# 第 6 章 抽样分布

## 6.1 内容提要

### 6.1.1 总体与样本

**1. 总体**

研究对象的某项数量指标的全体称为总体. 组成总体的每个元素称为个体. 总体是一个随机变量 $X$ , 而所取的每个值就是一个个体.

**2. 样本**

从总体 $X$ 中随机抽取的 $n$ 个个体 $X_1, X_2, \cdots, X_n$ 称为容量为 $n$ 的样本, 是 $n$ 维随机变量. 通常指的是简单随机样本, 即 $X_1, X_2, \cdots, X_n$ 相互独立, 且每一个 $X_i$ 与总体 $X$ 有相同的分布.

### 6.1.2 统计量

**1. 统计量**

是样本的函数 $T = T(X_1, X_2, \cdots, X_n)$ 且不含任何未知参数.

**2. 常用的统计量**

样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ .

样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

样本标准差 $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ .

样本的 $k$ 阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ .

样本的 $k$ 阶中心矩 $M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$.

顺序统计量 把样本 $X_1, X_2, \cdots, X_n$ 按从小到大的顺序排列起来得到 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ ,

其中

$$X_{(1)} = \min(X_1, X_2, \cdots, X_n),\ X_{(n)} = \max(X_1, X_2, \cdots, X_n),$$

$\boldsymbol{X_{(k)}}$称为第$k$顺序统计量$(k=1,2,\cdots,n)$.

3. 经验分布函数

设$X_1, X_2, \cdots, X_n$是来自总体$X$的样本, 对于任意的实数$x$, 用$S(x)$表示样本中不大于$x$的随机变量的个数, 则$S(x)$表示事件$\{X \le x\}$出现的频数, 而它出现的频率

$$F_n(x) = \frac{1}{n}S(x) = \begin{cases} 0, & x < X_{(1)}, \\ \dfrac{k}{n}, & X_{(k)} \le x \le X_{(k+1)}, k = 1,2,\ldots,n-1, \\ 1, & x > X_{(n)} \end{cases}$$

称之为经验分布函数.

### 6.1.3 数理统计中几个常用的分布

1. 正态分布

(1) $X \sim N(\mu, \sigma^2)$的密度函数, 密度曲线, 期望, 方差.

(2) $X \sim N(\mu, \sigma^2)$, $a, b \in R, a \ne 0$, 则

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2),$$

特别地, $X^* = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

(3) $X_i \sim N(\mu_i, \sigma_i^2)\ (i=1,2)$, 且$X_1, X_2$相互独立, 则

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)\text{(正态分布可加性)}.$$

(4) $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则

$$EX_i = \mu_i, \quad DX_i = \sigma_i^2, \quad (i=1,2)$$

$$\rho_{X_1X_2} = \rho,$$

并且, $X_1, X_2$相互独立的充要条件是$\rho = 0$.

(5) 分位点 设$X \sim N(0,1)$, 对给定的$\alpha\ (0<\alpha<1)$, 满足

$$P\{X > U_\alpha\} = \alpha$$

的数值$U_\alpha$称为标准正态分布的上$\alpha$分位点. 因为标准正态分布是对称分布, 所以在统计推断中常常要用双侧$\alpha$分位点$U_{\alpha/2}$, 它满足

$$P\{|X| > U_{\alpha/2}\} = \alpha.$$

2. $\chi^2$ 分布

(1) 定义 设 $X_1, X_2, \cdots, X_n$ 相互独立, 且均服从 $N(0,1)$ 分布, 则

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n),$$

(2) 期望与方差 若 $\chi^2 \sim \chi^2(n)$, 则

$$E\chi^2 = n, \quad D\chi^2 = 2n.$$

(3) 可加性 若 $X_i \sim \chi^2(n_i)\ (i=1,2)$, 且 $X_1$ 与 $X_2$ 相互独立, 则

$$X_1 + X_2 \sim \chi^2(n_1 + n_2).$$

(4) 分位点 设 $\chi^2 \sim \chi^2(n)$, 对给定的 $\alpha\ (0<\alpha<1)$, 满足 $P\{\chi^2 > \chi_\alpha^2(n)\} = \alpha$ 的数值 $\chi_\alpha^2(n)$ 称为 $\chi^2(n)$ 分布的上 $\alpha$ 分为点.

3. $t$ 分布

(1) 定义 设 $X \sim N(0,1), Y \sim \chi^2(n)$, 且 $X$ 与 $Y$ 相互独立, 则

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n).$$

(2) 分位点 设 $T \sim t(n)$, 对给定的 $\alpha\ (0<\alpha<1)$, 满足

$$P\{T > t_\alpha(n)\} = \alpha$$

的数值 $t_\alpha(n)$ 称为 $t(n)$ 分布的上 $\alpha$ 分位点. 因为 $t(n)$ 分布也是对称分布, 其双侧 $\alpha$ 分位点为 $t_{\alpha/2}(n)$, 满足

$$P\{|T| > t_{\alpha/2}(n)\} = \alpha.$$

4. $F$ 分布

(1) 定义 设 $X_i \sim \chi^2(n_i), i=1,2$, 且 $X_1$ 与 $X_2$ 相互独立, 则

$$F = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2).$$

(2) 分位点 设 $F \sim F(n_1, n_2)$, 对给定的 $\alpha\ (0<\alpha<1)$, 满足

$$P\{F > F_\alpha(n_1, n_2)\} = \alpha$$

的数值 $F_\alpha(n_1, n_2)$ 称为 $F(n_1, n_2)$ 分布的上 $\alpha$ 分位点. 由于 $\frac{1}{F} \sim F(n_2, n_1)$, 因此有,

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}.$$

### 6.1.4 一个正态总体的抽样分布

设总体 $X \sim N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 是 $X$ 的样本，$\overline{X}$，$S^2$ 分别是样本均值,样本方差,则有,

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

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

3. $\overline{X}$ 与 $S^2$ 相互独立.

4. $\frac{\overline{X}-\mu}{\sqrt{S^2/n}} \sim t(n-1)$.

### 6.1.5 两个正态总体的抽样分布

设 $X_i \sim N(\mu_i,\sigma_i^2)$ $(i=1,2)$ 是两个相互独立的总体，$X_{i1},X_{i2}$， $\cdots,X_{in_i}$ 是 $X_i$ 的样本,$\overline{X}_i$，$S_i^2$ 是 $X_i$ 的样本均值和样本方差$(i=1,2)$,则有

1. $\overline{X}_1+\overline{X}_2 \sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$,进而

$$U=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

2. 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时,

$$\frac{(n_1+n_2-2)S_W^2}{\sigma^2} \sim \chi^2(n_1+n_2-2),$$

其中 $S_W^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ 称为联合样本方差.

3. 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时,

$$\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_W^2}} \sim t(n_1+n_2-2).$$

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

# 6.2 习题详解

## 6.1 练习题

### 1. 填空题

设总体 $X \sim N(0,1)$，（$X_1, X_2, \cdots, X_n$）是取自总体 $X$ 的容量为 $n$ 的简单随机样本, 则 $\sum_{i=1}^{n} X_i \sim$ ______________.

**解** 由于 $X_1, X_2, \cdots, X_n$ 相互独立, 且 $X_i \sim N(0,1)$, $i = 1,2,\cdots,n$, 则根据第三章正态分布的可加性定理知, $\sum_{i=1}^{n} X_i \sim N(0,n)$.

### 2. 选择题

设总体 $X \sim U(0,1)$，（$X_1, X_2, \cdots, X_n$）是取自总体 $X$ 的容量为 $n$ 的简单随机样本, 则 $\max(X_1, X_2, \cdots, X_n)$ 的概率密度为（　　）.

A. $f(x) = \begin{cases} x^n, & 0 < x < 1 \\ 0, & 其他 \end{cases}$　　B. $f(x) = \begin{cases} nx^{n-1}, & 0 < x < 1 \\ 0, & 其他 \end{cases}$

C. $f(x) = \begin{cases} nx^n, & 0 < x < 1 \\ 0, & 其他 \end{cases}$　　D. $f(x) = \begin{cases} nx^{n+1}, & 0 < x < 1 \\ 0, & 其他 \end{cases}$

**解** 由于 $X_1, X_2, \cdots, X_n$ 独立同分布, 且与总体 $X$ 的分布相同. $X$ 的分布函数为及密度函数分别为

$$G(x) = \begin{cases} 1, & x \ge 1, \\ x, & 0 < x < 1, \\ 0, & x \le 0, \end{cases} \qquad g(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & 其他, \end{cases}$$

即 $F(x) = \max(X_1, X_2, \cdots, X_n)$, 则由第三章最大值和最小值的分布定理, 知故 $F(x) = [G(x)]^n$, 从而其密度函数为

$$f(x) = n[G(x)]^{n-1} g(x) = \begin{cases} nx^{n-1}, & 0 < x < 1, \\ 0, & 其他, \end{cases}$$

故本题应选 B.

## 6.2 练习题

### 1. 填空题

(1) 设随机变量 $X \sim t(n)$, 则 $P\{|X| < t_\alpha(n)\} =$ ____________________.

**解** $P\{|X|<t_{\alpha}(n)\}=1-P\{|X|\geq t_{\alpha}(n)\}=1-2P\{X\geq t_{\alpha}(n)\}=1-2\alpha$.

(2) 设随机变量 $X$ 与 $Y$ 相互独立, 且 $X\sim N(0,4)$, $Y\sim N(1,9)$, 当 $C=$________时, $\dfrac{CX^2}{(Y-1)^2}$ 服从 $F$ 分布, 参数为__________________.

**解** 由于 $X$ 与 $Y$ 的独立性及其分布知, $\left(\dfrac{X-0}{2}\right)^2$, $\left(\dfrac{X-1}{3}\right)^2$ 相互独立, 且均服从 $\chi^2(1)$ 分布, 由 $F$ 分布的定义知 $\dfrac{9X^2}{4(Y-1)^2}\sim F(1,1)$, 从而 $C=\dfrac{9}{4}$.

2. 选择题

(1) 用 $X_n$ 表示将一枚硬币随意投掷 $n$ 次“正面”出现的次数, 则(　　　).

A. $X+Y$ 服从正态分布　　　　　　B. $X^2+Y^2$ 服从 $\chi^2$ 分布

C. $X^2$ 与 $Y^2$ 均服从 $\chi^2$ 分布　　　　D. $X^2/Y^2$ 服从 $F$ 分布

**解** 因为标准正态分布变量的平方服从自由度为 1 的 $\chi^2$ 分布. 当随机变量 $X$ 和 $Y$ 独立时可以保证选项 A, B, D 成立, 但是题中并未要求随机变量 $X$ 和 $Y$ 独立, 选项 A, B, D 未必成立. 故本题应选 C.

(2) 设随机变量 $X\sim t(n)$, $Y=\dfrac{1}{X^2}$, 则(　　　).

A. $Y\sim\chi^2(n)$　　B. $Y\sim\chi^2(n-1)$　　C. $Y\sim F(n,1)$　　D. $Y\sim F(1,n)$

**解** 设 $X=\dfrac{Y}{\sqrt{Z/n}}\sim t(n)$, 其中 $Y\sim N(0,1)$, $Z\sim\chi^2(n)$, 则 $Y^2\sim\chi^2(1)$, 从而由 $F$ 分布的定义知 $\dfrac{1}{X^2}=\dfrac{Z/n}{Y^2}\sim F(n,1)$, 故本题应选 C.

## 6.3 练习题

1. 填空题

(1) 已知某种能力测试的得分服从正态分布 $N(\mu,\sigma^2)$, 随机取 10 个人参与这一测试, 则他们得分的平均值小于 $\mu$ 的概率为____________.

**解** 设 $X_i$ 表示能力测试中第 $i$ 人的得分, 则 $X_i\sim N(\mu,\sigma^2)$, $i=1,2,\cdots,n$, 他们的平均分为 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i\sim N(\mu,\dfrac{\sigma^2}{n})$, 故所求概率为 $P\{\overline{X}<\mu\}=0.5$.

(2) 设随机变量 $X$ 和 $Y$ 相互独立且都服从正态分布 $N(0,3^2)$, 而 $X_1,X_2,\cdots,X_9$ 和 $Y_1,Y_2,\cdots,Y_9$ 分别是来自总体 $X$ 和 $Y$ 的样本, 则统计量 $U=\dfrac{X_1+\cdots+X_9}{\sqrt{Y_1^2+\cdots+Y_9^2}}$ 服从_______分布,

参数为________.

**解** 由于 $X_1, X_2, \cdots, X_9$ 是来自正态总体的样本, 且都服从 $N(0, 3^2)$, 故样本均值

$$\overline{X} = \frac{1}{9}(X_1 + \cdots + X_9) \sim N(0,1),$$

由于 $Y_1, Y_2, \cdots, Y_9$ 相互独立, 且都服从 $N(0, 3^2)$, 则

$$\frac{Y_i}{3} \sim N(0,1), \qquad (i = 1, \cdots, 9)$$

故

$$\chi^2 = \left(\frac{Y_1}{3}\right)^2 + \cdots + \left(\frac{Y_9}{3}\right)^2 = \frac{1}{9}(Y_1^2 + \cdots + Y_9^2) \sim \chi^2(9).$$

又因为 $\overline{X}$ 与 $\chi^2$ 相互独立, 由 $t$ 分布的定义知,

$$\frac{\overline{X}}{\sqrt{\frac{\chi^2}{9}}} = \frac{\frac{1}{9}(X_1 + \cdots + X_9)}{\sqrt{\frac{1}{9^2}(Y_1^2 + \cdots + Y_9^2)}} = U \sim t(9),$$

即统计量 $U$ 服从自由度为 9 的 $t$ 分布.

2. 选择题

(1) 假设 $(X_1, X_2, \cdots, X_n)$ 是来自正态总体 $X \sim N(0, \sigma^2)$ 的样本, $\bar{X}$ 与 $S^2$ 分别是样本均值和样本方差, 则(　　).

A. $\dfrac{\bar{X}^2}{\sigma^2} \sim \chi^2(1)$　B. $\dfrac{S^2}{\sigma^2} \sim \chi^2(n-1)$　C. $\dfrac{\bar{X}}{S} \sim t(n-1)$　D. $\dfrac{S^2}{n\bar{X}^2} \sim F(n-1,1)$

**解** 由正态总体的抽样分布相关定理及常见分布的定义知, 本题应选 D.

(2) 假设 $(X_1, X_2, \cdots, X_n)$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本, $\bar{X}$ 是样本均值, 记

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2, \qquad S_2^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2,$$

$$S_3^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \mu)^2, \qquad S_4^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2,$$

则服从 $t(n)$ 的随机变量是(　　).

A. $t = \dfrac{\bar{X} - \mu}{S_1/\sqrt{n-1}}$　B. $t = \dfrac{\bar{X} - \mu}{S_2/\sqrt{n-1}}$　C. $t = \dfrac{\bar{X} - \mu}{S_3/\sqrt{n}}$　D. $t = \dfrac{\bar{X} - \mu}{S_4/\sqrt{n}}$

**解** $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\dfrac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$, 再由 $t$ 分布的定义知, 本题应选 D.

**注** 注意正态总体的两个样本的函数

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \quad 与 \quad \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$$

的区别, 它们分别服从自由度为$n$与$n-1$的$\chi^2$分布.

**习题六**

1. 设$X_1,X_2,\cdots,X_9$是来自正态总体$N(\mu,4)$的简单随机样本, $\overline{X}$是样本均值, 已知$P\{|\overline{X}-\mu|<\mu\}=0.95$, 试确定$\mu$的数值.

**解** 因为$X\sim N(\mu,4)$, 故$\overline{X}\sim N(\mu,\frac{4}{9})$, 所以

$$P\{|\overline{X}-\mu|<\mu\}=2\Phi(\frac{\mu}{2/3})-1=0.95,$$

即$\Phi(\frac{3\mu}{2})=0.975$, 查表得, $\frac{3\mu}{2}=1.96$, 从而解得$\mu=1.3067$.

2. 在天平上重复称量一个重为$a$的物品, 假设各次称量结果相互独立而且同服从正态分布$N(a,0.2^2)$, 若以$\overline{X}_n$表示$n$次称量结果的算术平均值, 则为使$P\{|\overline{X}_n-a|<0.1\}\geq 0.95$, $n$至少应等于多少?

**分析** 对该物品进行独立重复称量的所有可能结果, 看成总体$X$, 则$n$次称量结果$X_1,X_2,\cdots,X_n$就是$X$的一容量为$n$的样本, $\overline{X}_n$即样本均值.

**解** 由题意可知$\frac{\overline{X}_n-a}{0.2/\sqrt{n}}\sim N(0,1)$, 又

$$0.95\leq P\{|\overline{X}_n-a|<0.1\}=P\left\{\left|\frac{\overline{X}_n-a}{0.2/\sqrt{n}}\right|<\frac{0.1}{0.2/\sqrt{n}}\right\}=2\Phi\left(\frac{\sqrt{n}}{2}\right)-1,$$

故有$\Phi(\frac{\sqrt{n}}{2})\geq 0.975$, 查标准正态分布表, 得$\frac{\sqrt{n}}{2}\geq 1.96$, 从而$n\geq 15.3664$, 因此$n$至少应等于 16.

3. 设$X_1,X_2,X_3,X_4$是来自正态总体$N\left(0,2^2\right)$的容量为 4 的简单随机样本,

$$X=a\left(X_1-2X_2\right)^2+b\left(3X_3-4X_4\right)^2,$$

则当$a$, $b$各取什么值时, 统计量$X$服从自由度为 2 的$\chi^2$分布.

**分析** 统计量$X$服从$\chi^2$分布, 自由度只能为$2$, 且要$\sqrt{a}(X_1-2X_2)$与$\sqrt{b}(X_3-4X_4)$相互独立, 均服从$N(0,1)$.

**解** 由正态分布的性质及样本的独立性知, $X_1-2X_2$和$3X_3-4X_4$均服从正态分布且相互独立. 由于

$$E(X_1-2X_2)=0,\quad D(X_1-2X_2)=D(X_1)+4D(X_2)=20,$$

以及

$$E(3X_3-4X_4)=0,\quad D(3X_3-4X_4)=9D(X_3)+16D(X_4)=100,$$

故有

$$X_1-2X_2\sim N(0,20),\qquad 3X_3-4X_4\sim N(0,100),$$

从而

$$\frac{X_1-2X_2}{\sqrt{20}}\sim N(0,1),\qquad \frac{3X_3-4X_4}{\sqrt{100}}\sim N(0,1),$$

于是由 $\chi^2$ 分布的定义知, 当 $a=\dfrac{1}{20}, b=\dfrac{1}{100}$ 时, 有

$$X=a(X_1-2X_2)^2+b(3X_3-4X_4)^2=\left(\frac{X_1-2X_2}{\sqrt{20}}\right)^2+\left(\frac{3X_3-4X_4}{10}\right)^2\sim\chi^2(2).$$

4. 设总体 $X\sim B(1,p)$, $X_1,X_2,\cdots,X_n$ 是来自该总体的样本. (1) 求样本均值 $\overline{X}$ 的分布律; (2) 求 $E(\overline{X})$, $D(\overline{X})$, $E(S^2)$.

**解** (1) 由题意知, $\sum\limits_{i=1}^{n}X_i\sim B(n,p)$, 故 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 的可能取值为 0, 1, 2, ⋯, $n$. 其分布律为

$$P\{\overline{X}=\frac{k}{n}\}=P\{\sum_{i=1}^{n}X_i=k\}=C_n^k p^k(1-p)^{n-k},\quad k=0,1,2,\cdots,n.$$

(2) 由数学期望与方差的性质知:

$$E(\overline{X})=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\cdot np=p,$$

$$D(\overline{X})=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\cdot np(1-p)=\frac{p(1-p)}{n},$$

$$E(S^2)=D(X)=p(1-p).$$

5. 设总体 $X\sim N(\mu,\sigma^2)$, 从中抽取一样本 $X_1,X_2,\cdots,X_n,X_{n+1}$, 记 $\overline{X}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$, $S_n^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X}_n)^2$, 试证: $\sqrt{\dfrac{n}{n+1}}\cdot\dfrac{X_{n+1}-\overline{X}_n}{S}\sim t(n-1)$.

**证** 首先对所给统计量作变换, 在统计量的表达式中将分子和分母同除以 $\sigma$, 得

$$\sqrt{\frac{n}{n+1}}\cdot\frac{X_{n+1}-\overline{X}_n}{S_n}=\frac{U}{\sqrt{\chi^2/n-1}},$$

其中 $U=\frac{X_{n+1}-\overline{X}_n}{\sigma}\sqrt{\frac{n}{n+1}}$ ，$\chi^2=\frac{(n-1)S_n^2}{\sigma^2}$ . 由于总体 $X\sim N(\mu,\sigma^2)$ ，可见 $X_{n+1}\sim N(\mu,\sigma^2)$，$\overline{X}_n\sim N\left(\mu,\frac{\sigma^2}{n}\right)$, 从而

$$X_{n+1}-\overline{X}_n\sim N\left(0,(1+\frac{1}{n})\sigma^2\right), U=\frac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\frac{n}{n+1}}\sim N(0,1).$$

对于正态总体, $\overline{X}_n$ 和 $S_n^2$ 独立, 随机变量 $\chi^2=\frac{(n-1)S_n^2}{\sigma^2}$ 服从自由度为 $n-1$ 的 $\chi^2$ 分布.

现在证明, $X_{n+1}$，$\overline{X}_n$ 和 $S_n^2$ 独立. 首先它们显然两两独立, 其次对于任意实数 $u,v,w$，有

$$P\{X_{n+1}\le u,\overline{X}_n\le v,S_n^2\le w\}=P\{X_{n+1}\le u\}P\{\overline{X}_n\le v,S_n^2\le w\}$$
$$=P\{X_{n+1}\le u\}P\{\overline{X}_n\le v\}P\{S_n^2\le w\},$$

其中第一个等式成立, 因为 $X_1,\cdots,X_n$ 和 $X_{n+1}$ 独立；第二个等式成立, 因为正态总体的样本均值和样本方差独立. 从而 $X_{n+1}$，$\overline{X}_n$ 和 $S_n^2$ 独立.

于是, 由服从 t 分布的随机变量的典型模式, 知统计量

$$\sqrt{\frac{n}{n+1}}\cdot\frac{X_{n+1}-\overline{X}_n}{S}\sim t(n-1).$$

6. 设总体 $X$ 服从正态分布 $N(\mu,\sigma^2)$ $(\sigma>0)$，从该总体中抽取简单随机样本 $X_1,X_2,\cdots,X_{2n}$ $(n\ge 2)$, 其样本均值为 $\overline{X}=\frac{1}{2n}\sum_{i=1}^{2n}X_i$，求统计量

$$Y=\sum_{i=1}^{n}(X_i+X_{n+i}-2\overline{X})^2$$

的数学期望 $E(Y)$.

**解法 1** 考虑 $X_1+X_{n+1},X_2+X_{n+2},\cdots,X_n+X_{2n}$，将其视为取自正态总体 $N(2\mu,2\sigma^2)$ 的简单随机样本, 则其样本均值为

$$\frac{1}{n}\sum_{i=1}^{n}(X_i+X_{n+i})=\frac{1}{n}\sum_{i=1}^{2n}X_i=2\overline{X},$$

样本方差为 $\frac{1}{n-1}Y$ . 由于 $E\left(\frac{1}{n-1}Y\right)=2\sigma^2$, 所以

$$E(Y)=(n-1)(2\sigma^2)=2(n-1)\sigma^2.$$

**解法 2** 记 $\overline{X}'=\frac{1}{n}\sum_{i=1}^{n}X_i$，$\overline{X}''=\frac{1}{n}\sum_{i=1}^{n}X_{n+i}$，显然有 $2\overline{X}=\overline{X}'+\overline{X}''$，因此

$$E(Y)=E\left[\sum_{i=1}^{n}(X_i+X_{n+i}-2\overline{X})^2\right]=E\left\{\sum_{i=1}^{n}\left[(X_i-\overline{X}')+(X_{n+i}-\overline{X}'')\right]^2\right\}$$

$$=E\left\{\sum_{i=1}^{n}\left[(X_i-\overline{X}')^2+2(X_i-\overline{X}')(X_{n+i}-\overline{X}'')+(X_{n+i}-\overline{X}'')^2\right]\right\}$$

$$=(n-1)\sigma^2+0+(n-1)\sigma^2=2(n-1)\sigma^2.$$

7. $X_1,X_2,\cdots,X_9$ 是取自正态总体 $X$ 的简单随机样本, 记

$$Y_1=\frac{1}{6}(X_1+\cdots+X_6),\qquad Y_2=\frac{1}{3}(X_7+X_8+X_9),$$

$$S^2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2,\qquad Z=\frac{\sqrt{2}(Y_1-Y_2)}{S},$$

证明:统计量 $Z$ 服从自由度为 2 的 $t$ 分布.

**解** 记 $D(X)=\sigma^2$（未知）, 易见 $E(Y_1)=E(Y_2)$, $D(Y_1)=\dfrac{\sigma^2}{6}$, $D(Y_2)=\dfrac{\sigma^2}{3}$. 由于 $Y_1,Y_2$ 相互独立, 故有

$$E(Y_1-Y_2)=0,\quad D(Y_1-Y_2)=\frac{\sigma^2}{6}+\frac{\sigma^2}{3}=\frac{\sigma^2}{2}.$$

从而 $U=\dfrac{Y_1-Y_2}{\sigma/\sqrt{2}}\sim N(0,1)$. 又 $\chi^2=\dfrac{2S^2}{\sigma^2}\sim\chi^2(2)$, 由于 $Y_1$ 与 $Y_2$ 相互独立, $Y_1$ 与 $S^2$ 独立, $Y_2$ 与 $S^2$ 独立, 所以 $Y_1-Y_2$ 与 $S^2$ 独立, 于是由 $t$ 分布的定义, 知

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

8. 设 $X,Y$ 为两个正态总体, 又 $X\sim N(\mu_1,\sigma_1^2)$, $(X_1,\cdots,X_n)$ 为取自 $X$ 的样本, $\overline{X},S_1^2$ 分别为其样本均值和样本方差, $Y\sim N(\mu_2,\sigma_2^2)$, $(Y_1,\cdots,Y_n)$ 为取自 $Y$ 的样本, $\overline{Y},S_2^2$ 分别为其样本均值和样本方差, 且两样本独立, 求统计量

$$U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{S_1^2+S_2^2-2S_{12}}}\sqrt{n}$$

所服从的分布. 其中 $S_{12}=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})$.

**解** 令 $Z=X-Y$, 则 $Z\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$, 视 $Z$ 为样本, 则 $Z_i=X_i-Y_i$（$i=1,2,\cdots,n$）是取自总体 $Z$ 的样本, 则其样本均值为 $\overline{Z}=\overline{X}-\overline{Y}$, 样本方差为

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(Z_i-\overline{Z})=\frac{1}{n}\sum_{i=1}^{n}(X_i-Y_i-(\overline{X}-\overline{Y}))=S_1^2+S_2^2-2S_{12},$$

故

$$U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{S_1^2+S_2^2-2S_{12}}}\sqrt{n}=\frac{\overline{Z}-(\mu_1-\mu_2)}{S/\sqrt{n}}\sim t(n-1).$$

9. 设 $(X_1,\cdots,X_5)$ 是取自正态总体 $N(0,\sigma^2)$ 的一个样本, 试问当 $k$ 为何值时, $\dfrac{k(X_1+X_2)^2}{X_3^2+X_4^2+X_5^2}$ 服从 $F$ 分布.

**解** 因为 $X_1+X_2\sim N(0,2\sigma^2)$, 所以

$$\frac{X_1+X_2}{\sqrt{2}\sigma}\sim N(0,1), \qquad \left(\frac{X_1+X_2}{\sqrt{2}\sigma}\right)^2\sim \chi^2(1),$$

又由于 $\dfrac{X}{\sigma}\sim N(0,1)$, 故

$$\left(\frac{X_3}{\sigma}\right)^2+\left(\frac{X_4}{\sigma}\right)^2+\left(\frac{X_5}{\sigma}\right)^2\sim \chi^2(3),$$

而 $(X_1+X_2)^2$ 与 $X_3^2+X_4^2+X_5^2$ 独立, 故

$$\frac{\dfrac{\left(\dfrac{X_1+X_2}{\sqrt{2}\sigma}\right)^2}{1}}{\dfrac{\left(\dfrac{X_3}{\sigma}\right)^2+\left(\dfrac{X_4}{\sigma}\right)^2+\left(\dfrac{X_5}{\sigma}\right)^2}{3}}=\frac{3}{2}\cdot\frac{(X_1+X_2)^2}{X_3^2+X_4^2+X_5^2}\sim F(1,3),$$

从而应取 $k=\dfrac{3}{2}$.

10. 从两个正态总体中分别抽取容量为 25 和 20 的两个独立样本, 算得样本方差依次为 $s_1^2=62.7, s_2^2=25.6$, 若两总体方差相等, 求随机抽取的样本的样本方差比 $\dfrac{S_1^2}{S_2^2}$ 大于 $\dfrac{62.7}{25.6}$ 的概率是多少?

**解** 由于正态总体的方差相等, 且 $n_1=25$, $n_2=20$, 故有 $\dfrac{S_1^2}{S_2^2}\sim F(24,19)$, 从而查 $F$ 分布表可得

$$P\left\{\frac{S_1^2}{S_2^2}>\frac{62.7}{25.6}\right\}=P\left\{\frac{S_1^2}{S_2^2}>2.45\right\}=0.025.$$

**注** 用 $F$ 分布表求相应事件的概率, 在查表时应先按自由度找出上分位数 $F_\alpha(n_1,n_2)$, 再

反查概率$\alpha = P\{F > F_{\alpha}(n_1, n_2)\}$. 类似的问题, 用$t$分布表或$\chi^2$分布表的方法亦一样. 有时可能在表中只能找出邻近的两个分位点$F_{\alpha_1}(n_1, n_2)$, $F_{\alpha_2}(n_1, n_2)$, 即

$$F_{\alpha_1}(n_1, n_2) < F_{\alpha}(n_1, n_2) < F_{\alpha_2}(n_1, n_2),$$

则概率$\alpha$应在$\alpha_1$, $\alpha_2$之间用(线性)插值的方法, 求得其近似值.

11. 设总体$X \sim N(12, 2^2)$, 先抽取容量为 5 的样本$X_1, \cdots X_5$, 试求: (1)样本的最小次序统计量小于 10 的概率; (2)最大次序统计量大于 15 的概率.

**解** (1) 所求的概率为

$$\begin{aligned}
P\{X_{(1)} < 10\} &= 1 - P\{X_{(1)} \ge 10\} \\
&= 1 - P\{X_1 \ge 10, X_2 \ge 10, \cdots, X_5 \ge 10\} \\
&= 1 - P\{X_1 \ge 10\} \cdot P\{X_2 \ge 10\} \cdots P\{X_5 \ge 10\} \\
&= 1 - [P\{X \ge 10\}]^5 = 1 - \left[P\left(\frac{X-12}{2} > \frac{10-12}{2}\right)\right]^5 \\
&= 1 - [1 - \Phi(-1)]^5 = 1 - [\Phi(1)]^5 \\
&= 1 - (0.8413)^5 \approx 0.5785.
\end{aligned}$$

(2) 所求的概率为

$$\begin{aligned}
P\{X_{(5)} > 15\} &= 1 - P\{X_{(5)} \le 15\} \\
&= 1 - P\{X_1 \le 15, X_2 \le 15, \cdots, X_5 \le 15\} \\
&= 1 - P\{X_1 \le 15\} \cdot P\{X_2 \le 15\} \cdots P\{X_5 \le 15\} \\
&= 1 - [P\{X \le 15\}]^2 = 1 - \left[P\{\frac{X-12}{2} \le \frac{15-12}{2}\}\right]^5 \\
&= 1 - [\Phi(1.5)]^5 = 1 - (0.9332)^5 \approx 0.2923.
\end{aligned}$$

**注** 关于样本最小顺序统计量$X_{(1)}$与样本最大顺序统计量$X_{(n)}$的分布, 可做如下推导:

设总体$X$的分布函数为$F(x)$, 因为

$$P\{X_{(1)} > a\} = P\{X_1 > a, X_2 > a, \cdots, X_n > a\} = [1 - F(a)]^n,$$

$$P\{X_{(n)} \le a\} = P\{X_1 \le a, X_2 \le a, \cdots, X_n \le a\} = [F(a)]^n,$$

所以, $X_{(1)}$的分布函数为

$$F_1(x) = P\{X_{(1)} \le x\} = 1 - [1 - F(x)]^n.$$

$X_{(n)}$的分布函数为

$$F_2(x) = P\{X_{(2)} \le x\} = [F(x)]^n.$$

12. 设 $X_1, X_2$ 为取自正态总体 $N(\mu,\sigma^2)$ 的样本, (1) 证明 $X_1+X_2$ 与 $X_1-X_2$ 相互独立; (2) 假定 $\mu=0$, 求 $\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2}$ 的分布, 并求 $P\left\{\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2}<4\right\}$.

**分析** 根据正态分布的性质, $X_1+X_2$ 与 $X_1-X_2$ 服从二维正态分布, 所以要证明它们相互独立, 只需证它们不相关即可.

**解** (1) 由于

$$E[(X_1+X_2)(X_1-X_2)]=E(X_1^2)-E(X_2^2)=0,$$

$$E(X_1+X_2)E(X_1-X_2)=0,$$

所以

$$\operatorname{cov}(X_1+X_2, X_1-X_2)=0,$$

即 $X_1+X_2$ 与 $X_1-X_2$ 相互独立.

(2) 由于 $\mu=0$, 所以

$$X_1+X_2 \sim N(0,2\sigma^2) \Rightarrow \frac{X_1+X_2}{\sqrt{2}\sigma} \sim N(0,1),$$

$$X_1-X_2 \sim N(0,2\sigma^2) \Rightarrow \frac{X_1-X_2}{\sqrt{2}\sigma} \sim N(0,1),$$

从而

$$\frac{1}{2}\left(\frac{X_1+X_2}{\sigma}\right)^2 \sim \chi^2(1), \qquad \frac{1}{2}\left(\frac{X_1-X_2}{\sigma}\right)^2 \sim \chi^2(1),$$

由上面证明的独立性, 再由 $F$ 分布的定义知

$$F=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}=\frac{\left(\dfrac{X_1+X_2}{\sigma}\right)^2/2}{\left(\dfrac{X_1-X_2}{\sigma}\right)/2} \sim F(1,1),$$

所以

$$P\left\{\frac{(X_1+X_2)^2}{(X_1-X_2)^2}<4\right\}=P\{F<4\}<P\{F<5.83\}=0.25.$$