\bar{X}_2$ 作为 $E(X)$ 的估计量都没有“系统偏差”，为什么 $\bar{X}_2$ 比 $\bar{X}_1$ 好呢？

我们自然想到，一个好的估计量应该取值稳定，因而要求方差小．可以证明 $D(\bar{X}_1)>D(\bar{X}_2)$（除非 $D(X)=0$），实际上有

定理 4.2 设 X 的期望、方差都存在，则

$$D(\bar{X}_n)=\frac{D(X)}{n}$$

证 $D(\bar{X}_n)=\frac{1}{n^2}[D(X_1)+D(X_2)+\cdots+D(X_n)]$

$$=\frac{D(X)}{n}$$

从这个定理知道，n 越大，$D(\bar{X})$ 就越小．这也就解释了前面提到的事实：n 越大，$\bar{X}$ 对 $E(X)$ 的估计就越好．

我们还可以把这一点说得更清楚一些．利用切比雪夫不等式：

$$P\{|\bar{X}-E(\bar{X})|<\varepsilon\}\geqslant 1-\frac{D(\bar{X})}{\varepsilon^2}$$

以及 $E(\bar{X})=E(X)$，$D(\bar{X})=\frac{D(X)}{n}$，我们得到：

$$P\{|\bar{X}-E(X)|<\varepsilon\}\geqslant 1-\frac{D(X)}{n\varepsilon^2} \tag{4.1}$$

故

$$\lim_{n\to\infty}P\{|\bar{X}-E(X)|<\varepsilon\}=1 \tag{4.2}$$

这说明，只要 n 足够大，就能以充分大的把握保证：

$|\bar{X}-E(X)|<\varepsilon$，即 $\bar{X}\approx E(X)$．

关系式(4.2)就是所谓大数定律．我们在第四章已讨论过．

估计量的优良性

设 X 的分布密度是 $p(x;\theta)$，其中 $\theta=(\theta_1,\theta_2,\cdots,\theta_m)\in\Theta$，$\Theta$ 是 m 维空间 R^m 中的某个集合(当 X 是离散型时，可作类似的讨论)．设 $g(\theta)$是参数(向量)θ 的函数，$X_1,X_2,\cdots,X_n$ 是 X 的样本．如何利用样本值对 $g(\theta)$进行估计?

定义 4.1 称样本的函数 $\varphi(X_1,X_2,\cdots,X_n)$为 $g(\theta)$的估计量．

φ 的不同选择就得到不同的估计量．什么样的 φ 是最优的呢? 这就涉及到优良性的标准了．

由于 $X_1,X_2,\cdots,X_n$ 的联合密度与 θ 有关，故 $\varphi(X_1,X_2,\cdots,X_n)$的数学期望与 θ 有关，以下记作 $E_\theta[\varphi(X_1,X_2,\cdots,X_n)]$．

定义 4.2 称 $\varphi(X_1,X_2,\cdots,X_n)$是 $g(\theta)$的无偏估计，若

$$E_\theta[\varphi(X_1,X_2,\cdots,X_n)]=g(\theta)\quad(\text{一切 }\theta\in\Theta)$$

定义 4.3 若 $\varphi_1(X_1,X_2,\cdots,X_n)$和 $\varphi_2(X_1,X_2,\cdots,X_n)$都是 $g(\theta)$的估计量，满足

$$E_\theta[\varphi_1(X_1,X_2,\cdots,X_n)-g(\theta)]^2\leqslant E_\theta[\varphi_2(X_1,X_2,\cdots,X_n)-g(\theta)]^2$$

(对一切 $\theta\in\Theta$)，且存在 $\theta_0\in\Theta$ 使上式左端严格小于右端，则说 φ_1 比 φ_2 有效．

从定理 2.1，2.2 知 $\bar{X}_k=\frac{1}{k}\sum_{i=1}^{k}X_i\ (k\leqslant n)$是 $E(X)$(即 $E_\theta(X)$)的无偏估计量，而且 $\bar{X}_k$ 比 $\bar{X}_{k-1}$有效(当 $D(X)\neq0$ 时)．

定义 4.4 如果 $\varphi(X_1,X_2,\cdots,X_n)$是 $g(\theta)$的无偏估计量，而且不存在无偏估计量比 φ 有效，则称 φ 是 $g(\theta)$的最小方差无偏估计量．

"最小方差无偏估计量"就是一种最优的估计量，可惜它有时并不存在．还有别的优良性标准，这里就不介绍了．

2．方差的点估计

$D(X)$ 是描述 X 取值的分散程度，也就是 X 取值偏离$E(X)$的程度．设 X 的样本值是 $x_1,x_2,\cdots,x_n$，这些数大小不一的程度(分散性)显然是反映了 X 取值的分散性．怎样描写 n 个数 $x_1,x_2,\cdots,x_n$ 大小不一的程度呢? 我们说，可以用这样的量：$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$．这个量越大，表明这组数很参差不齐；这个量越小，就表明这组数大小差不多．特别地，这 n 个数要是全相等，

则$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=0$,这是显而易见的．为什么要除以 $n-1$ 不除以 n 呢?这个道理见下面定理．量 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 叫做“样本方差”,记作 S^2 或小写的 s^2.

定理 4.3 设 X 的方差存在,则

$$E(S^2)=D(X)$$

证
$$\begin{aligned}\sum_{i=1}^{n}(X_i-\bar{X})^2&=\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)\\&=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-n\bar{X}^2\end{aligned}\tag{4.3}$$

于是

$$\begin{aligned}E(S^2)&=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]\\&=\frac{1}{n-1}\sum_{i=1}^{n}E(X_i^2)-\frac{n}{n-1}E(\bar{X}^2)\\&=\frac{n}{n-1}[E(X^2)-E(\bar{X}^2)]\end{aligned}$$

但 $E(\eta^2)=D(\eta)+(E\eta)^2$．故

$$\begin{aligned}E(S^2)&=\frac{n}{n-1}\{D(X)+(EX)^2-[D(\bar{X})+(E\bar{X})^2]\}\\&=\frac{n}{n-1}\left[D(X)+(EX)^2-\frac{D(X)}{n}-(EX)^2\right]\\&=D(X)\end{aligned}$$

这个定理告诉我们,用 S^2 估计方差 $D(X)$,虽然有时大些、有时小些,但没有“系统偏差”．如果采用估计量 $\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$,则这个估计量的期望等于 $\frac{n-1}{n}D(X)$,总比 $D(X)$ 小．不过

n 比较大时，这个估计量与 S^2 差异就不大了．所以在 n 比较大时，也常采用 $\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 作为 $D(X)$ 的估计量．

3．标准差的估计

如何估计总体的“标准差” $\sqrt{D(X)}$（常记作 σ）呢？既然 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是方差 $D(X)$ 的无偏估计量，自然想到用“样本标准差” $S\triangleq\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$ 作为 $\sqrt{D(X)}$ 的估计量．可以用这个办法估计 $\sqrt{D(X)}$，但要注意的是，S 一般不是 $\sqrt{D(X)}$ 的无偏估计量．实际上，对于正态总体（利用附录二定理 5 的系）可以证明

$$E\left[\frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\Gamma\left(\frac{n}{2}\right)\sqrt{2}}S\right]=\sqrt{D(X)}=\sigma$$

换句话说，$\frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\Gamma\left(\frac{n}{2}\right)\sqrt{2}}S$ 才是 σ 的无偏估计量．（在应用中将 S 前面的系数的数值列成表，以便查用．）

4．样本平均值 $\bar{x}$ 及样本方差 S^2 的简化算法

当样本量 n 很大时，如何计算出 $\bar{x}$，S^2 很值得考究．技巧运用得好，既省事又准确．下面介绍一种有用的笔算法．

例 4.1 设有下列样本值：

$$0.497,0.506,0.518,0.524,0.488$$
$$0.510,0.510,0.515,0.512$$

求 $\bar{x}$ 和 S^2．

解 令 $y_i=x_i\times 1\,000-500$，则 $y_i(i=1,2,\cdots,9)$ 为 $-3,6,18,24,-12,10,10,15,12$．

这些数是绝对值较小的整数，便于计算，易知

$$\bar{y}=\frac{1}{9}\sum_{i=1}^{9}y_i=\frac{80}{9}=8.9$$

$$\sum_{i=1}^{9}(y_i-\bar{y})^2=\sum_{i=1}^{9}y_i^2-\frac{1}{9}\left(\sum_{i=1}^{9}y_i\right)^2$$

$$=[(-1)^2+6^2+\cdots+12^2]-\frac{80^2}{9}$$

$$=946.9$$

容易求得[①]:

$$\bar{x}=\frac{500+\bar{y}}{1\,000}=\frac{500+8.9}{1\,000}=0.508\,9$$

$$s^2=\frac{1}{1\,000^2}\cdot\frac{1}{9-1}\sum_{i=1}^{9}(y_i-\bar{y})^2$$

$$=\frac{1}{10^6}\cdot\frac{1}{8}\cdot946.9$$

$$=0.000\,118$$

作为本节的末尾,我们来简略地介绍一下所谓矩估计法(简称矩法).

设随机变量 X 的分布密度是 $p(x;\theta_1,\theta_2,\cdots,\theta_m)$,当然 X 的 k 阶原点矩 $\nu_k=E(X^k)$ 也是 $\theta_1,\theta_2,\cdots,\theta_m$ 的函数,记 $g_k(\theta_1,\theta_2,\cdots,\theta_m)=E(X^k)(k=1,2,\cdots,m)$. 假定从方程组:

$$\begin{cases}g_1(\theta_1,\theta_2,\cdots,\theta_m)=\nu_1\\ g_2(\theta_1,\theta_2,\cdots,\theta_m)=\nu_2\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ g_m(\theta_1,\theta_2,\cdots,\theta_m)=\nu_m\end{cases}$$

---

① 我们用到下列事实:若 $y_i=bx_i-a(b\neq0,i=1,2,\cdots,n)$,则

$$\frac{\sum_{1}^{n}x_i}{n}=\frac{1}{b}\left(\frac{\sum_{1}^{n}y_i}{n}+a\right)$$

$$\sum_{1}^{n}(x_i-\bar{x})^2=\frac{1}{b^2}\sum_{1}^{n}(y_i-\bar{y})^2$$

可求出：

$$\begin{cases}\theta_1 = f_1(\nu_1, \nu_2, \cdots, \nu_m) \\ \theta_2 = f_2(\nu_1, \nu_2, \cdots, \nu_m) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \theta_m = f_m(\nu_1, \nu_2, \cdots, \nu_m)\end{cases}$$

设 $x_1, x_2, \cdots, x_n$ 是 X 的样本值，用

$$\hat{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

来估计 $\nu_k (k = 1, 2, \cdots, m)$. 然后，用

$$\hat{\theta}_k = f_k(\hat{\nu}_1, \hat{\nu}_2, \cdots, \hat{\nu}_m)$$

来估计 $\theta_k (k = 1, 2, \cdots, m)$.

这种估计未知参数的办法叫做矩估计法 .(刚才考虑的是原点矩，某些 ν_k 也可用中心矩 $\mu_k = E[X - E(X)]^k$ 代替，然后进行相类似的讨论，仍称矩法 .)

例 4.2 设 $X \sim N(\mu, \sigma^2)$, $x_1, x_2, \cdots, x_n$ 是其样本值 . 求 μ, σ^2 的矩估计量 .

解 易知 $\nu_1 = E(X) = \mu$, $\nu_2 = E(X^2) = \sigma^2 + \mu^2$，由这两个方程解得

$$\begin{cases}\mu = \nu_1 \\ \sigma^2 = \nu_2 - \nu_1^2\end{cases} \tag{4.4}$$

用 $\hat{\nu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ 和 $\hat{\nu}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ 分别估计 ν_1 和 ν_2，代入 (4.4)就可得到 μ, σ^2 的矩估计量：

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\sigma}^2 = \hat{\nu}_2 - \hat{\nu}_1^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

对于正态分布的参数 μ 和 σ^2 来说，矩估计量和最大似然估计量完全相同．但对不少分布，它们并不一样，通常用矩法估计参数较方便，但样本量 n 较大时，矩估计量的精度一般不及最大似然估计量的高．

例 4.3 设 X 的密度函数是

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

这里 θ 是未知的正数．设 $x_1,x_2,\cdots,x_n$ 是 X 的样本．不难看出 $E(X)=\dfrac{\theta}{2}$．故 θ 的矩估计 $\tilde{\theta}=\dfrac{2}{n}\sum\limits_{i=1}^{n}x_i$，可以证明，$\tilde{\theta}$ 与 $\hat{\theta}$ 并不一样．

***例 4.4** 台风可以引起内陆降雨．下列 36 个数是 24 小时降雨量的实际观测数据(单位：mm)：

31.00, 2.82, 3.98, 4.02, 9.50, 4.50, 11.40,
10.71, 6.31, 4.95, 5.64, 5.51, 13.40, 9.72,
6.47, 10.16, 4.21, 11.60, 4.75, 6.85, 6.25,
3.42, 11.80, 0.80, 3.69, 3.10, 22.22, 7.43,
5.00, 4.58, 4.46, 8.00, 3.73, 3.50, 6.20, 0.67

凭以往知识知道这种降雨量一般服从 Γ 分布，其分布密度为

$$p(x;\alpha,\beta)=\begin{cases}0, & x\leqslant 0\\ \dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x}, & x>0\end{cases}$$

我们可用矩法估计参数 α,β．用 X 表示降雨量，$x_1,x_2,\cdots,x_{36}$ 表示上述 36 个数．易知 $\nu_1=E(X)=\alpha/\beta$，$\nu_2=E(X^2)=\alpha(\alpha+1)/\beta^2$，$\hat{\nu}_1=\dfrac{1}{36}\sum\limits_{i=1}^{36}x_i=7.29$，$\hat{\nu}_2=\dfrac{1}{36}\sum\limits_{i=1}^{36}x_i^2=85.59$．

解方程组

$$\begin{cases}\alpha/\beta=7.29\\ \alpha(\alpha+1)/\beta^2=85.59\end{cases}$$

得 $\hat{\alpha}=1.64,\hat{\beta}=0.22$,这些分别是 α,β 的估计值.

§5 期望的置信区间

从上节知道,可以用 $\bar{X}$ 来估计 $E(X)$,用 S^2 来估计 $D(X)$,并且这些估计是相当好的.读者对此可能还会有不满足之处.到底 $\bar{X}$ 与 $E(X)$ 相差多少?(还有,S^2 与 $D(X)$ 相差多少?)这个问题可以换成一种提法:估计 $E(X)$ 所在的范围(区间),而且希望范围越小越好(对 $D(X)$ 也一样).

这就是对期望和方差的区间估计问题.

我们下面先讨论如何对期望 $E(X)$ 进行区间估计,这在实际应用中相当重要.至于方差,下节再讲.

我们的讨论分两种情形进行:

(1) 已知方差 $D(X)$,对 $E(X)$ 进行区间估计;

(2) 未知方差 $D(X)$,对 $E(X)$ 进行区间估计;

由于正态随机变量广泛存在,特别是很多产品的指标服从正态分布,我们重点研究正态随机变量情形的区间估计.先研究第一种情形,即已知方差 $D(X)$ 的情形.设 X 是一个正态随机变量,可以证明样本平均 $\bar{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n)$ 也是正态随机变量,且 $E(\bar{X})=E(X),D(\bar{X})=\frac{1}{n}D(X)$(参看习题十四第 9 题,也可参看附录二定理 5 的系).

于是随机变量

$$\eta=\frac{\bar{X}-E(X)}{\sqrt{\frac{D(X)}{n}}}$$

是服从标准正态分布的.查附表 1 知

$$P\{|\eta|\leqslant 1.96\}=0.95$$

即

$$P\left\{|\bar{X}-E(X)|\leqslant 1.96\sqrt{\frac{D(X)}{n}}\right\}=0.95 \tag{5.1}$$

从(5.1)式看出,我们有 95 %的把握保证:

$$|\bar{X}-E(X)|\leqslant 1.96\sqrt{\frac{D(X)}{n}}$$

即

$$\bar{X}-1.96\sqrt{\frac{D(X)}{n}}\leqslant E(X)\leqslant \bar{X}+1.96\sqrt{\frac{D(X)}{n}}$$

这就是说:虽然 $\bar{X}$ 是随机变量(取值随抽样的具体结果而定).但是随机区间

$$\left[\bar{X}-1.96\sqrt{\frac{D(X)}{n}},\bar{X}+1.96\sqrt{\frac{D(X)}{n}}\right] \tag{5.2}$$

以很大的概率包含 $E(X)$. 具体来说,如果做 100 次抽样(每次抽 n 个样品),则从平均的意义讲,算出的 $\bar{x}$ 值将有 95 次,使得区间(5.2)包含 $E(X)$.

根据上述,我们可以总结如下:随便作一次抽样,得到样本值 $x_1,x_2,\cdots,x_n$. 计算 $\bar{x}$. 我们可以认为 $E(X)$是落在区间(5.2)中,这就给出了 $E(X)$的区间估计,该区间称为置信区间 .

当然,也可能碰上这个区间并不包含 $E(X)$的偶然情形,此时我们就犯了错误 . 不过,出现这种情况的可能性比较小,约为 5 %.

我们还要注意,置信区间的长度与 n 有关 . 当然希望置信区间的长度越短越好,但为此需花费代价:即 n 必须大 . 故在实际问题里要具体分析,适当掌握,不能走极端 .

例 5.1 某车间生产滚珠,从长期实践中知道,滚珠直径 X 可以认为是服从正态分布的 . 从某天的产品里随机抽取 6 个,量得直径如下(单位:mm):

14.70,15.21,14.90,14.91,15.32,15.32

试估计该天产品的直径的平均值?

如果知道该天产品的直径的方差是0.05,试找出平均直径的置信区间?

解 用 X 表示该天产品的直径,要估计的就是 $E(X)$.

根据所给的样本值进行计算 .

$$\bar{x} = \frac{1}{6}(14.70 + 15.21 + 14.90 + 14.91 + 15.32 + 15.32)$$

$$= 15.06(\text{mm})$$

这就是 $E(X)$的近似值 .

为了找 $E(X)$的置信区间,我们来计算 $1.96\sqrt{\frac{D(X)}{n}}$. 现在 $n=6, D(X)=0.05$,于是

$$1.96\sqrt{\frac{D(X)}{n}} = 1.96 \times \sqrt{\frac{0.05}{6}} = 0.18$$

由(5.2)式,可以认为 $E(X)$在区间$[15.06-0.18, 15.06+0.18]$里 . 换句话说,滚珠直径的均值 $E(X)$的置信区间是$[14.88, 15.24]$.

对于不是服从正态分布的随机变量,如果 n 相当大,即所谓大样本(或大子样)的情形,仍可用(5.2)来对 $E(X)$进行比较准的估计 . 这是为什么呢? 原因在于有这样一个重要事实:无论 X 是怎样的随机变量,只要 n 充分大,随机变量 $\eta = \frac{\bar{X} - E(X)}{\sqrt{\frac{D(X)}{n}}}$就和标准正态随机变量差别很小 .

这就是概率论中有名的中心极限定理,我们在第四章里已介绍过了 .

由此看来,只要 n 充分大,就可认为 η 是服从标准正态分布 . 于是 $P\{|\eta| \leqslant 1.96\} = 0.95$,故又得到(5.2)式 . 这样,可以用 $\left[\bar{X} - 1.96 \times \sqrt{\frac{D(X)}{n}}, \bar{X} + 1.96 \times \sqrt{\frac{D(X)}{n}}\right]$作为 $E(X)$的置信区

间．

n 多大可以算作是“充分大”呢？很难提个绝对的标准，一般认为 n 不应小于 50．

以上找出的置信区间的可靠程度是 95%．我们也说置信水平(或置信度)是 95%．通常的工业生产和科学研究中是采取这个置信水平的，但有时嫌 95%偏低或偏高，而采取 99%或 90%的置信水平．置信水平定得不同，置信区间的长短就不同，但求置信区间的办法完全类似．请读者自己想一想．

上面的讨论是在已知方差 $D(X)$的情况下进行的．在实际应用中经常遇到不知道方差的情况，此时怎样对 $E(X)$找置信区间呢？现在就来研究与解决这个重要问题．

一个很自然的想法是，利用 $D(X)$的估计量

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

来代替 $D(X)$．就是说，研究

$$T=\frac{\bar{X}-E(X)}{\sqrt{\frac{S^2}{n}}}$$

的分布．

我们说，当 X 是正态随机变量时，随机变量 T 的分布确实能算出来．更确切些说，设 $X\sim N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 相互独立且与 X 有同样的概率分布．经过较长的数学推导(见附录二定理 7)，可以证明

$$T=\frac{\frac{X_1+X_2+\cdots+X_n}{n}-\mu}{\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n(n-1)}}} \tag{5.3}$$

的分布密度是

$$p_n(t)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} \tag{5.4}$$

显然,这个分布密度函数关于 $t=0$ 是对称的,它的图形如下:

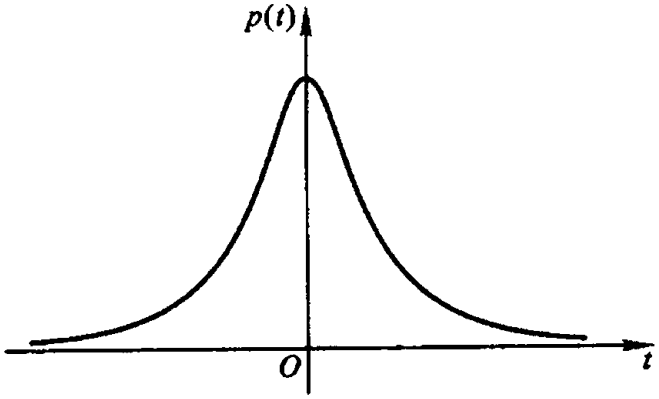

图 5.3

值得注意的是,这个分布密度与原来随机变量的期望 μ,方差 σ^2 无关,只与样本容量 n 有关 .n 是惟一的参数 . 知道了 n,$p_n(t)$就完全确定了 .

既然 $T=\dfrac{\bar{X}-E(X)}{\sqrt{\dfrac{S^2}{n}}}$的分布密度是(5.4),即:

$$P\{a\leqslant T\leqslant b\}=\int_a^b p_n(t)\mathrm{d}t$$

取数 λ 满足:

$$\int_{-\lambda}^{\lambda} p_n(t)\mathrm{d}t=0.95 \tag{5.5}$$

于是

$$P\left\{\left|\frac{\bar{X}-E(X)}{\sqrt{S^2/n}}\right|\leqslant\lambda\right\}=0.95 \tag{5.5}'$$

这就是说,以 95% 的把握保证:$E(X)$ 在区间 $\left[\bar{X}-\lambda\sqrt{\dfrac{S^2}{n}},\right.$

$\bar{X}+\lambda\sqrt{\frac{S^2}{n}}\Big]$中．这就给出了 $E(X)$的置信区间．

怎样具体找 λ 呢？经过数学工作者的研究，已把满足(5.5)的数值 λ 计算好了，并且列成了表，见本讲义附表 2．注意，对不同的 n，λ 的值也不同．这里我们来介绍一个名词：t 分布．

定义 5.1 如果随机变量 Y 的分布密度是：

$$p(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \tag{5.6}$$

则称 Y 服从 n 个自由度的 t 分布．

“自由度”的名字有点怪，大家记住这里是指表达式(5.6)中的参数 n．从(5.4)知道，我们关心的 T 正是服从 $n-1$ 个自由度的 t 分布．注意：自由度是样本容量减 1．

我们举例说明如何用上述理论去找置信区间．

例 5.2 用某仪器间接测量温度，重复测量 5 次，得到的结果如下(单位：℃)：

1 250，1 265，1 245，1 260，1 275

试问，温度的真值在什么范围内？

我们很容易把这个问题化成数学问题．用 μ 表示温度的真值，X 表示测量值．X 通常是一个正态随机变量，$E(X)=\mu$(假定仪器无系统偏差)．现在重复测量 5 次，得到 X 的 5 个值 $x_1=1\ 250,\cdots,x_5=1\ 275$，这就是样本值．问题就是未知方差(仪器的精度)的情况下，找期望(真值)的置信区间．

利用上述一般理论，知 μ 在区间 $\left[\bar{x}-\lambda\sqrt{\frac{S^2}{n}},\bar{x}+\lambda\sqrt{\frac{S^2}{n}}\right]$ 中．现在 $n=5$．

$$\bar{x}=\frac{1\ 250+1\ 265+1\ 245+1\ 260+1\ 275}{5}=1\ 259$$

$$\begin{aligned}S^2&=\frac{1}{5-1}[(1\ 250-1\ 259)^2+(1\ 265-1\ 259)^2+\\&\quad(1\ 245-1\ 259)^2+(1\ 260-1\ 259)^2+\\&\quad(1\ 275-1\ 259)^2]\\&=\frac{1}{4}[9^2+6^2+14^2+1^2+16^2]=\frac{570}{4}\end{aligned}$$

于是$\sqrt{\frac{S^2}{n}}=\sqrt{\frac{570}{5\times4}}=\sqrt{28.5}=5.339$. 自由度 = 样本容量 − 1 = 5 − 1 = 4,查 t 分布的临界值表($\alpha=0.05$),得 $\lambda=2.776$. 故

$$\lambda\sqrt{\frac{S^2}{n}}=2.776\times5.339\approx14.8$$

$$\bar{x}-\lambda\sqrt{\frac{S^2}{n}}=1\ 259-14.8=1\ 244.2$$

$$\bar{x}+\lambda\sqrt{\frac{S^2}{n}}=1\ 259+14.8=1\ 273.8$$

于是得到温度真值的置信度为 0.95 的置信区间[1 244.2, 1 273.8].

例 5.3 对飞机的飞行速度进行 15 次独立试验,测得飞机的最大飞行速度($m\cdot s^{-1}$)如下:

422.2,418.7,425.6,420.3,425.8

423.1,431.5,428.2,438.3,434.0

412.3,417.2,413.5,441.3,423.7

根据长期的经验,可以认为最大飞行速度服从正态分布,试对最大飞行速度的期望进行区间估计?

用 X 表示最大飞行速度,现在不知道 $D(X)$,要找 $E(X)$的置信区间. 由上面的一般讨论知,$E(X)$的置信区间是

$$\left[\bar{x}-\lambda\sqrt{\frac{S^2}{n}},\bar{x}+\lambda\sqrt{\frac{S^2}{n}}\right].$$

现在来具体计算 $\bar{x}$ 和 S^2，先将数据简化，令 $y_i = x_i - 420(i = 1,2,\cdots,15)$，则有

$$\begin{aligned}\sum_{i=1}^{15} y_i &= 2.2 - 1.3 + 5.6 + 0.3 + 5.8 + 3.1 + 11.5 + 8.2 + \\ &\quad 18.3 + 14.0 - 7.7 - 2.8 - 6.5 + 21.3 + 3.7 \\ &= 75.7\end{aligned}$$

$$\bar{x} = \bar{y} + 420 = \frac{75.7}{15} + 420 = 425.047$$

且有

$$\begin{aligned}S^2 &= \frac{1}{14}\sum_{i=1}^{15}(x_i - \bar{x})^2 = \frac{1}{14}\sum_{i=1}^{15}(y_i - \bar{y})^2 \\ &= \frac{1}{14}\left[\sum_{i=1}^{15} y_i^2 - \frac{1}{15}\left(\sum_{i=1}^{15} y_i\right)^2\right] \\ &= \frac{1}{14}\left[1\,388.37 - \frac{(75.7)^2}{15}\right] \\ &= \frac{1\,006.34}{14}\end{aligned}$$

查自由度为 14 的 t 分布临界值表，得 $\lambda = 2.145$. 于是

$$\begin{aligned}\bar{x} - \lambda\sqrt{\frac{S^2}{n}} &= 425.047 - 2.145\sqrt{\frac{1\,006.34}{15\times 14}} \\ &= 425.047 - 2.145\times\sqrt{4.792} \\ &= 425.047 - 4.696 = 420.35\end{aligned}$$

$$\bar{x} + \lambda\sqrt{\frac{S^2}{n}} = 425.047 + 4.696 = 429.74$$

故得 $E(X)$的置信度为 0.95 的置信区间为[420.35,429.74].

现将这部分内容总结如下：设 $X \sim N(\mu,\sigma^2)$，未知方差，找 μ 的置信区间的步骤是

① 由样本值 $x_1, x_2, \cdots, x_n$ 计算出 $\bar{x}, S^2$.

② 查 t 分布临界值表(本讲义附表 2)，注意自由度 $= n - 1$，$\alpha = 1 -$ 置信度，得临界值 λ.

③ 计算 $\lambda\sqrt{\frac{S^2}{n}}$(记作 d).

④ 得 μ 的置信区间$[\bar{x}-d,\bar{x}+d]$(置信度为 $1-\alpha$).

§6 方差的置信区间

上节我们研究了期望 $E(X)$的区间估计,找出了 $E(X)$的置信区间 . 但有的实际问题是要求对方差 $D(X)$(或标准差 $\sqrt{D(X)}$进行区间估计,即根据样本找出 $D(X)$的置信区间,这在研究生产的稳定性与精度问题时是需要的 .

例 6.1 某自动车床加工零件 . 抽查 16 个零件,测得长度如下(单位:mm):

12.15,12.12,12.01,12.08,12.09,12.16
12.03,12.01,12.06,12.13,12.07,12.11
12.08,12.01,12.03,12.06

怎样去估计该车床所加工零件的长度的方差?

按 §4 中的办法,当然可给出方差的一个近似值 . 先算样本平均值,得 $\bar{x}=12.075$. 再用简化法计算方差 $D(X)$的点估计值 .

$$S^2=\frac{1}{10\,000(16-1)}[15^2+12^2+\cdots+6^2-16\times7.5^2]$$

$$=\frac{366}{15\times10\,000}=0.002\,44$$

这是零件长度的真实方差的近似值 . 到底这近似值与真值相差多少呢? 这就需要给出方差真值的置信区间 .

现在来研究一般性的理论,然后再把理论用到刚才所举的例子上去 . 我们只研究总体是正态随机变量的情形 .

设 $X\sim N(\mu,\sigma^2)$, $X_1,X_2,\cdots,X_n$ 是来自这个总体的样本,我们的任务就是利用样本值 $x_1,x_2,\cdots,x_n$ 来给出 σ^2 的置信区间 .

我们已经知道,可用样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 来对

σ^2 进行估计，这是一个无偏估计．但是不知道 S^2 离 σ^2 差多少．

容易看出，如果把 $\frac{S^2}{\sigma^2}$ 看成随机变量，又能够找出它的概率分布，则我们的问题便迎刃而解了．

经过仔细的研究，可以证明（可参阅附录二定理 5 的系）随机变量 $\eta=\frac{(n-1)S^2}{\sigma^2}$ 的分布密度 $p(u)$ 是这样的：

$$p(u)=\begin{cases}\dfrac{1}{2^{\frac{n-1}{2}}\Gamma\left(\dfrac{n-1}{2}\right)}u^{\frac{n-3}{2}}\mathrm{e}^{-\frac{u}{2}} & u>0\\ 0 & u\leqslant 0\end{cases} \tag{6.1}$$

其图形如下（$n>3$）：

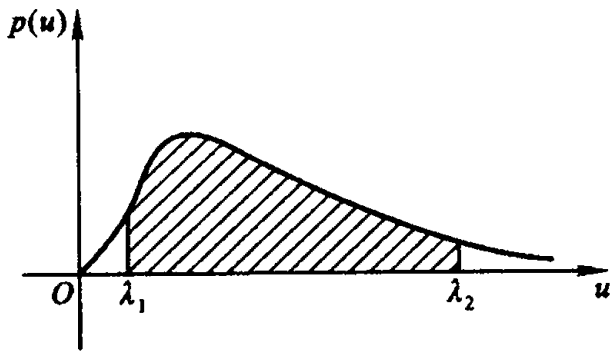

图 5.4

我们可选 $\lambda_1,\lambda_2(0<\lambda_1<\lambda_2)$ 满足：

$$P\{\lambda_1\leqslant\eta\leqslant\lambda_2\}=0.95 \tag{6.2}$$

即

$$\int_{\lambda_1}^{\lambda_2}p(u)\mathrm{d}u=0.95$$

换句话说，选 λ_1,λ_2 使得上面图中阴影部分的面积等于 0.95.

但合乎这个要求的 λ_1,λ_2 有很多对，究竟怎样选呢？[①]通常的办法是，使得阴影部分的左方的面积与右方的面积相等，都是 0.025. 用式子来写，就是选 λ_1,λ_2 满足：

$$\int_0^{\lambda_1} p(u)\mathrm{d}u = 0.025 \tag{6.3}$$

$$\int_{\lambda_2}^{+\infty} p(u)\mathrm{d}u = 0.025 \tag{6.4}$$

注意，(6.3)相当于

$$\int_{\lambda_1}^{+\infty} p(u)\mathrm{d}u = 0.975 \tag{6.5}$$

数学工作者已经把满足(6.5)和(6.4)的 λ_1,λ_2 计算出来了，我们只要学会查表就行了．这里我们介绍一个名词，它在统计里常遇到．

定义 6.1 如果随机变量 Y 的分布密度函数是这样的：

$$k_n(u)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}u^{\frac{n}{2}-1}\mathrm{e}^{-\frac{u}{2}} & u>0\\ 0 & u\leqslant 0\end{cases}$$

---

① 为了得到 σ^2 的 $1-\alpha$ 水平置信区间，应该选 $\lambda_1,\lambda_2(0<\lambda_1<\lambda_2)$ 满足

$$P(\lambda_1\leqslant\eta\leqslant\lambda_2)=1-\alpha \tag{注 6.1}$$

其中 η 的分布密度是(6.1)．从而可以得到 σ^2 的 $1-\alpha$ 水平置信区间

$$\left[\frac{1}{\lambda_2}\sum_{i=1}^{n}(X_i-\bar{X})^2,\frac{1}{\lambda_1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \qquad \text{(参看(6.6))}$$

这个区间的长度为 $\left(\dfrac{1}{\lambda_1}-\dfrac{1}{\lambda_2}\right)\sum_{i=1}^{n}(X_i-\bar{X})^2$．很自然想到，应选择 λ_1,λ_2 不仅满足(注 6.1)，而且应使得 $\dfrac{1}{\lambda_1}-\dfrac{1}{\lambda_2}$ 达到最小值．数学上可以证明，当 λ_1,λ_2 满足(注 6.1)且 $\lambda_1^{\frac{n+3}{2}}\mathrm{e}^{-\frac{\lambda_1}{2}}=\lambda_2^{\frac{n+3}{2}}\mathrm{e}^{-\frac{\lambda_2}{2}}$ 时 $\dfrac{1}{\lambda_1}-\dfrac{1}{\lambda_2}$ 达到了最小值．但用这个最优化原则确定 λ_1 和 λ_2 很不方便(当然，利用计算机总可以这样确定 λ_1,λ_2)，在实际工作中常常不追求“最优”，而是采用平分法，即选 λ_1 和 λ_2 分别满足：

$$\int_0^{\lambda_1} p(u)\mathrm{d}u=\frac{\alpha}{2},\int_{\lambda_2}^{+\infty} p(u)\mathrm{d}u=\frac{\alpha}{2}.$$

则称 Y 服从 n 个自由度的 χ^2 分布 .

从(6.1)式来看，$\eta=\dfrac{(n-1)S^2}{\sigma^2}$ 正是服从 $n-1$ 个自由度的 χ^2 分布 .

χ^2 分布的临界值 λ 可从附表 3 中查到 .

既然有了 λ_1 和 λ_2，根据(6.2)知，以 95 % 的把握保证：

$$\lambda_1\leqslant\frac{(n-1)S^2}{\sigma^2}\leqslant\lambda_2$$

换句话说，

$$\frac{(n-1)S^2}{\lambda_2}\leqslant\sigma^2\leqslant\frac{(n-1)S^2}{\lambda_1}$$

但

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

故得

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\lambda_2}\leqslant\sigma^2\leqslant\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\lambda_1}\tag{6.6}$$

这就给出了 σ^2 的置信度是 0.95 的置信区间 . 顺便还看出，σ 的置信区间是

$$\left[\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\lambda_2}},\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\lambda_1}}\right]$$

现在把上述理论用到上面的例 6.1 上去 .

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=0.036\,6$$

自由度是 15，查表知 $\lambda_1=6.26$，$\lambda_2=27.5$.

$$\frac{0.036\,6}{\lambda_1}=0.005\,8,\frac{0.036\,6}{\lambda_2}=0.001\,3$$

故 σ^2 的置信区间是[0.001 3,0.005 8]，σ 的置信区间是[0.036,

0.076].

*§7 寻求置信区间和置信限的一般方法

上面我们就正态分布介绍了期望和方差的置信区间的寻找方法.那么对于其他分布情形呢?而且未知参数也不一定是期望或方差.怎样去求未知参数或参数的函数的置信区间?更确切地说,设随机变量 X 的分布函数是 $F(x,\theta_1,\theta_2,\cdots,\theta_m)$,其中 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_m)$ 是未知参数向量,只知 $\boldsymbol{\theta}$ 属于某个集合 G,$g(\theta)$ 是实值函数.又设 $X_1,X_2,\cdots,X_n$ 是 X 的样本,我们有如下定义:

定义 7.1[①] 设 $\alpha\in(0,1)$,$\varphi_1(X_1,X_2,\cdots,X_n)$ 和 $\varphi_2(X_1,X_2,\cdots,X_n)$ 是两个统计量,$\varphi_1\leqslant\varphi_2$,称 $[\varphi_1,\varphi_2]$ 是 $g(\theta)$ 的置信水平是 $1-\alpha$ 的置信区间(或叫区间估计),若对一切 θ 均有

$$P[\varphi_1(X_1,X_2,\cdots,X_n)\leqslant g(\theta)\leqslant\varphi_2(X_1,X_2,\cdots,X_n)]\geqslant 1-\alpha, \quad (7.1)$$

应注意的是,(7.1)式中的概率 P 与 θ 有关,因为总体 X 的分布依赖于 θ.有时为确切计,用 P_θ 代替 P.若(7.1)式的左端的下确界(对一切 θ)恰好是 $1-\alpha$,则称区间 $[\varphi_1,\varphi_2]$ 的置信系数是 $1-\alpha$.

置信区间 $[\varphi_1,\varphi_2]$ 的好处在于:它以一定把握保证该区间包含 $g(\theta)$.

在(7.1)中的 φ_1,φ_2 分别称为 $g(\theta)$ 的置信下限、置信上限.在某些问题里只关心置信下限(例如产品的强度),此时取 $\varphi_2=+\infty$;在另一些问题里只关心置信上限(例如食品中某种有害细菌的数量),此时取 $\varphi_1=-\infty$.

寻找置信区间是一件重要的工作,当然应该限于寻找优良的置信区间.若不考虑优良性,取 $\varphi_1\to-\infty$,$\varphi_2\to+\infty$,$(-\infty,+\infty)$ 永远是 $g(\theta)$ 的 $1-\alpha$ 水平置信区间.很明显,这个置信区间毫无用处,它没有提供 $g(\theta)$ 的任何信息.那么,什么是优良的置信区间呢?这就涉及优良性的标准.我们这里不进行深入的讨论,只是指出:优良的置信区间其区间长度(即 $\varphi_2-\varphi_1$)应该是比

---

① 粗一看,在定义 7.1 的(7.1)式中出现"$\geqslant$"有些不顺眼(在前两节讨论正态分布期望和方差的置信区间时,均是"="号).能否把(7.1)中的"$\geqslant$"改为"="呢?我们指出,若在定义里将"$\geqslant$"改为"=",则对某些常见的分布(例如伯努利分布),参数的置信区间不存在.这一点以后将会看到.

较小的(如果只关心置信下限,则这种下限越大越好;如果只关心置信上限,则这种上限越小越好).

怎样寻找优良的置信区间呢? 这不是容易的事,要具体问题具体分析,有三个一般性方法可指导我们对具体问题进行分析,有助于找出优良的置信区间.本节介绍枢轴量方法和统计量方法,第三个方法是借助于假设检验理论的接受域方法,将在第六章中叙述.

(一) 枢轴量方法

这个方法是初等统计学中最常用的,前两节我们对正态分布寻求期望和方差的置信区间时用的就是这个方法.这个一般性方法叙述如下.为了寻找 $g(\theta)$ 的置信区间,我们设法选择与样本 $X_1, X_2, \cdots, X_n$ 及 $g(\theta)$ 有关的函数 $h[X_1, X_2, \cdots, X_n; g(\theta)]$,使得这个函数(实际是随机变量)的概率分布函数 $H(x)$ 与 θ 无关.在此基础上,找 $\lambda_1 < \lambda_2$ 满足 $H(\lambda_2) - H(\lambda_1) \geqslant 1-\alpha$.于是 $P\{\lambda_1 \leqslant h[X_1, X_2, \cdots, X_n; g(\theta)] \leqslant \lambda_2\} \geqslant 1-\alpha$,解不等式 $\lambda_1 \leqslant h(x_1, x_2, \cdots, x_n; u) \leqslant \lambda_2$,得到 $\varphi_1(x_1, x_2, \cdots, x_n) \leqslant u \leqslant \varphi_2(x_1, x_2, \cdots, x_n)$.于是 $[\varphi_1(X_1, X_2, \cdots, X_n), \varphi_2(X_1, X_2, \cdots, X_n)]$ 便是 $g(\theta)$ 的置信水平为 $1-\alpha$ 的置信区间.

上述的 $h[X_1, X_2, \cdots, X_n; g(\theta)]$ 一般称为枢轴量(pivotal),它含有样本 $X_1, X_2, \cdots, X_n$ 及 $g(\theta)$,但其概率分布与未知参数 θ 无关.如何找到合适的枢轴量就是问题的关键.在前两节里关于正态分布的讨论中正是由于找到了合适的枢轴量,才顺利地求出了期望和方差的置信区间.这里再举一例.

例 7.1(指数分布的参数的置信区间) 设 X 的密度函数是

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

其中 θ 是未知的正数.如何从样本 $X_1, X_2, \cdots, X_n$ 出发找出 θ 的置信区间?

易知 θ 的最大似然估计为 $\hat{\theta}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$.可以求出 $\sum_{i=1}^{n}X_i$ 的分布密度为

$$g(x,\theta)=\begin{cases}\dfrac{1}{(n-1)!}\dfrac{1}{\theta^n}x^{n-1}\exp\left\{-\dfrac{x}{\theta}\right\}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

(读者可用数学归纳法验证)

令 $h(X_1,X_2,\cdots,X_n;\theta)=2\sum_{i=1}^{n}X_i/\theta$,则 h 的分布密度是

$$p(x)=\begin{cases}\dfrac{1}{2^n\Gamma(n)}x^{n-1}e^{-\frac{x}{2}} & x>0\\ 0 & x\leqslant 0\end{cases} \tag{7.2}$$

这表明 h 是一个枢轴量,而且 h 服从 $2n$ 个自由度的 χ^2 分布.取 $\lambda_1,\lambda_2(0<\lambda_1<\lambda_2)$满足 $P(\lambda_1\leqslant 2\sum_{i=1}^{n}X_i/\theta\leqslant\lambda_2)=1-\alpha$(通过查 χ^2 分布的数值表可找出 λ_1,λ_2).解不等式 $\lambda_1\leqslant 2\sum_{i=1}^{n}X_i/\theta\leqslant\lambda_2$,可得到 θ 的 $1-\alpha$ 水平置信区间

$$\left[\frac{2}{\lambda_2}\sum_{i=1}^{n}X_i,\frac{2}{\lambda_1}\sum_{i=1}^{n}X_i\right].$$

这个置信区间的长度是 $\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right)2\sum_{i=1}^{n}X_i$.很自然想到,选择 $\lambda_1<\lambda_2$ 还应使得 $\frac{1}{\lambda_1}-\frac{1}{\lambda_2}$ 到最小值.可以证明,达到最小值的 λ_1,λ_2 满足两个方程

$$\begin{cases}\int_{\lambda_1}^{\lambda_2}p(x)\mathrm{d}x=1-\alpha\\ \lambda_1^{n+1}e^{-\frac{\lambda_1}{2}}=\lambda_2^{n+1}e^{-\frac{\lambda_2}{2}}\end{cases}$$

这里 $p(x)$是 $2n$ 个自由度的 χ^2 分布的密度函数(见(7.2)).从这两个方程就可以确定 λ_1 和 λ_2,但在实际工作中颇嫌不便.人们通常利用 χ^2 分布数值表,找 λ_1 和 λ_2 使之分别满足

$$\int_0^{\lambda_1}p(x)\mathrm{d}x=\frac{\alpha}{2},\int_{\lambda_2}^{+\infty}p(x)\mathrm{d}x=\frac{\alpha}{2}$$

而不去追求 λ_1,λ_2 的最优选择.

还应指出,指数分布常用来描述产品的寿命或生物的生存时间,这时参数 θ 就是平均寿命或平均生存时间.实际工作中最关心的是 θ 的置信下限,即要找 $\varphi_1(X_1,X_2,\cdots,X_n)$满足 $P[\theta\geqslant\varphi_1(X_1,X_2,\cdots,X_n)]\geqslant 1-\alpha$(对一切 θ),而且要 φ_1 尽可能的大.既然上述的 $h=2\sum_{i=1}^{n}X_i/\theta$ 是枢轴量,利用 χ^2 分布数值表可找到 λ_0 满足 $P\left(2\sum_{i=1}^{n}X_i/\theta\leqslant\lambda_0\right)=1-\alpha$ 解不等式 $2\sum_{i=1}^{n}X_i/\theta$

$\leqslant\lambda_0$，得 $\theta\geqslant\frac{2}{\lambda_0}\sum_{i=1}^{n}X_i$. 这表明 θ 的 $1-\alpha$ 水平置信下限为 $\frac{2}{\lambda_0}\sum_{i=1}^{n}X_i$.

枢轴量方法有两个明显的缺点：① 怎样寻找枢轴量？没有统一的方法；② 利用枢轴量方法得到的置信区间有何优良性质？没有一般性的结论.

（二）统计量方法

设 $X_1,X_2,\cdots,X_n$ 是来自分布函数为 $F(x,\theta)(\theta\in\Theta)$ 的总体的样本，Θ 是任意的非空集合，$g(\theta)$ 是 Θ 上的实值函数. 设 $\varphi(X_1,X_2,\cdots,X_n)$ 是任何统计量①（即样本 $X_1,X_2,\cdots,X_n$ 的函数），令

$$G(u,\theta)=P_\theta[\varphi(X_1,X_2,\cdots,X_n)\geqslant u] \tag{7.3}$$

$$H(u,\theta)=P_\theta[\varphi(X_1,X_2,\cdots,X_n)>u] \tag{7.4}$$

这里 $P_\theta(A)$ 是参数为 θ 时事件 A 的概率，$-\infty\leqslant u\leqslant+\infty$.

给定 $0<\alpha<1$，令

$$g_L(u)=\inf\{g(\theta):\theta\in\Theta\text{ 且 }G(u,\theta)>\alpha\} \tag{7.5}$$

$$g_U(u)=\sup\{g(\theta):\theta\in\Theta\text{ 且 }H(u,\theta)<1-\alpha\} \tag{7.6}$$

（当集合 $\{\theta:\theta\in\Theta$ 且 $G(u,\theta)>\alpha\}$ 是空集时，定义 $g_L(u)=+\infty$；当 $\{\theta:\theta\in\Theta$ 且 $H(u,\theta)<1-\alpha\}$ 是空集时，定义 $g_U(u)=-\infty$）.

我们有下列重要结论.

定理 7.1 (1) $g_L[\varphi(X_1,X_2,\cdots,X_n)]$ 是 $g(\theta)$ 的 $1-\alpha$ 水平（单侧）置信下限，即

$$P_\theta\{g(\theta)\geqslant g_L[\varphi(X_1,X_2,\cdots,X_n)]\}\geqslant 1-\alpha(\text{一切 }\theta\in\Theta) \tag{7.7}$$

(2) $g_U[\varphi(X_1,X_2,\cdots,X_n)]$ 是 $g(\theta)$ 的 $1-\alpha$ 水平（单侧）置信上限，即

$$P_\theta\{g(\theta)\leqslant g_U[\varphi(X_1,X_2,\cdots,X_n)]\}\geqslant 1-\alpha\quad(\text{一切 }\theta\in\Theta) \tag{7.8}$$

(3) 设 $g_1=\min\{g_L[\varphi(X_1,X_2,\cdots,X_n)],g_U[\varphi(X_1,X_2,\cdots,X_n)]\}$，$g_2=\max\{g_L[\varphi(X_1,X_2,\cdots,X_n)],g\setminus -U[\varphi(X_1,X_2,\cdots,X_n)]\}$，则 $[g_1,g_2]$ 是 $g(\theta)$ 的 $1-2\alpha$ 水平置信区间（当 $0<\alpha<0.5$）.

证 固定 $\theta\in\Theta$. 我们来证明

$$P_\theta\{g(\theta)<g_L[\varphi(X_1,X_2,\cdots,X_n)]\}\leqslant\alpha \tag{7.9}$$

为简单计，记 $Z=\varphi(X_1,X_2,\cdots,X_n)$，则 $G(u,\theta)=P_\theta(Z\geqslant u)$ 是 u 的减

---

① 这里的统计量是广义实值的，即取值是实数，也可以是 $+\infty$ 或 $-\infty$. 这种较广的定义有好处. 初学者不妨只考虑 $\varphi(X_1,X_2,\cdots,X_n)$ 取实数值的情形.

函数.令

$$C=\inf\{u:-\infty\leqslant u\leqslant+\infty, G(u,\theta)\leqslant\alpha\}$$

分两种情况进行讨论.

(Ⅰ) $G(C,\theta)\leqslant\alpha$.

此时,$P_\theta[g(\theta)<g_l(Z)]\leqslant P_\theta[G(Z,\theta)\leqslant\alpha]=P_\theta(Z\geqslant C)=G(C,\theta)\leqslant\alpha$,故(7.9)成立.

(Ⅱ) $G(C,\theta)>\alpha$

此时 $P_\theta[g(\theta)<g_L(Z)]\leqslant P_\theta[G(Z,\theta)\leqslant\alpha]=P_\theta(Z>C)\xlongequal{记}A$.分三种情形.

(i) $C=+\infty$.此时 $A=0$;

(ii) $C=-\infty$,此时 $A=\lim\limits_{n\to\infty}P_\theta(Z\geqslant -n)=\lim\limits_{n\to\infty}G(-n,\theta)\leqslant\alpha$;

(iii) $-\infty<C<+\infty$,此时 $A=\lim\limits_{n\to\infty}P_\theta(Z\geqslant C+\frac{1}{n})=\lim\limits_{n\to\infty}G\left(C+\frac{1}{n},\theta\right)\leqslant\alpha$.总之,$A\leqslant\alpha$.故(7.9)成立.从(7.9)直接推知(7.7)成立.

同理可证明(7.8)成立.

从(7.7)和(7.8)知 $P_\theta[g(\theta)\geqslant g_1]\geqslant 1-\alpha$,$P_\theta[g(\theta)\leqslant g_2]\geqslant 1-\alpha$.于是 $P_\theta[g(\theta)<g_1]\leqslant\alpha$,$P_\theta[g(\theta)>g_2]\leqslant\alpha$.从而 $P_\theta[g(\theta)<g_1$ 或 $g(\theta)>g_2]\leqslant 2\alpha$.故 $P_\theta[g_1\leqslant g(\theta)\leqslant g_2]\geqslant 1-2\alpha$.

定理 7.1 全部证完.

从一个统计量 $\varphi=\varphi(X_1,X_2,\cdots,X_n)$出发,利用定理 7.1 得到 $g(\theta)$的置信下(上)限及置信区间的方法,叫做统计量方法.实际应用此方法时,通常取 $g(\theta)$的一个估计量或估计量的一个增函数作为统计量 $\varphi(X_1,X_2,\cdots,X_n)$,而且要使得 $G(u,\theta)$和 $H(u,\theta)$,$g_L(u)$,$g_U(u)$都比较好计算,才便于获得具体的置信限或置信区间.这些都需要具体问题具体分析.

例 7.2 设 X 服从伯努利分布,即

$$P(X=1)=p=1-P(X=0)$$

其中 p 是未知参数,$0\leqslant p\leqslant 1$,问:如何从 X 的样本 $X_1,X_2,\cdots,X_n$ 找出 p 的置信下限(置信水平是 $1-\alpha$)?

易知 p 的矩估计 $\hat{p}=\sum\limits_{i=1}^{n}X_i/n$,取统计量 $\varphi(X_1,X_2,\cdots,X_n)=\sum\limits_{k=1}^{n}X_k$.令 $g(p)=p$,

$G(k,p)=P[\varphi(X_1,X_2,\cdots,X_n)\geqslant k](k=0,1,\cdots,n)$则

$G(k,p)=P\left(\sum_{i=1}^{n}X_i\geqslant k\right)=\sum_{i=k}^{n}C_n^i p^i(1-p)^{n-i}$多次使用分部积分公式得

$$G(k,p)=\frac{n!}{(k-1)!\ (n-k)!}\int_0^p x^{k-1}(1-x)^{n-k}\mathrm{d}x\quad(k\geqslant 1).\quad(7.10)$$

由此可见，$1\leqslant k\leqslant n$ 时，$G(k,p)$是 p 的严格增连续函数. 设 $p(k)$是方程 $G(k,p)=\alpha$ 的惟一根，则 $k\geqslant 1$ 时从(7.5)知

$$\begin{aligned}g_L(k)&=\inf\{p:0\leqslant p\leqslant 1,G(k,p)>\alpha\}\\&=p(k)\end{aligned}$$

另一方面，$g_L(0)=\inf\{p:0\leqslant p\leqslant 1,G(0,p)>\alpha\}=0$(因 $G(0,p)=1$). 令 $p(0)=0$. 从定理7.1 知 p 的 $1-\alpha$ 水平置信下限 $p_L=p\left(\sum_{i=1}^{n}X_i\right)$.

易知，$G(1,p)=1-(1-p)^n$，$G(n,p)=p^n$，故很易求出 $p(1)=1-(1-\alpha)^{\frac{1}{n}}$，$p(n)=\alpha^{\frac{1}{n}}$，当 $1<k<n$ 时，$p(k)$无显式表达，下一章将给出计算公式.

我们特别指出，$p(n)=\alpha^{\frac{1}{n}}$是工程上应用颇广的重要公式. 例如，为了估计某种炮弹的发射成功率 p，进行了 20 次试验，结果每次都成功，则 p 的 0.80水平置信下限 $p_L=(0.2)^{\frac{1}{20}}=0.922\ 7$.

$\alpha^{\frac{1}{n}}$是无失效情形下成功率的 $1-\alpha$ 水平置信下限.

(7.7) 告诉我们 $P_p\left[p\geqslant p\left(\sum_{i=1}^{n}X_i\right)\right]\geqslant 1-\alpha$. 数学上可以证明，$\inf\limits_{0<p<1}P_p\left[p\geqslant p\left(\sum_{i=1}^{n}X_i\right)\right]=1-\alpha$.(由于证明较长，从略).

上述结论自然会引起这样的问题：是否有 $\psi(X_1,X_2,\cdots,X_n)$满足

$$P_p[p\geqslant\psi(X_1,X_2,\cdots,X_n)]\equiv 1-\alpha(\text{一切 }0<p<1).\quad(7.11)$$

我们指出，这样的 $\psi(X_1,X_2,\cdots,X_n)$是不存在的. 我们用反证法加以证明. 设有这样的 ψ. 记 $a=\psi(1,1,\cdots,1)$(自变量全是 1). 分两种情况：

(Ⅰ) $a<1$，此时当 $p\in(a,1)$时 $1-\alpha\equiv P_p[p\geqslant\psi(X_1,X_2,\cdots,X_n)]\geqslant P_p[p\geqslant\psi(X_1,X_2,\cdots,X_n),X_1=X_2=\cdots=X_n=1]=P_p[p\geqslant a,X_1=X_2=\cdots=X_n=1]=p^n$. 令 $p\to 1$ 得 $1-\alpha\geqslant 1$. 这就产生了矛盾.

(Ⅱ) $a\geqslant 1$，此时对一切 $p\in(0,1)$，有 $\alpha\equiv P_p[p<\psi(X_1,X_2,\cdots,X_n)]\geqslant$

$P_p[p<\psi(X_1,X_2,\cdots,X_n),X_1=X_2=\cdots=X_n=1]=P_p[p<\alpha,X_1=X_2=\cdots=X_n=1]=p^n$. 令 $p\to1$ 得 $\alpha\geqslant1$,这与 $\alpha<1$ 相矛盾. 总之,不可能有 $\psi(X_1,X_2,\cdots,X_n)$使(7.11)成立.

我们指出,对任何 $0<\alpha<1$,不存在 $\psi_1(X_1,X_2,\cdots,X_n)\leqslant\psi_2(X_1,X_2,\cdots,X_n)$满足

$$P_p[\psi_1(X_1,X_2,\cdots,X_n)\leqslant p\leqslant\psi_2(X_1,X_2,\cdots,X_n)]\equiv1-\alpha \tag{7.12}$$

(对一 0 切 $p\in(0,1)$)

我们用反证法证明这一点. 假设有这样的 $\psi_1(X_1,X_2,\cdots,X_n)$和 $\psi_2(X_1,X_2,\cdots,X_n)$. 不妨设 $\psi_2(X_1,X_2,\cdots,X_n)\leqslant1$, $\psi_1(X_1,X_2,\cdots,X_n)\geqslant0$, 记 $a=\psi_1(1,1,\cdots,1)$, $b=\psi_2(1,1,\cdots,1)$(自变量全是 1). 则 $0\leqslant a\leqslant b\leqslant1$.

分三种情况讨论.

(Ⅰ) $b<1$

从(7.12)知,对一切 $p\in(b,1)$有

$$\begin{aligned}\alpha&=P_p[\psi_1>p\text{ 或 }\psi_2(X_1,X_2,\cdots,X_n)<p]\\&\geqslant P_p[\psi_2(X_1,X_2,\cdots,X_n)<p,X_1=X_2=\cdots=X_n=1]\\&=P_p(X_1=X_2=\cdots=X_n=1)=p^n\end{aligned}$$

令 $p\to1$ 得 $\alpha\geqslant1$,与已知条件 $0<\alpha<1$ 矛盾.

(Ⅱ) $b=1,a<1$.

从(7.12)知,对一切 $p\in(a,1)$有

$1-\alpha\geqslant P_p[\psi_1(X_1,X_2,\cdots,X_n)\leqslant p\leqslant\psi_2(X_1,X_2,\cdots,X_n),X_1=X_2=\cdots=X_n=1]=P_p(X_1=X_2=\cdots=X_n=1)=p^n$ 令 $p\to1$ 得 $1-\alpha\geqslant1$. 与 $0<\alpha<1$ 矛盾.

(Ⅲ) $b=1,a=1$.

从(7.12)知, $\alpha\equiv P_p[p<\psi_1(X_1,X_2,\cdots,X_n)$或 $p>\psi_2(X_1,X_2,\cdots,X_n)]\geqslant P_p[p<\psi_1(X_1,X_2,\cdots,X_n)]\geqslant P_p[p<\psi_1(X_1,X_2,\cdots,X_n),X_1=X_2=\cdots=X_n=1]=P_p(X_1=X_2=\cdots=X_n=1)=p^n$.

令 $p\to1$ 得 $\alpha\geqslant1$. 这与 $\alpha<1$ 相矛盾.

可见,不存在 $\psi_1(X_1,X_2,\cdots,X_n)\leqslant\psi_2(X_1,X_2,\cdots,X_n)$满足(7.12).

例 7.3 设 X 服从泊松分布,

$$P_\lambda(X=i)=\frac{e^{-\lambda}\lambda^i}{i!}\quad(i=0,1,\cdots)$$

其中 λ 是未知的正数.问:如何从 X 的样本 $X_1, X_2, \cdots, X_n$ 找出 λ 的单侧置信下限和置信上限?(置信水平是 $1-\alpha$).

易知 λ 的矩估计是 $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$. 取统计量 $\varphi(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} X_i$,令

$$G(k,\lambda) = P_\lambda(\varphi(X_1, X_2, \cdots, X_n) \geqslant k) \quad (k = 0,1,2,\cdots)$$

$$H(k,\lambda) = P_\lambda(\varphi(X_1, X_2, \cdots, X_n) > k)$$

由于 $\sum_{i=1}^{n} X_i$ 服从参数是 $n\lambda$ 的泊松分布,故

$$G(k,\lambda) = \sum_{m=k}^{\infty} \frac{e^{-n\lambda}(n\lambda)^m}{m!}$$

$$H(k,\lambda) = \sum_{m=k+1}^{\infty} \frac{e^{-n\lambda}(n\lambda)^m}{m!}$$

可以证明恒等式:

$$\sum_{m=k}^{\infty} \frac{e^{-A}A^m}{m!} = \int_0^{2A} g_{2k}(x)\mathrm{d}x \quad (\text{一切 } A>0) \tag{7.13}$$

这里 $g_{2k}(x)$ 是 $2k$ 个自由度的 χ^2 分布的密度函数,即

$$g_{2k}(x) = \frac{x^{k-1}e^{-\frac{x}{2}}}{2^k\Gamma(k)} \quad (x>0)$$

最简单的证明方法是直接验证(7.13)式中等号两端的函数对 A 微商后处处相等.从略.

利用(7.13)得

$$G(k,\lambda) = \int_0^{2n\lambda} g_{2k}(x)\mathrm{d}x$$

$$H(k,\lambda) = \int_0^{2n\lambda} g_{2k+2}(x)\mathrm{d}x$$

令

$$\lambda_L(k) = \inf\{\lambda : \lambda>0 \text{ 且 } G(k,\lambda)>\alpha\}$$

$$\lambda_U(k) = \sup\{\lambda : \lambda>0 \text{ 且 } H(k,\lambda)<1-\alpha\}$$

则

$\lambda_L(k)$ = 方程"$G(k,\lambda)=\alpha$"的惟一根,$\lambda_U(k)$ = 方程"$H(k,\lambda)=1-\alpha$"的惟一根.于是 $2n\lambda_L(k) = \chi^2_\alpha(2k)$, $2n\lambda_U(k) = \chi^2_{1-\alpha}(2k+2)$. 这里 $\chi^2_r(m)$ 是 m

个自由度的 χ^2 分布的 r 分位数(可通过查表得到具体数值),即 $\chi_r^2(m)$ 是满足下列等式的惟一的数.

$$\int_0^{\chi_r^2(m)} g_m(x)\mathrm{d}x = r \quad (g_m(x)\text{是 } m \text{ 个自由度的}\chi^2 \text{ 分布的密度函数}).$$

由此可见,$\lambda_L(k)=\frac{1}{2n}\chi_\alpha^2(2k)$,$\lambda_U(k)=\frac{1}{2n}\chi_{1-\alpha}^2(2k+2)$.跟据定理 7.1,$\lambda$ 的 $1-\alpha$ 水平置信下限是

$$\lambda_L=\frac{1}{2n}\chi_\alpha^2\left(2\sum_{i=1}^{n}X_i\right) \tag{7.14}$$

λ 的 $1-\alpha$ 水平置信上限是

$$\lambda_U=\frac{1}{2n}\chi_{1-\alpha}^2\left(2\sum_{i=1}^{n}X_i+2\right) \tag{7.15}$$

λ 的 $1-2\alpha$ 水平置信区间是$[\lambda_L,\lambda_U]$,泊松分布的用处很广.例如,为了考查某工厂所生产的布(或毛料)的质量,常用一定面积(如每平方米)上的疵点数来刻画.疵点数 X 一般服从泊松分布,参数 λ 是平均疵点数.利用(7.15)就可得到 λ 的置信上限.

从定理 7.1 的证明过程知道,我们并没有利用 $X_1,X_2,\cdots X_n$ 是"简单随机样本"的性质,只要$(X_1,X_2,\cdots,X_n)$是随机向量其概率分布依赖于参数 θ 即可.因而,统计量方法应用极广.既然从任何一个统计量 $\varphi(X_1,X_2,\cdots,X_n)$出发都可用来寻找 $g(\theta)$的置信限(下限或上限),那么自然要问:这样得到的置信限有何优良性?

为了表述优良性,先下一定义:

定义 7.2 设 $\varphi(x_1,x_2,\cdots,x_n)$和 $\psi(x_1,x_2,\cdots,x_n)$是两个函数.称 ψ 对 φ 是保序的,若对任何$(x_1,x_2,\cdots,x_n)$和$(x'_1,x'_2,\cdots,x'_n)$只要 $\varphi(x_1,x_2,\cdots,x_n)\leqslant\varphi(x'_1,x'_2,\cdots,x'_n)$就一定成立 $\psi(x_1,x_2,\cdots,x_n)\leqslant\psi(x'_1,x'_2,\cdots,x'_n)$.

我们可证明下列定理:

定理 7.2 设 $\varphi(X_1,X_2,\cdots,X_n)$是任何统计量,$g(\theta)$是 θ 的函数.$g_l[\varphi(X_1,X_2,\cdots,X_n)]$,$g_U[\varphi(X_1,X_2,\cdots,x_n)]$由(7.5)和(7.6)确定.若 $\psi_1(X_1,X_2,\cdots,X_n)$是 $g(\theta)$的任何 $1-\alpha$ 水平置信下限,且 $\psi_1(x_1,x_2,\cdots,x_n)$对$\varphi(x_1,x_2,\cdots,x_n)$保序,则

$$\psi_1(X_1,X_2,\cdots,X_n)\leqslant g_L[\varphi(X_1,X_2,\cdots,X_n)]$$

若 $\psi_2(X_1, X_2, \cdots, X_n)$ 是 $g(\theta)$ 的任何 $1-\alpha$ 水平置信上限，且 $\psi_2(x_1, x_2, \cdots, x_n)$ 对 $\varphi(x_1, x_2, \cdots, x_n)$ 保序，则

$$\psi_2(X_1, X_2, \cdots, X_n) \geqslant g_U[\varphi(X_1, X_2, \cdots, X_n)]$$

换句话说，由统计量 $\varphi(X_1, X_2, \cdots, X_n)$ 确定的置信下限是所有对 φ 保序的置信下限中最大的，所确定的置信上限是所有对 φ 保序的置信上限中最小的.

我们不叙述这个定理的证明了，读者如有兴趣，可参看[18]中的第六章.

习题十六

1. 某食品厂为加强质量管理，对某天生产的罐头抽查了 100 个(数据如下表). 试画直方图；它是否近似服从正态分布？

100 个罐头样品的净重数据(单位：g)：

342	340	348	346	343
342	346	341	344	348
346	346	340	344	342
344	345	340	344	344
343	344	342	343	345
339	350	337	345	349
336	348	344	345	332
342	342	340	350	343
347	340	344	353	340
340	356	346	345	346
340	339	342	352	342
350	348	344	350	335
340	338	345	345	349
336	342	338	343	343
341	347	341	347	344
339	347	348	343	347
346	344	345	350	341
338	343	339	343	346
342	339	343	350	341

346　341　345　344　342

2. 设 $x_1, x_2, \cdots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本值，μ 已知，求 σ^2 的最大似然估计量．

3. 设 $x_1, x_2, \cdots, x_n$ 是来自正态分布 $N(\mu, 1)$ 的样本值，求 μ 的最大似然估计量．

4. 设 X 服从区间 $[0, \lambda]$ $(\lambda > 0)$ 上的均匀分布，λ 是未知参数．而 $x_1, x_2, \cdots, x_n$ 是 X 的样本值，试求出 λ 的最大似然估计量和矩估计量．

5. 对 §5 的例 5.1，分别对置信度 0.99，0.90 找出均值的置信区间．

6. 对 §5 的例 5.2，分别对置信度 0.99 与 0.90，找出均值的置信区间．

7. 已知样本 3.3，−0.3，−0.6，−0.9，求具有 $\sigma = 3$ 的正态分布的均值的置信区间（置信度为 0.95）．如果 σ 未知，问均值的置信区间为何？

8. 对某一距离进行 5 次独立测量，得（单位：m）：

2 781，2 836，2 807，2 763，2 858

已知测量无系统误差，求该距离的置信度为 0.95 的置信区间（测量值可认为服从正态分布）．

9. 为了估计灯泡使用时数的均值 μ 及标准差 σ，测试 10 个灯泡．得 $\bar{x} = 1\,500$ h，$S = 20$ h．如果已知灯泡使用时数是服从正态分布的，求 μ 及 σ 的置信区间（置信度为 0.95）．

10. 测量铝的比重 16 次，测得 $\bar{x} = 2.705$，$S = 0.029$，试求铝的比重的置信区间（设测量值服从正态分布，置信度为 0.95）．

11. 设 $X \sim N(\mu, \sigma^2)$，$x_1, x_2, \cdots, x_n$ 是其样本值．如果 σ^2 已知，问：n 取多大时方能保证 μ 的置信度为 0.95 的置信区间的长度不大于给定的 L？

12. 随机地从甲批导线中抽取 4 根，从乙批导线中抽取 5 根，测得其电阻为（单位：Ω）：

甲批导线：0.143，0.142，0.143，0.137

乙批导线：0.140，0.142，0.136，0.138，0.140

设甲、乙两批导线的电阻分别服从 $N(\mu_1, \sigma^2)$、$N(\mu_2, \sigma^2)$（并且它们相互独立），σ^2 已知，等于 $0.002\,5^2$，但 μ_1、μ_2 均未知．试求 $\mu_1 - \mu_2$ 的置信度为 0.95 的置信区间．

第六章　假设检验

§1　问题的提法

上一章我们介绍了估计参数和估计分布密度的方法，但在实践中还有许多重要问题与估计问题的提法不同，也需要我们去解决.请看下列简单的例子.

例 1.1　某厂有一批产品，共 200 件，须经检验合格才能出厂，按国家标准，次品率不得超过 1 %，今在其中任意抽取 5 件，发现这 5 件中含有次品.问这批产品是否能出厂？

从直观上看，这批产品是不能出厂的.但理由何在？

设这批产品的次品率是 p.问题化为：如何根据抽样的结果来判断不等式“$p\leqslant 0.01$”成立与否？

例 1.2　用某仪器间接测量温度，重复五次，所得数据如下：(单位：℃)1 250，1 265，1 245，1 260，1 275，而用别的精确办法测得温度为 1 277(可看作温度的真值)，试问此仪器间接测量有无系统偏差？

用 X 代表用这个仪器测得的数值，当然这是一个随机变量，得到的 5 个数据是 X 的一个样本.问题化为：如何判断等式“$E(X)=1\ 277$”成立与否？

例 1.3　某工厂近 5 年来发生了 63 次事故，这些事故在工作日的分布如下

星期	一	二	三	四	五	六
次数	9	10	11	8	13	12

问:事故的发生是否与星期几有关?

用 X 表示这样的随机变量:若事故发生在星期 i,则 $X=i$. 显然 X 的可能值是 $1,2,\cdots,6$(星期日是该厂厂休日). 问题化为如何判断 $P(X=i)\equiv\frac{1}{6}(i=1,2,\cdots,6)$是否成立?

例 1.4 在针织品的漂白工艺过程中,要考察温度对针织品断裂强力(主要质量指标)的影响. 为了比较 70℃ 与 80 ℃ 的影响有无差别,在这两个温度下,分别重复作了八次试验,得数据如下:(单位:千克力)

70℃ 时的强力:20.5,18.8,19.8,20.9,21.5,19.5,21.0,21.2

80℃ 时的强力:17.7,20.3,20.0,18.8,19.0,20.1,20.2,19.1

究竟 70 ℃ 下的强力与 80℃ 下的强力有没有差别?

用 X 表示 70 ℃ 下的强力,Y 表示 80 ℃ 下的强力,问题变成:如何判断等式"$E(X)=E(Y)$"成立与否?(还可进一步问等式"$D(X)=D(Y)$"成立与否?)

例 1.5 某公司生产一种头发干燥机(吹风机),销售情况一向良好,但现在面临激烈的市场竞争,压力很大. 该公司研究与开发部研制出一种新型干燥机,单机的成本比原先的减少 15 %,但公司的副总裁不愿批准此项新产品上市销售,担心新产品的可靠性不如原产品.(该公司对商品销售有一年的保质期,在保质期内失效的商品(即失去规定功能的商品)可以免费更换). 为此该公司进行了可靠性试验. 将新产品和原产品各取 250 件在模拟一年使用的条件下进行试验,发现新产品中有 11 个失效,原产品中有 20 个失效. 问:新产品的可靠性是否不比原产品的差?

用 p_1 表示新产品的失效率,p_2 表示原产品的失效率,问题化为判断"$p_1\leqslant p_2$"是否成立?

例 1.6 怎样根据一个随机变量的样本值,判断该随机变量是否服从正态分布 $N(\mu,\sigma^2)$?

更一般地,如何根据样本的特性去判断随机变量是否以给定

的函数 $F(x)$为其分布函数?

这些例子所代表的问题是很广泛的,其共同点就是要从样本值出发去判断一个"看法"是否成立.例 1.1 的看法是"次品率 $p\leqslant 0.01$",例 1.2 的看法是"$E(X)=1\ 277$",例 1.3 的看法是"$P(X=i)\equiv\frac{1}{6}(i=1,2,\cdots,6)$",例 1.4 的看法是"$E(X)=E(Y)$",例 1.5 的看法是"$p_1\leqslant p_2$",例 1.6 的看法是"$X$ 的分布函数是$F(x)$".

"看法"又叫"假设".这些就是所谓假设检验问题(或叫假设的鉴定问题).

本章的任务就是介绍一些常用的检验办法,判断所关心的"假设"是否成立.

例 1.1、例 1.2 和例 1.3 中的"假设"都是关于一个随机变量的参数的判断,这叫做一个总体的参数检验问题.例 1.6 也是一个总体的检验问题,不过它一般不是参数检验,而是概率分布的检验问题.

例 1.4 和例 1.5 中的"假设"是关于两个随机变量的判断,这叫二总体的检验问题.也可以考虑三个或更多个总体的检验问题.

怎样对"假设"进行检验呢? 无论"假设"的类型多么复杂,进行检验的基本思想却是很简单的,是某种带有概率性质的反证法.掌握这个基本思想是很重要的.下面我们通过例 1.1 来说明假设检验的基本思想.

例 1.1 要检验的假设是"$p\leqslant 0.01$".如果假设 $p\leqslant 0.01$ 成立,看看会出现什么后果.此时,200 件中最多有两件是次品,任抽取 5 件,我们先来求这 5 件中"无次品"的概率.在第一章中我们已熟知这类问题的解法.

$$P\{\text{无次品}\}=\begin{cases}\dfrac{C_{198}^{5}}{C_{200}^{5}} & \text{当 200 件中有两件次品时}\\ \dfrac{C_{199}^{5}}{C_{200}^{5}} & \text{当 200 件中有一件次品时}\\ \dfrac{C_{200}^{5}}{C_{200}^{5}} & \text{当 200 件中没有次品时}\end{cases}$$

显然

$$P\{\text{无次品}\}\geqslant\frac{C_{198}^{5}}{C_{200}^{5}}=\frac{198\times197\times\cdots\times194}{200\times199\times\cdots\times196}\geqslant0.95$$

于是，任抽 5 件，“出现次品”的概率$\leqslant1-0.95=0.05$. 以上结果表明，如果次品率$\leqslant0.01$，那么抽 5 件样品，出现次品的机会是很少的，平均在 100 回抽样中，出现不到 5 回. 也就是说，如果 $p\leqslant0.01$ 成立，则在一次抽样中，人们实际上很少遇到出现次品的情形. 然而，现在的事实是，在这一次具体的抽样实践中，竟然发生了这种情形. 这是“不合理”的. 产生这种不合理现象的根源在于假设 $p\leqslant0.01$；因此假设“$p\leqslant0.01$”是不能接受的. 故按国家标准，这批产品不能出厂.

从上面的分析讨论中，可以看到，我们的推理方法有两个特点：

(1) 用了反证法的思想.

为了检验一个“假设”(“$p\leqslant0.01$”)是否成立，我们就先假定这个“假设”是成立的，而看由此会产生什么后果. 如果导致了一个不合理现象的出现，那就表明原先的假定是不正确的，也就是说，“假设”是不能成立的. 因此，我们拒绝这个“假设”. 如果由此没有导出不合理的现象发生，则**不能拒绝**原来的“假设”，称原假设是相容的.

(2) 它又区别于纯数学中的反证法. 因为我们这里的所谓“不合理”，并不是形式逻辑中的绝对的矛盾，而是基于人们在实践中广泛采用的一个原则：小概率事件在一次观察中可以认为基本上

不会发生.

这个原则在我们日常生活中是不自觉地使用的.就以刚才举的产品验收问题来看,每个稍有经验的人都会否定假设"$p \leqslant 0.01$",其原因实际上就是利用了上述原则.

自然会产生这样的问题:概率小到什么程度才能当作"小概率事件"呢?通常把概率不超过 0.05 的事件当作"小概率事件",有时把不超过 0.01(也有把不超过 0.10)的事件当作"小概率事件".

以上讲的关于假设检验的基本思想的两个特点,可以概括成一句话:"概率性质的反证法"①.

以下各节就是把这个基本思想运用到各种类型的问题中去.由于正态随机变量最经常出现,所以,我们重点讨论有关正态总体的假设检验问题.下面 §2 先讲一个正态总体的情形,然后在 §3 中介绍假设检验的某些一般概念与数学描述;§4 中再讲两个正态总体的假设检验;§5 中叙述比率的假设检验(包括一个总体和两个总体的情形),最后在 §6 中讲关于一般概率分布的检验.

§2 一个正态总体的假设检验

设 $X \sim N(\mu, \sigma^2)$,关于它的假设检验问题,主要是下列四种:

① 已知方差 σ^2,检验假设 $H_0: \mu = \mu_0$(μ_0 是已知数).

② 未知方差 σ^2,检验假设 $H_0: \mu = \mu_0$(μ_0 是已知数).

③ 未知期望 μ,检验假设 $H_0: \sigma^2 = \sigma_0^2$($\sigma_0$ 是已知数).

④ 未知期望 μ,检验假设 $H_0: \sigma^2 \leqslant \sigma_0^2$($\sigma_0$ 是已知数).

1.已知方差 σ^2,检验假设 $H_0: \mu = \mu_0$.

先从具体例子谈起.

例 2.1 某车间生产铜丝.铜丝的主要质量指标是折断力大

① 我们在这里对"假设检验"采取初学者易于理解的说法.至于数学上比较确切的陈述,请参看 §3 中"检验法与功效函数".

小.用 X 表示该车间生产的铜丝的折断力.根据过去的资料来看，可以认为 X 服从正态分布，期望是 570 千克力，标准差是 8 千克力.今换了一批原材料，从性能上看，估计折断力的方差不会有什么变化，但不知道折断力的大小和原先有无差别？这个问题就是已知方差 $\sigma^2=8^2$，检验假设 H_0："$\mu=570$".设抽出 10 个样品，测得折断力(千克力)为：578，572，570，568，572，570，570，572，596，584，怎样进行检验？

按 §1 中说的基本思想，我们先提出假设 H_0："$\mu=570$"，看在 H_0 成立的条件下，会不会产生不合理的现象.

在"$\mu=570$"的条件下，$X\sim N(570,8^2)$，设 $X_1,X_2,\cdots,X_n$ 是 X 的一个样本，把它们看成随机变量①，则

$$\bar{X}=\frac{X_1+X_2+\cdots+X_{10}}{10}\sim N(570,8^2/10)$$

于是

$$U=\frac{\bar{X}-570}{\sqrt{8^2/10}}\sim N(0,1)$$

查正态分布表知

$$P\left\{\left|\frac{\bar{X}-570}{\sqrt{8^2/10}}\right|>1.96\right\}=0.05$$

这就是说，事件 $\left\{\left|\frac{\bar{X}-570}{\sqrt{8^2/10}}\right|>1.96\right\}$ 是一个小概率事件.

---

① 我们一般用大写拉丁字母 $X_1,X_2,\cdots,X_n$ 表示随机变量，用相应的小写拉丁字母 $x_1,x_2,\cdots,x_n$ 表示观察值(样本值).类似地，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，后者是前者的值.下面不再声明.但在不引起误会的情况下，有时同一个符号一会儿用来代表随机变量，一会儿又用来表示随机变量取的某个值.

现在根据所给的 10 个样本值计算 $\bar{x}$，知 $\bar{x}=575.2$，且

$$\left|\frac{\bar{x}-570}{\sqrt{8^2/10}}\right|=\frac{(575.2-570)\times 3.16}{8}=2.05$$

这说明小概率事件竟在一次观察中发生了，故认为是不合理的．这就表明原先的假设 $H_0:\mu=570$ 不能成立．习惯上说“折断力的大小和原先有显著性差异”．这就解决了例 2.1 的问题．

例 2.2 根据长期的经验和资料的分析，某砖瓦厂所生产的砖的“抗断强度”X 服从正态分布，方差 $\sigma^2=1.21$，今从该厂所生产的一批砖中，随便抽取 6 块，测得抗断强度（单位：$kg\cdot cm^{-2}$）如下：

$$32.56,29.66,31.64,30.00,31.87,31.03$$

现在问：这一批砖的平均抗断强度可否认为是 32.50？

解 待检验的假设是：$\mu=32.50$．

根据所给的样本值，计算统计量 U．

$$U=\frac{\bar{x}-32.50}{\sqrt{\frac{\sigma^2}{n}}}=\frac{31.13-32.50}{\sqrt{\frac{1.21}{6}}}$$

$$=\frac{-1.37}{1.1}\times\sqrt{6}=\frac{-1.37}{1.1}\times 2.45\approx -3$$

既然 $|U|=3>1.96$，故应否定假设 $H_0:\mu=32.50$．也就是说这批砖的平均抗断强度不能认为是 32.50[①]．

在例 2.1 和例 2.2 中，我们都是把概率为 0.05 的事件当作“小概率事件”．就是说，取的**检验标准**是 $\alpha=0.05$．而有些实际问

---

① 本例中对于 $\alpha=0.05$ 否定了假设“$\mu=32.50$”，实际上利用不等式 $\bar{x}=31.13<32.50$ 还可以否定假设“$\mu\geqslant 32.50$”．因而可以认为这批砖的平均抗断强度比 32.50 公斤/厘米^2 小．

同样地，对于例 2.1，不仅否定了假设“$\mu=570$”，利用样本平均 $\bar{x}=575.2>570$，还可以否定假设“$\mu\leqslant 570$”．这样做的理由请读者自己想一想．

题中，应把检验标准取得更小些，例如取 $\alpha=0.01$. 同一问题，采用不同的检验标准，常常得到不同的结论. 以例 2.1 来说，如果取 $\alpha=0.01$，查正态分布表得到临界值 2.58，即

$$P\left\{\left|\frac{\bar{X}-570}{\sqrt{8^2/10}}\right|>2.58\right\}=0.01$$

从样本值算出

$$\bar{x}=575.2,\left|\frac{\bar{x}-570}{\sqrt{8^2/10}}\right|=2.05$$

就是说概率为 0.01 的“小概率事件”没有发生. 可见，当 $\alpha=0.01$ 时未发现不合理的现象，此时也说假设 H_0 与数据是相容的，简称 H_0 是相容的. 于是我们不否定 H_0. 这与 $\alpha=0.05$ 时的结论不同. 可见，检验的结果依赖于 α 的选择.

检验标准 α 的直观意义在于：把概率不超过 α 的事件当作一次观察时不会发生的“小概率事件”. α 通常取为 0.05，有时也取作 0.01(或 0.10).

α 的意义，还可以解释得更确切些：对于 α，找临界值 λ，满足(在 H_0 成立的假定下)：

$$P\left\{\left|\frac{\bar{X}-570}{\sqrt{8^2/n}}\right|>\lambda\right\}=\alpha \tag{2.1}$$

根据我们的规则，当不等式

$$\left|\frac{\bar{x}-570}{\sqrt{8^2/n}}\right|>\lambda$$

成立时，就否定假设 H_0.

这样下结论当然不能保证绝对不犯错误(请读者注意，我们是通过样本来推断总体的性质，也就是由部分来推断整体. 这本身就决定了：“不能保证绝对不犯错误”). 而从(2.1)看，α 正是犯这样

一种错误的概率；这种把客观上符合假设 H_0 判为不符合假设 H_0，即“以真为假”的错误，称为**第一类错误**. α 就是犯第一类错误的概率；称为**检验标准**或**检验水平**. 自然，我们希望 α 小些. 不过，也还有另一方面的问题.

当不等式

$$\left|\frac{\bar{x}-570}{\sqrt{8^2/n}}\right|\leqslant\lambda$$

成立时，假设 $H_0(\mu=570)$ 是相容的[①]（即未发现什么不合理的现象），我们不能否定 H_0. 这时如果接受假设 H_0 也可能犯错误，因为当 H_0 不成立时，也可能出现满足不等式：

$$\left|\frac{\bar{x}-570}{\sqrt{8^2/n}}\right|\leqslant\lambda$$

的样本值 $x_1, x_2, \cdots, x_n$.

这样一种把不符合 H_0 的总体当作符合 H_0 而加以接受所犯的错误，即所谓犯“以假为真”的错误，叫做犯**第二类错误**. 用 β 表示犯第二类错误的概率. 自然我们也希望越小越好.

遗憾的是，对一定的样本容量 n 来讲，一般而论，α 小时 β 就大，β 小时 α 就大. 因而不能做到 α，β 同时非常的小. 所以问题的正确提法是：α，β 要尽量小些. 对于固定的 α，主要通过增加样本容量来减小 β.

通常取 $\alpha=0.05$ 或 0.01. 样本容量 n 不能太小，n 不能小于 5（最好是 $n\geqslant10$，n 越大越好），否则 β 就会太大了.

---

① 这里以及下面的各种假设检验问题中，遇到这种相容的情形，应如何对待假设 H_0 呢？在实际工作需要我们迅速作出明确表态的情况下，常常采取接受假设 H_0 的态度. 有时为了更慎重些，暂不表态，继续进行一些观察（即增加样本容量），再进行检验. 当然，在样本容量较大时，不应该再不表态了.

上面关于例 2.1 所说的话，对于其他类型的假设检验问题也是适用的.

2. 未知方差 σ^2，检验假设 $H_0:\mu=\mu_0$.

这类问题在实际中更常见，§1 中的例 1.2 就是一个代表. 例 1.2 是用仪器间接测量温度，得到 5 个数据：1 250，1 265，1 245，1 260，1 275. 测量值 X 是服从 $N(\mu,\sigma^2)$ 的. 现在根据别的精确方法得到温度的真值是 1 277，我们的问题是，这台仪器测量温度有无系统偏差？

前面说过，这就是检验假设 $H_0:\mu=1\ 277$ 是否成立的问题. 注意，σ^2 等于多少不知道(即仪器的精度不知道)，怎么办呢？

记样本为 $X_1,X_2,\cdots,X_5$，很自然想到，用 σ^2 的估计量

$$S^2=\frac{1}{5-1}\sum_{i=1}^{5}(X_i-\bar{X})^2$$

代替 σ^2，选用样本的函数

$$T=\frac{\bar{X}-1\ 277}{\sqrt{S^2/5}}$$

作为检验的统计量. 从第五章我们知道，当 $X\sim N(1\ 277,\sigma^2)$ 时，

$$\frac{\bar{X}-1\ 277}{\sqrt{S^2/5}}$$

服从 4 个自由度的 t 分布. 换句话说，如果假设 $H_0:\mu=1\ 277$ 成立，则统计量 T 的概率分布能够求出来，而且是有表可查的 t 分布.

查附表 2，知

$$P\left\{\left|\frac{\bar{X}-1\ 277}{\sqrt{S^2/5}}\right|>2.776\right\}=0.05$$

换句话说，如果 H_0 成立，则事件 $\left\{\left|\frac{\bar{X}-1\ 277}{\sqrt{S^2/5}}\right|>2.776\right\}$ 是一个概率为 0.05 的小概率事件.

根据所给的样本值，可算得

$$\bar{x}=\frac{1\ 250+\cdots+1\ 275}{5}=1\ 259$$

$$\sqrt{\frac{S^2}{5}}=\sqrt{\frac{570}{5\times 4}}=\sqrt{28.5}=5.339$$

于是

$$\left|\frac{\bar{x}-1\ 277}{\sqrt{S^2/5}}\right|=\left|\frac{-18}{5.339}\right|>\frac{18}{6}=3>2.776$$

这说明，所给的样本值竟使“小概率事件”发生了，这是不合理的.产生这个不合理现象的根源在于假定了 H_0 是成立的.故应否定假设 H_0.换句话说，该仪器间接测量有系统偏差.

我们把上列检验方法加以总结和概括，得到：

当正态总体的方差未知时，关于期望的检验程序：

(1) 提出待检验的假设 $H_0:\mu=\mu_0$（μ_0 已知）

(2) 根据样本值 $x_1,x_2,\cdots,x_n$ 计算统计量

$$T=\frac{\bar{x}-\mu_0}{\sqrt{S^2/n}}$$

的数值.

(3) 对于检验水平 α，自由度 $=n-1$，查 t 分布临界值表(附表 2)，得临界值 λ.

(4) 将 $|T|$ 与 λ 进行比较，作出判断.[当 $|T|>\lambda$ 时拒绝 H_0；当 $|T|\leqslant\lambda$ 时，H_0 是相容的(此时常接受 H_0).]

关于正态总体方差已知时期望的检验程序，请读者自己给出.

例 2.3 根据长期资料的分析，知道某种钢生产出的钢筋的强度服从正态分布.今随机抽取六根钢筋进行强度试验，测得强度为(单位：$kg\cdot mm^{-2}$)：

48.5,49.0,53.5,49.5,56.0,52.5

问：能否认为该种钢生产的钢筋的平均强度为 52.0?

解 用 X 表示钢筋强度，$X\sim N(\mu,\sigma^2)$.

(1) 要检验的假设是 $H_0:\mu=52.0$.

(2) 计算统计量 T 的值.算得 $\bar{x}=51.5, S^2=\frac{44.50}{5}$,

$$T=\frac{\bar{x}-52.0}{\sqrt{S^2/n}}\approx -0.41$$

(3) 查附表 2,$\alpha=0.05$,自由度 $=6-1=5$,得 $\lambda=2.571$.

(4) 下判断.现在 $|T|\approx 0.41<2.571$,故 H_0 是相容的.即不能否定钢筋的平均强度为 52.0 $kg\cdot mm^{-2}$.

顺便指出,这个检验与 $E(X)$的区间估计之间有密切的联系:

由上一章 §5 的(5.5′)式知,$E(X)$的置信度为 $1-\alpha$ 的置信区间是满足不等式

$$\left|\frac{\bar{x}-\mu}{\sqrt{S^2/n}}\right|\leqslant\lambda$$

的 μ 值的集合.因此"$H_0:\mu=\mu_0$"的检验等价于下述检验:找出总体均值的置信区间,如果 μ_0 不在置信区间中$\left(\text{亦即}\left|\frac{\bar{x}-\mu_0}{\sqrt{S^2/n}}\right|>\lambda\right)$,则拒绝 H_0;否则 H_0 就是相容的.

(读者试将上面例 1.2 的假设检验与上章 §5 例 5.2 的区间估计进行比较.)

3.未知期望 μ,检验假设 $H_0:\sigma^2=\sigma_0^2$.

我们还是从一个具体例子谈起.我们把例 2.1 的提法改变一下.

例 2.4 某车间生产铜丝,生产一向比较稳定,今从产品中随便抽出 10 根检查折断力,得数据如下(单位:千克力):

578,572,570,568,572,570,570,572,596,584

问:是否可相信该车间的铜丝的折断力的方差为 64?

用 X 表示铜丝的折断力,当然 $X\sim N(\mu,\sigma^2)$,我们的任务就是根据上述 10 个样本值,来检验假设 $H_0:\sigma^2=64$.

很自然想到，看 σ^2 的估计量 S^2 有多大. 如果 $\frac{S^2}{64}$ 很大或很小，则应该否定 H_0，为了数学上处理方便，我们取统计量

$$W=\frac{\sum_{i=1}^{10}(X_i-\bar{X})^2}{64}$$

注意 $W=9\cdot\frac{S^2}{64}$. 显然 W 很大或很小时，应该否定 H_0.

从第五章知道，如果 H_0 成立则 W 服从 9 个自由度的 χ^2 分布. 和区间估计时的情况一样，通过查 χ^2 分布的临界值表，找到 λ_1,λ_2 满足：

$$P\{W<\lambda_1\}=0.025$$

$$P\{W>\lambda_2\}=0.025$$

于是事件 $\{W<\lambda_1$ 或 $W>\lambda_2\}$ 是小概率事件.

现在根据所给的样本值，计算统计量 W 的数值：

$$\bar{x}=575.2,\ \sum_{i=1}^{10}(x_i-\bar{x})^2=681.6$$

故

$$W=\frac{681.6}{64}=10.65$$

查 χ^2 分布表，得 $\lambda_1=2.70,\lambda_2=19.0$. 现在 $\lambda_1=2.70<10.65<19.0=\lambda_2$，故下结论：假设 $H_0:\sigma^2=64$ 是相容的.

（如果算得 W 的值比 $\lambda_1=2.70$ 小或比 $\lambda_2=19.0$ 大，则否定假设 H_0.）

4. 未知期望 μ，检验假设 $H_0:\sigma^2\leqslant\sigma_0^2$.

这种情况在实际应用中比 3 更重要. 生产中为了了解加工精度有无变化，进行抽样，如算得样本方差 S^2 比原先的方差 σ_0^2 大，这时可检验假设 $H_0:\sigma^2\leqslant\sigma_0^2$. 经过检验，如能否定 H_0，说明精度变差了，须停产检查原因.

我们来分析一下这个问题. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本.

很自然想到,如果 $\frac{S^2}{\sigma_0^2}$ 很大,则有理由否定假设:$\sigma^2 \leqslant \sigma_0^2$,否则,可以接受这个假设.

但在假设 $\sigma^2 \leqslant \sigma_0^2$ 成立的条件下,比值

$$\frac{S^2}{\sigma_0^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)\sigma_0^2}$$

的概率分布并不能算出来. 因而我们遇到前所未有的困难. 怎么办呢? 命

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

我们从第五章知道(可参看附录二定理 5 的系),Y 服从 $n-1$ 个自由度的 χ^2 分布,于是可找到 λ 满足:

$$P\{Y > \lambda\} = \alpha$$

于是 $\left\{ \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} > \lambda \right\}$ 是一个"小概率事件". 可惜的是,$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}$ 算不出来(因不知 σ 等于多少). 但是在假设 $\sigma^2 \leqslant \sigma_0^2$ 之下,有不等式:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_0^2} \leqslant \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}$$

因此

$$P\left\{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma_0^2}>\lambda\right\}\leqslant P\left\{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}>\lambda\right\}=\alpha$$

这就表明，事件 $\left\{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma_0^2}>\lambda\right\}$ 更是一个“小概率事件”.

如果根据所给的样本值 $x_1,x_2,\cdots,x_n$，算出

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}>\lambda$$

则应该否定假设“$\sigma^2\leqslant\sigma_0^2$”.

如果 $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}\leqslant\lambda$，则假设“$\sigma^2\leqslant\sigma_0^2$”是相容的.

现将上面 3、4 两段所讨论的内容加以总结，得到下列关于正态总体的方差 σ^2 的假设检验程序：

(1) 提出待检验的假设 $H_0:\sigma^2=\sigma_0^2$（或 $\sigma^2\leqslant\sigma_0^2$）.

(2) 计算统计量

$$W=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}$$

的数值.

(3) 查 χ^2 分布临界值表（附表 3），注意自由度 $=n-1$，得 λ_1,λ_2（或 λ）满足

$$P\{\chi^2<\lambda_1\}=P\{\chi^2>\lambda_2\}=\frac{\alpha}{2}$$

（或 $P\{\chi^2>\lambda\}=\alpha$）

其中 α 是检验水平.

(4) 比较 W 与 λ_1,λ_2(或 λ)的值,作出判断.

作为本节的末尾,我们再举一个单边检验的例子.

例 2.5 已知罐头番茄汁中,维生素 C(Vc)含量服从正态分布 .按照规定,Vc 的平均含量不得少于 21 毫克.现从一批罐头中抽了 17 罐,算得 Vc 含量的平均值 $\bar{x}=23, S^2=3.98^2$,问该批罐头 Vc 含量是否合格?

解 Vc 含量 $X\sim N(\mu,\sigma^2)$.我们来检验假设 H_0:

$$\mu<21.$$

如果能否定 H_0,则可以认为 $\mu\geqslant 21$,从而该批罐头 Vc 含量合格.

现来分析一下这个假设 H_0 如何进行检验.设 $X_1,\cdots,X_n$ 是来自 X 的样本(现在 $n=17$),如果 H_0 成立,即 $\mu<21$,则有

$$\frac{\bar{X}-21}{\sqrt{\frac{S^2}{n}}}<\frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}}$$

于是

$$P\left\{\frac{\bar{X}-21}{\sqrt{S^2/n}}>\lambda\right\}\leqslant P\left\{\frac{\bar{X}-\mu}{\sqrt{S^2/n}}>\lambda\right\} \tag{2.2}$$

但 $\dfrac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}}$ 服从 $n-1$ 个自由度的 t 分布,查附表 2 知

$$P\left\{\left|\frac{\bar{X}-\mu}{\sqrt{S^2/n}}\right|>1.746\right\}=0.10\ .$$

从而

$$P\left\{\frac{\bar{X}-\mu}{\sqrt{S^2/n}}>1.746\right\}=0.05$$

从(2.2)知

$$P\left\{\frac{\bar{X}-21}{\sqrt{S^2/n}}>1.746\right\}\leqslant 0.05$$

这表明，如果 H_0 成立，则 $\left\{\frac{\bar{X}-21}{\sqrt{S^2/n}}>1.746\right\}$ 是一个小概率事件.

现将 $\bar{x}=23, S^2=3.98^2$ 代入，有

$$\frac{\bar{x}-21}{\sqrt{\frac{S^2}{n}}}=2.07>1.746$$

这个小概率事件竟发生了，于是否定 H_0. 从而认为该批罐头合格.

以上我们讨论了一个正态总体的假设检验问题，这是实际工作中碰到比较多的情形. 至于非正态总体的情形，除了运用其他检验方法以外，当样本容量 n 很大时，可以这样近似地考虑，把它形式上当作正态总体来处理，检验的方法和步骤和本节讲的完全一样. 样本容量 n 一般不得小于 30，最好是 50 以上，或 100 以上. 这样做的理论根据，是基于概率论中的“中心极限定理”，详细情况我们不说了.

* §3 假设检验的某些概念和数学描述

本节对假设检验问题及有关的基本概念进行概括的数学描述. 由于叙述较抽象，初学者不必阅读.

首先介绍检验法与功效函数，然后介绍临界值与 p 值，最后介绍假设检验与置信区间的联系.

1. 检验法与功效函数

把我们要检验的“假设”记作 H_0（有时叫做零假设）. H_0 是关于随机变量 X（总体）的分布的一个“看法”. 说得更确切些，设 X 的分布函数为 $F(x,\theta)$，其中 θ 属于 Θ，这里 Θ 是实数（或向量，或其他符号）组成的已知集合. “看法” H_0 通常可表示成这样的形式：$\theta\in\Theta_0$，这里 Θ_0 是 Θ 的非空子集，且 $\Theta_0\neq\Theta$. 通常也把“$\theta\in\Theta-\Theta_0$”叫做对立假设（或叫做备择假设），记作 H_a.

例如，在例 2.1 中，$X\sim N(\mu,8^2)$，$\mu=\theta$，$\Theta=(-\infty,+\infty)$，$\Theta_0=\{570\}$，

$\Theta - \Theta_0 = (-\infty, 570) \cup (570, +\infty)$. 要检验的假设"$\mu = 570$"是"$\theta \in \Theta_0$".

怎样根据样本值对 H_0 进行检验呢？这就需要对"检验法"给出合理的定义. 直观上说，所谓一个检验法，就是给出一个规则，对给定的样本值 $x_1, x_2, \cdots, x_n$ 进行明确表态：接受假设 H_0 还是拒绝假设 H_0.

这一点用数学语言可以说得更清楚些. 设 S 是所有可能的样本值$(x_1, x_2, \cdots, x_n)$(n 固定)组成的集合(样本空间)，不失一般性，常设 $S = R^n$(我们讨论连续型随机变量的情形). 所谓一个检验法就是指空间 S 的一个划分：$S = S_1 \cup S_2$(这里 S_1 与 S_2 无共同元素). 当$(x_1, x_2, \cdots, x_n) \in S_1$ 时，接受假设 H_0；当$(x_1, x_2, \cdots, x_n) \in S_2$ 时，拒绝 H_0. 这 S_1 叫接受域，S_2 叫否定域. 因为 $S_1 = S - S_2$，故只要知道了否定域，就知道了检验法. 每个检验法对应一个否定域；反之，任给定 S 的一个子集 W，则有惟一的检验法以 W 作为它的否定域. 故研究检验法就相当于研究否定域.

S 中的子集太多了，因而否定域多得很①. 究竟应该选哪一个对于检验 H_0 是最合适的呢？这就涉及到优良性的标准. 为了分析这个问题，我们看看在取定一个否定域 W(即一个检验法)之后，有什么后果.

零假设 H_0 在客观上只有两种可能性：真、假. 样本值$(x_1, x_2, \cdots, x_n)$只有两种可能性：属于否定域 W、不属于 W. 若采用 W 作否定域，则在观察样本值$(x_1, x_2, \cdots, x_n)$时只可能有下列四种情况：

① H_0 真，但$(x_1, x_2, \cdots, x_n)$属于 W；

② H_0 真，但$(x_1, x_2, \cdots, x_n)$不属于 W；

③ H_0 假，但$(x_1, x_2, \cdots, x_n)$属于 W；

④ H_0 假，但$(x_1, x_2, \cdots, x_n)$不属于 W.

根据我们的规则，在情形①、③应拒绝 H_0，在情形②、④应接受 H_0. 情形②、③当然很好，对 H_0 的表态与客观实际相符. 但在①、④两种情形下，表态犯了错误：与客观实际不相符.

在情形①下出现的错误就是所谓第一类错误(它把本来真实的看法 H_0 进行了否定)，在情形④下出现的错误是所谓第二类错误(它把本来虚假的看

---

① 为了便于数学处理，通常要求否定域是 n 维空间中的所谓"Borel 集". 学过实变函数论的读者知道，Borel 集是个很广的概念，我们通常遇到的规则集合都是 Borel 集.

法 H_0 接受下来).由于样本取值有随机性,这两种错误一般难以避免.为了描述这两类错误出现的机会,需要使用一些记号.

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本,显然当且仅当事件 $\{(X_1, X_2, \cdots, X_n) \in W\}$ 发生时拒绝假设 H_0.这个事件的概率当然与参数 θ 有关.当总体的分布函数是 $F(x,\theta)$时,我们把这个事件的概率记作 $M_W(\theta)$,它是参数集 Θ 上处处有定义的函数,通常叫做否定域(检验法)W 的功效函数(或叫做势函数).

容易看出,当 $\theta \in \Theta_0$ 时,$M_W(\theta)$表示犯第一类错误的概率;当 $\theta \in \Theta_1$ 时,$1 - M_W(\theta)$表示犯第二类错误的概率.我们希望犯两类错误的概率越小越好,也就是说,希望找到这样的否定域 W,当 $\theta \in \Theta_0$ 时 $M_W(\theta)$的值很小很小,当 $\theta \in \Theta_1$ 时 $M_W(\theta)$很接近于 1.

由此可见,功效函数 $M_W(\theta)$是用来刻画否定域的优良程度的.通常用

$$\alpha_W(\theta) = M_W(\theta) \qquad (\theta \in \Theta_0)$$

$$\beta_W(\theta) = 1 - M_W(\theta) \qquad (\theta \in \Theta_1)$$

表示犯两类错误的概率的大小.

定义 3.1 给定小数 $\alpha(0<\alpha<1)$,如果对一切 $\theta \in \Theta_0$,$\alpha_W(\theta) \leqslant \alpha$,则称 W 的检验水平(也称显著性水平)是 α.[①]也称 W 的检验水平不超过 α.

在实际工作中常常这样提出问题:在所有检验水平是 α 的否定域中,如何找出犯第二类错误的概率尽可能小的否定域来?

这个问题是比较复杂的,需要对总体 X 的概率性质以及零假设 H_0 的具体结构作具体分析.在许多情形下已有很好的答案,本章中介绍的一些检验法都是比较好的.还有许多情形尚未研究清楚或答案令人不够满意.

我们还要强调一下,在回答这个问题时,即使找到了犯第二类错误概率最小的否定域,也并不表明这个否定域犯第二类错误的概率一定很小.显然这个概率还和样本容量 n 有关,n 取得大,它就会小.正是由于这个缘故,"零假设 H_0"与"备择假设 H_a"的地位是不对称的.对于给定的小正数 α,若 W 是检验水平是 α 的否定域,当样本值落入 W 时拒绝 H_0 是有力的,因为此

---

① 上确界 $\sup_{\theta \in \Theta_0} \alpha_W(\theta)$叫做 W 的精确检验水平.即零假设成立时拒绝零假设的最大概率.