西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第五章 数理统计的基本概念与抽样分布

第三节 抽样分布

- 一、问题的提出
- 二、抽样分布定理

一、问题的提出

总体 X： 随机变量 $\sim E(X), D(X), F(x,\lambda)$

例如：成绩、温度、时间、质量、长度

估计

样本 $(X_1, X_2, \cdots, X_n)$： n维随机向量

例如：（第1个学生成绩，第2个学生成绩…第n个学生成绩）

（第1天温度，第2天温度…第n天温度）

样本观测值：$(86, 93, \cdots, 65, 72)$

统计量 $f(X_1, X_2, \cdots, X_n)$：随机变量

抽样分布：统计量的分布 $f(X_1, X_2, \cdots, X_n) \sim$ 分布？

抽样分布 $\begin{cases} \text{精确抽样分布} & \text{(小样本问题中使用)} \\ \text{渐近分布} & \text{(大样本问题中使用)} \end{cases}$

例如：

$\overline{X}$ 平均成绩、平均温度…

S_n^2 成绩的方差、温度的方差…

这一节，我们来讨论正态总体的抽样分布.

二、抽样分布定理

1. 样本来自单个正态总体

定理5.3 设样本$(X_1, X_2, \cdots, X_n)$是来自总体X，而

$$X \sim N(\mu, \sigma^2),$$

则 (1) 样本均值

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n),$$

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

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

或

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

标准化样本均值

$X \sim N(\mu,\sigma^2)$, $\Rightarrow X_i \sim N(\mu,\sigma^2)$, $\overline{X} \sim N(\mu,\sigma^2/n)$,

$\boxed{令 Y_i = X_i - \mu} \sim N(0,\sigma^2)$

$\boxed{令 \overline{Y} = \overline{X} - \mu} \sim N(0,\sigma^2/n)$

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}Y_i^2 - \overline{Y}^2$$

$\boxed{令 P_i = \dfrac{Y_i}{\sigma}} \sim N(0,1)$

$$= \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{\sigma}\right)^2 - \frac{\sigma^2}{n}\left(\frac{\overline{Y}}{\sigma/\sqrt{n}}\right)^2$$

$\boxed{令 Q = \dfrac{\overline{Y}}{\sigma/\sqrt{n}}} \sim N(0,1)$

$$\therefore \frac{S_n^2}{\sigma^2/n} = \sum_{i=1}^{n}P_i^2 - Q^2 \sim \chi(n-1)$$

(2) $$V=\frac{nS_n^2}{\sigma^2}=\frac{(n-1)S_n^{*2}}{\sigma^2}$$

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

其中S_n^2是样本方差.

注 1° 减少一个自由度的原因:

$$\{\frac{X_i-\overline{X}}{\sigma}\}(i=1,2,\cdots,n)$$不相互独立.

事实上，它们受到一个条件的约束：

$$\boxed{\sum_{i=1}^{n} X_i = n\overline{X}}$$

$$\therefore \sum_{i=1}^{n} \frac{X_i - \overline{X}}{\sigma} = \frac{1}{\sigma}\left(\sum_{i=1}^{n} X_i - n\overline{X}\right) = \frac{1}{\sigma}\cdot 0 = 0.$$

(3) $\overline{X}$ 与 S_n^2 独立.

注 2° 若X不服从正态分布，由中心极限定理知，

当$n>>1$ (一般$n\geq 30$)时，

$$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\overset{\text{近似}}{\sim} AN(0,1),$$

其中 $\mu=E(X)$, $\sigma^2=D(X)$.

3° 在实际问题中，总体方差σ^2常常是未知的，若将标准样本均值U中的σ用S_n^*代替，则有如下推论：

（4） 设$(X_1, X_2, \cdots, X_n)$是总体$N(\mu,\sigma^2)$的样本，$\bar{X}, S_n^{*2}$ 分别是样本均值和修正样本方差， 则有

$$T = \frac{\bar{X}-\mu}{S_n^* / \sqrt{n}} = \frac{\bar{X}-\mu}{S_n / \sqrt{n-1}} \sim t(n-1).$$

证 Q $U = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $V = \dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$,

且两者独立, 由 t 分布的定义知

$$T = \frac{U}{\sqrt{\dfrac{V}{n-1}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S_n^{*2}}{\sigma^2(n-1)}} = \frac{\bar{X}-\mu}{S_n^*/\sqrt{n}}$$

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

例1 设X和$Y_1,\cdots,Y_n$分别来自正态总体$N(0,\sigma_1^2)$、$N(\mu_2,\sigma_2^2)$且相互独立的样本，试求

$$F=\frac{X^2\sigma_2^2}{S_n^{*2}\sigma_1^2},\quad 其中S_n^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

的概率分布, 并写出分布参数.

解 由卡方分布的定义有

$$\frac{X^2}{\sigma_1^2}\sim\chi^2(1)$$

$$\frac{(n-1)S_n^{*2}}{\sigma_2^2}=\sum_{i=1}^{n}\left(\frac{Y_i-\bar{Y}}{\sigma_2}\right)^2\sim\chi^2(n-1)$$

又因为X^2与$\dfrac{(n-1)S_n^{*2}}{\sigma_2^2}$相互独立，

由F分布性质知

$$F=\frac{X^2\sigma_2^2}{S_n^{*2}\sigma_1^2}=\frac{\dfrac{X^2}{\sigma_1^2}/1}{\dfrac{(n-1)S_n^{*2}}{\sigma_2^2}/(n-1)}$$

$$\sim F(1,n-1).$$

例2 某厂生产的灯泡使用寿命$X \sim N(2250, 250^2)$现进行质量检查，方法如下：任意挑选若干个灯泡，如果这些灯泡的平均寿命超过2200h，就认为该厂生产的灯泡质量合格，若要使通过检验的概率超过0.997，问至少检查多少只灯泡?

解　以$\bar{X}$记样本均值，则　$\bar{X} \sim N(2250, \dfrac{250^2}{n})$

$$P(\text{合格}) = P(\bar{X} > 2200)$$

$$= P\left(\frac{\bar{X} - 2250}{250/\sqrt{n}} > \frac{2200 - 2250}{250/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(\frac{2200 - 2250}{250/\sqrt{n}}\right) > 0.997$$

即

$$1-\Phi\left(-\frac{\sqrt{n}}{5}\right)$$

$$=\Phi\left(\frac{\sqrt{n}}{5}\right)>0.997$$

$$\Rightarrow \frac{\sqrt{n}}{5}>u_{0.003}=2.75 \Rightarrow n \geq 190$$

所以，要是检查能通过的概率超过0.997，至少应该检查190只灯泡.

2. 样本来自两个正态总体

定理5.4 若$X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(\mu_2,\sigma_2^2)$，

X与Y相互独立. 样本$(X_1, X_2, \cdots, X_{n_1})$

与$(Y_1, Y_2, \cdots, Y_{n_2})$分别来自总体$X$和$Y$，则

$$(1)\quad \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

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

例3 设$(X_1, X_2, \cdots, X_{10})$和$(Y_1, Y_2, \cdots, Y_{15})$是来自总体$N(20,3)$的两个独立的样本，求

$$P\{|\overline{X}-\overline{Y}|>0.3\}.$$

解 $$\overline{X}=\frac{1}{10}\sum_{i=1}^{10}X_i \sim N(20, \frac{3}{10}),$$

$$\overline{Y}=\frac{1}{15}\sum_{i=1}^{15}Y_i \sim N(20, \frac{3}{15}),$$

$$\therefore \quad \overline{X}-\overline{Y} \sim N(0, \frac{3}{10}+\frac{3}{15}) = N(0, \frac{1}{2}),$$

故 $\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{1}{2}}} \sim N(0,1)$

从而 $P\{|\overline{X}-\overline{Y}|>0.3\}=1-P\{|\overline{X}-\overline{Y}|\le 0.3\}$

$$=1-P\left\{\left|\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{2}}}\right|\le\frac{0.3}{\sqrt{\frac{1}{2}}}\right\}=1-[\Phi(0.3\sqrt{2})-\Phi(-0.3\sqrt{2})]$$

$$\boxed{\because \Phi(-0.3\sqrt{2})=1-\Phi(0.3\sqrt{2})}$$

$$=2[1-\Phi(0.3\sqrt{2})]\approx 2(1-0.6628)=0.6744.$$

(2) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时,

由定理5.4(1)，知 $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\dfrac{\sigma^2}{n_1}+\dfrac{\sigma^2}{n_2})$

$$\therefore\quad U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N(0,1),$$

又由 $\dfrac{(n_1-1)S_1^{*2}}{\sigma^2}\sim\chi^2(n_1-1),\ \dfrac{(n_2-1)S_2^{*2}}{\sigma^2}\sim\chi^2(n_2-1),$

且它们相互独立, 故由 χ^2 分布的可加性知

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

由于 U 与 V 相互独立,按 t 分布的定义

$$T=\frac{U}{\sqrt{V/(n_1+n_2-2)}}$$

$$=\frac{\left[(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right]/\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\sqrt{((n_1-1)S_1^{*2}+(n_2-1)S_2^{*2})/\sigma^2(n_1+n_2-2)}}$$

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

其中 $S_w=\sqrt{\dfrac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}}.$

$\therefore$ 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时，

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2),$$

其中 $S_w=\sqrt{\dfrac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}}.$

$$=\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}}.$$

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

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

由假设 S_1^{*2}, S_2^{*2} 独立, 则由 F 分布的定义知

$$\frac{(n_1-1)S_1^{*2}}{(n_1-1)\sigma_1^2}\Bigg/\frac{(n_2-1)S_2^{*2}}{(n_2-1)\sigma_2^2}\sim F(n_1-1,n_2-1),$$

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

例3 设$X_1,X_2,\cdots,X_n$是来自正态总体$N(\mu,\sigma^2)$的样本

记
$$\boxed{Y_1=\frac{1}{6}\sum_{i=1}^{6}X_i, Y_2=\frac{1}{3}\sum_{i=7}^{9}X_i}$$

$$S^2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2$$

$$Z=\frac{\sqrt{2}(Y_1-Y_2)}{S}$$

试证明：$Z\sim t(2)$.

解
$$\because Y_1=\overline{X}_A\sim N\left(\mu,\frac{\sigma^2}{6}\right), Y_2=\overline{X}_B\sim N\left(\mu,\frac{\sigma^2}{3}\right)$$

且Y_1,Y_2相互独立

所以 $Y_1 - Y_2 \sim N(0, \dfrac{\sigma^2}{2})$

$$Z = \frac{\sqrt{2}(Y_1 - Y_2)}{S}$$

从而有 $\dfrac{\sqrt{2}(Y_1 - Y_2)}{\sigma} \sim N(0,1)$

$$S^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - Y_2)^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - \overline{X}_B)^2$$

又因为 $\dfrac{2S^2}{\sigma^2} \sim \chi^2(2)$, 且 $\dfrac{\sqrt{2}(Y_1 - Y_2)}{\sigma}$ 与 $\dfrac{2S^2}{\sigma^2}$ 独立

所以 $$Z = \frac{\sqrt{2}(Y_1 - Y_2)/\sigma}{\sqrt{2S^2/2\cdot\sigma^2}} = \frac{\sqrt{2}(Y_1 - Y_2)}{S} \sim t(2).$$

例4 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ $\bar{X}$ 和 S_n^2 分别为样本均值与方差,又设 $X_{n+1} \sim N(\mu, \sigma^2)$ 且与 $X_1, X_2, \cdots, X_n$ 相互独立，试求常数C 使得 $F = C(X_{n+1} - \bar{X})^2 / S_n^2$ 服从$F(1,n-1)$.

解 因为 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), X_{n+1} \sim N(\mu, \sigma^2)$

所以，由正态分布的线性性得

$$(X_{n+1} - \bar{X}) \sim N(0, \frac{n+1}{n}\sigma^2)$$

因此 $$\frac{(X_{n+1} - \bar{X})}{\sigma} \cdot \sqrt{\frac{n}{n+1}} \sim N(0,1)$$

从而有 $$U=\left[\frac{(X_{n+1}-\bar{X})}{\sigma}\cdot\sqrt{\frac{n}{n+1}}\right]^2\sim\chi^2(1)$$

另一方面，有样本方差的性质知

$$V=\frac{nS_n^2}{\sigma^2}\sim\chi^2(n-1)$$ 且 $\bar{X}$ 与 S_n^2 相互独立

$\therefore U$ 与 V 相互独立

所以 C=(*n*-1)/(*n*+1).

1)

内容小结

抽样分布定理

1、单正态总体的抽样分布 $X \sim N(\mu, \sigma^2)$,

$$(1)\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2 / n)$$

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

$$(3)\ \frac{nS_n^2}{\sigma^2} = \frac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$$

$$(4)\ \frac{\bar{X} - \mu}{S_n^* / \sqrt{n}} = \frac{\bar{X} - \mu}{S_n / \sqrt{n-1}} \sim t(n-1).$$

2 两正态总体的抽样分布

若$X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$,

$$(1)\bar{X} \sim N(\mu_1,\frac{\sigma_1^2}{n_1}) \qquad \bar{Y} \sim N(\mu_2,\frac{\sigma_2^2}{n_2})$$

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

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

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

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

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

其中 $S_w=\sqrt{\frac{(n_1-1)S_1^{*2}+(n_2-1)S_2^{*2}}{n_1+n_2-2}}.$

$$(3)\ \frac{(n_1-1)S_1^{*2}}{\sigma_1^2} \sim \chi^2(n_1-1), \quad \frac{(n_2-1)S_2^{*2}}{\sigma_2^2} \sim \chi^2(n_2-1),$$

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

西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

5-3 抽样分布

Thank You !