概率论与数理统计第19讲

本讲义可在网址http://math.shekou.com

或

ftp://math.shekou.com

下载

§ 5.3 抽样分布

一, 抽样分布

有时, 总体分布的类型虽然已知, 但其中含有未知参数, 此时需对总体的未知参数或对总体的数字特征(如数学期望, 方差等)进行统计推断, 此类问题称为**参数统计推断**. 在参数统计推断问题中, 常需利用总体的样本构造出合适的统计量, 并使其服从或渐近地服从已知的分布. 统计学中泛称统计量分布为**抽样分布**.

讨论抽样分布的途径有两个. 一是精确地求出抽样分布, 并称相应的统计推断为**小样本统计推断**; 另一种方式是让样本容量趋于无穷, 并求出抽样分布的极限分布. 然后, 在样本容量充分大时, 再利用该极限分布作为抽样分布的近似分布, 进而对未知参数进行统计推断, 称与此相应的统计推断为**大样本统计推断**. 这里重点讨论正态总体的抽样分布, 属小样本统计范畴, 此外, 也简要介绍一般总体的某些抽样分布的极限分布, 属大样本统计范畴.

二, 单正态总体的抽样分布

设总体X的均值为μ, 方差为σ^2, $X_1, X_2, \cdots, X_n$是取自X的一个样本，$\bar{X}$与S^2分别为该样本的样本均值与样本方差, 则有

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu$$

$$D(\bar{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) =$$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

而

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

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)\right]$$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right]$$

$$=\sigma^2$$

定理1 设总体$X\sim N(\mu,\sigma^2)$, $X_1,X_2,\cdots,X_n$是取自X的一个样本，$\bar{X}$与S^2分别为该样本的样本均值与样本方差, 则有

$$(1)\ \bar{X}\sim N(\mu,\sigma^2/n); \qquad (3.1)$$

$$(2)\ U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1). \qquad (3.2)$$

定理2 设总体$X\sim N(\mu,\sigma^2)$, $X_1,X_2,\cdots,X_n$是取自X的一个样本，$\bar{X}$与S^2分别为该样本的样本均值与样本方差, 则有

(1)

$$\chi^2=\frac{n-1}{\sigma^2}S^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\sim\chi^2(n-1)\qquad(3.3)$$

(2) $\bar{X}$与S^2相互独立.

(证略)

定理3 设总体$X \sim N(\mu,\sigma^2)$, $X_1,X_2,\cdots,X_n$是取自X的一个样本，$\bar{X}$与S^2分别为该样本的样本均值与样本方差, 则有

$$(1)\quad \chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n) \qquad (3.4)$$

$$(2)\ T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1). \qquad (3.5)$$

证明 结论(1)是χ^2分布定义的直接推论.

对结论(2), 前面已知

$$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1) \qquad \frac{n-1}{\sigma^2}S^2\sim\chi^2(n-1)$$

由t分布的定义有

$$U\bigg/\sqrt{\frac{n-1}{\sigma^2}S^2\bigg/n-1}=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\bigg/\sqrt{\frac{n-1}{\sigma^2}S^2\bigg/n-1}$$

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

例1 设$X \sim N(21,2^2)$, $X_1, X_2, \cdots, X_{25}$为X的一个样本, 求

(1) 样本均值 $\overline{X}$的数学期望与方差;

(2) $P\{|\overline{X}-21| \le 0.24\}$.

解 (1) 因为 $E(\overline{X}) = 21, D(\overline{X}) = 2^2/25 = 0.4^2$.

$$\overline{X} \sim N\left(21, 0.4^2\right)$$

(2)

$$P\{|\overline{X}-21| \le 0.24\} = P\left\{\left|\frac{\overline{X}-21}{0.4}\right| \le 0.6\right\}$$

$$= 2\Phi(0.6) - 1 = 0.4504.$$

例 2 假设某物体的实际重量为μ，但它是未知的. 现在用一架天平去称它，共称了 n 次，得到 $X_1, X_2, \cdots, X_n$. 假设每次称量过程彼此独立且没有系统误差，则可以认为这些测量值都服从正态分布 $N(\mu, \sigma^2)$，方差σ^2 反映了天平测量过程的总精度，通常我们用样本均值 $\overline{X}$ 去估计μ，根据定理 1，$\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

再从正态分布的 3σ性质知

$$P\left\{|\bar{X}-\mu|<\frac{3\sigma}{\sqrt{n}}\right\}\geq 99.7\%.$$

这就是说，我们的估计值$\bar{X}$与真值μ的偏差不超过$3\sigma/\sqrt{n}$的概率为 99.7%，并且随着称量次数 n 的增加，这个偏差界限$3\sigma/\sqrt{n}$愈来愈小.

例如若σ=0.1, n=10. 则

$$P\left\{|\bar{X}-\mu|<\frac{3\times0.1}{\sqrt{10}}\right\}\approx P\{|\bar{X}-\mu|<0.09\}\geq 99.7\%$$

于是我们以99.7%的概率断言，$\bar{X}$与物体真正重量μ的偏差不超过0.09. 如果将称量次数n增加到100, 则

$$P\left\{|\bar{X}-\mu|<\frac{3\times0.1}{\sqrt{100}}\right\}\approx P\{|\bar{X}-\mu|<0.03\}\geq 99.7\%$$

这时我们以同样的概率断言，$\bar{X}$与物体真正重量μ的偏差不超过0.03.

例3 在设计导弹发射装置时, 重要事情之一是研究弹着点偏离目标中心的距离的方差. 对于一类导弹发射装置, 弹着点偏离目标中心的距离服从正态分布$N(\mu,\sigma^2)$, 这里σ^2=100米^2, 现在进行了25次发射试验, 用S^2记这25次试验中弹着点偏离目标中心的距离的样本方差, 试求S^2超过50米^{2}的概率.

解 因为 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$

于是

$$P\{S^2 > 50\} = P\left\{\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)50}{\sigma^2}\right\}$$

$$= P\left\{\chi^2(24) > \frac{24 \times 50}{100}\right\}$$

$$= P\{\chi^2(24) > 12\} > P\{\chi^2(24) > 12.401\}$$

$$= 0.975.$$

于是我们以超过97.5%的概率断言, S^2超过50米^2.

例4 从正态总体$N(\mu, 0.5^2)$中抽取容量为10的样本$X_1, X_2, \ldots, X_n$. $\bar{X}$是样本的均值. 若μ未知, 计算概率

$$P\left\{\sum_{i=1}^{10}(X_i-\mu)^2 \geq 1.68\right\}$$

$$P\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2 < 2.85\right\}$$

[分析] 计算与随机变量有关的事件的概率, 必须知道该随机变量的分布.

解 若μ未知, 由于$X_i \sim N(\mu, 0.5^2)$, 所以

$$\frac{X_i-\mu}{0.5} \sim N(0,1).$$

$$\sum_{i=1}^{10}\left(\frac{X_i-\mu}{0.5}\right)^2 = 4\sum_{i=1}^{10}(X_i-\mu)^2 \sim \chi^2(10).$$

$$4\sum_{i=1}^{10}(X_i-\bar{X})^2 \sim \chi^2(9).$$

故

$$P\left\{\sum_{i=1}^{10}(X_i-\mu)^2 \geq 1.68\right\} = P\left\{4\sum_{i=1}^{10}(X_i-\mu)^2 \geq 6.72\right\}$$

查χ^2分布表知: $\chi^2_{0.75}(10) = 6.737$

所以

$$P\left\{\sum_{i=1}^{10}(X_i-\mu)^2 \geq 1.68\right\} = 0.75;$$

$$P\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2<2.85\right\}=1-P\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 2.85\right\}$$

$$=1-P\left\{4\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 11.4\right\}$$

查χ^2分布表知， $\chi^2_{0.25}(9)=11.4,$

所以

$$P\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2<2.85\right\}=1-0.25=0.75$$

例5 从正态总体$X\sim N(\mu,\sigma^2)$中抽取容量为16的一个样本，$\bar{X}, S^2$分别为样本均值和样本方差, 若μ, σ^2均未知, 求S^2的方差$D(S^2)$及概率

$$(1)\ P\left\{\frac{S^2}{\sigma^2}\le 2.041\right\};$$

$$(2)\ P\left\{\frac{\sigma^2}{2}\le \frac{1}{16}\sum_{i=1}^{16}(X_i-\bar{X})^2\le 2\sigma^2\right\};$$

$$(3)\ P\left\{\frac{\sigma^2}{2}\le \frac{1}{16}\sum_{i=1}^{16}(X_i-\mu)^2\le 2\sigma^2\right\}$$

解

$$\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$$

$$D\left(\frac{n-1}{\sigma^2}S^2\right) = 2(n-1),$$

$$D(S^2) = \frac{2\sigma^4}{n-1} = \frac{2}{15}\sigma^4$$

$$(1)\ P\left\{\frac{S^2}{\sigma^2}\le 2.041\right\}=P\left\{\frac{15S^2}{\sigma^2}\le 30.615\right\}$$

$$=1-P\left\{\frac{15S^2}{\sigma^2}>30.615\right\}=1-0.01=0.99$$

$$(\chi^2_{0.01}(15)=30.578)$$

$$(2)\ P\left\{\frac{\sigma^2}{2} \le \frac{1}{16}\sum_{i=1}^{16}(X_i - \bar{X})^2 \le 2\sigma^2\right\}$$

$$= P\{8 \le \chi^2(15) \le 32\}$$

$$= P\{\chi^2(15) \ge 8\} - P\{\chi^2(15) \ge 32\}$$

$$= 0.95 - 0.005 = 0.945.$$

(3) 由于 $\dfrac{X_i-\mu}{\sigma}\sim N(0,1),\ (i=1,2,\cdots,n)$

$$\sum_{i=1}^{16}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(16),$$

$$P\left\{\frac{\sigma^2}{2}\le\frac{1}{16}\sum_{i=1}^{16}(X_i-\mu)^2\le 2\sigma^2\right\}$$

$$=P\{8\le\chi^2(16)\le 32\}$$

$$=P\{\chi^2(16)\ge 8\}-P\{\chi^2(16)>32\}$$

$$=0.95-0.01=0.94.$$

三, 双正态总体的抽样分布

定理 4 设$X \sim N(\mu_1, \sigma_1^2)$与$Y \sim N(\mu_2, \sigma_2^2)$是两个相互独立的正态总体, 又设$X_1, \ldots, X_{n_1}$是取自总体$X$的样本, $\bar{X}$与S_1^2分别为该样本的样本均值与样本方差. $Y_1, \ldots, Y_{n_2}$是取自总体Y的样本, $\bar{Y}$与S_2^2分别为此样本的样本均值与样本方差. 再记S_w^2是S_1^2与S_2^2的加权平均, 即

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}. \qquad (3.6)$$

则

$$(1)\ U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1); \qquad (3.7)$$

$$(2)\ F=\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1); \qquad (3.8)$$

(3) 当$\sigma_1{}^2=\sigma_2{}^2=\sigma^2$时,

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2). \quad (3.9)$$

证明 (1) 因

$$\bar{X} \sim N(\mu_1, \sigma_1^2 / n_1), \bar{Y} \sim N(\mu_2 \sigma_2^2 / n_2).$$

相互独立, 故

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)$$

即

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \sim N(0,1)$$

(2) 因

$$\frac{n_1-1}{\sigma_1^2}S_1^2\sim\chi^2(n_1-1),\frac{n_2-1}{\sigma_2^2}S_2^2\sim\chi^2(n_2-1).$$

且相互独立, 因此有

$$F=\frac{\left.\frac{n_1-1}{\sigma_1^2}S_1^2\right/(n_1-1)}{\left.\frac{n_2-1}{\sigma_2^2}S_2^2\right/(n_2-1)}=\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)$$

(3) 当$\sigma_1^2=\sigma_2^2=\sigma^2$时, 由(1)知

$$U_1=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/n_1+1/n_2}}\sim N(0,1);$$

$$V=\left(\frac{n_1-1}{\sigma^2}S_1^2\right)+\left(\frac{n_2-1}{\sigma^2}S_2^2\right)\sim\chi^2(n_1+n_2-2).$$

则

$$\frac{U_1}{\sqrt{\dfrac{V}{n_1+n_2-2}}}\sim t(n_1+n_2-2).$$

整理后即得结论.

例6 设两个正态总体X与Y都服从正态分布$N(20,3)$. 今从总体X与Y中分别抽得容量$n_1=10$, $n_2=15$的两个相互独立的样本, 求$P\{|\bar{X}-\bar{Y}|>0.3\}$.

解 由题设知

$$\frac{(\bar{X}-\bar{Y})-(20-20)}{\sqrt{\frac{3}{10}+\frac{3}{15}}}=\frac{\bar{X}-\bar{Y}}{\sqrt{0.5}}\sim N(0,1).$$

$$\frac{\bar{X}-\bar{Y}}{\sqrt{0.5}} \sim N(0,1).$$

于是

$$P\{|\bar{X}-\bar{Y}|>0.3\}=1-P\left\{\left|\frac{\bar{X}-\bar{Y}}{\sqrt{0.5}}\right| \le \frac{0.3}{\sqrt{0.5}}\right\}$$

$$=1-\left[2\Phi\left(\frac{0.3}{\sqrt{0.5}}\right)-1\right]$$

$$=2-2\Phi(0.42)=0.6744.$$

例 7 设总体 X 和 Y 相互独立且都服从正态分布 $N(30,3^2)$; $X_1,\ldots,X_{20}$; $Y_1,\ldots,Y_{25}$ 分别来自总体 X 和 Y 的样本，$\bar{X},\bar{X},S_1^2,S_2^2$分别是这两个样本的均值和方差．求$P\{S_1^2/S_2^2 \le 0.4\}$.

解 因$\sigma_1^2=\sigma_2^2=3^2$,则$\dfrac{S_1^2}{S_2^2}\sim F(20-1,25-1)$,

$$P\left\{\frac{S_1^2}{S_2^2}\le 0.4\right\}=P\left\{\frac{S_2^2}{S_1^2}\ge\frac{1}{0.4}\right\}$$

$$=P\{F(24,19)>2.5\}\approx 0.025$$

四, 一般正态总体抽样分布的极限分布

对于一般总体, 无论其服从什么分布(离散的或者连续的), 只要样本容量n足够大根据中心极限定理, 它的样本均值 $\bar{X}$都近似服从正态分布. 即近似有

$$U_n = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$T_n = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim N(0,1)$$

课堂练习

1. 设$X_1, X_2, \ldots, X_{15}$为正态总体$N(0,3^2)$的一个样本，$\overline{X}$为样本均值, 求:

$$P\left\{36.65 \le \sum_{i=1}^{15}(X_i - \overline{X})^2 \le 235\right\}.$$

2. 设$X_1,X_2,\ldots,X_n$为总体$X\sim N(\mu,\sigma^2)$的一个样本，$\bar{X}$和S^2为样本均值和样本方差. 又设新增加一个试验量X_n+1, 而且X_n+1与$X_1,\ldots,X_n$也相互独立, 求统计量

$$U=\frac{X_{n+1}-\bar{X}}{S}\sqrt{\frac{n}{n+1}}$$

的分布.

作业 习题5-3 第183页开始
第2,3,9题