

《概率论与数理统计》期末速通

6. 样本与抽样分布

6.1 随机样本

[定义6.1.1] 试验的所有可能的观察值或研究对象的全体称为**总体**, 每个可能的观察值称为**个体**. 总体中包含的个体的个数称为总体的**容量**, 容量有限的总体称为**有限总体**, 否则称为**无限总体**.

[注] 总体中的每个个体对应试验的一个观察值, 进而对应一个随机变量 X 的值, 即每个个体对应一个随机变量, 下面称总体 X .

[定义6.1.2] 从总体中抽取一部分个体, 根据得到的数据推断总体分布, 被抽出的部分个体称为总体的一个**样本**. 严谨地, 在相同条件下, 对总体 X 进行 n 次重复、独立的观察, n 次观察的结果按试验次序依次记作 X_1, \dots, X_n . 因 X_1, \dots, X_n 是对 X 观察的结果, 且各次观察在相同条件下独立进行, 则可认为 X_1, \dots, X_n 相互独立, 且与 X 同分布, 此时称 X_1, \dots, X_n 是来自 X 的一个**简单随机样本**, 简称**样本**, 其中 n 称为样本的**容量**.

[注]

(1) 对有限个体, 用放回抽样得到样本, 但使用不方便. 个体总数 $N \gg$ 样本容量 n 时, 可将不放回抽样近似为放回抽样.

(2) 对无限个体, 抽取一个个体不影响其分布, 故用不放回抽样得到样本.

(3) 有限个体的个体总数较多时, 可近似为无限个体.

[定义6.1.3] 设随机变量 X 的分布函数为 F . 若 X_1, \dots, X_n 是有同一分布函数 F 且相互独立的随机变量, 则称 X_1, \dots, X_n 为取自总体 X 、容量为 n 的**简单样本**, 简称**样本**, 它们的观察值 x_1, \dots, x_n 称为**样本值**, 又称为 X 的 n 个**独立的观察值**. 可将样本视为一个随机向量, 记作 (X_1, \dots, X_n) , 此时样本值记作 (x_1, \dots, x_n) .

[注] 若 X_1, \dots, X_n 是取自 X 的一个样本, 则有: ① X_1, \dots, X_n 相互独立; ② X_1, \dots, X_n 与 X 同分布.

[定理6.1.1] 设 X_1, \dots, X_n 是取自总体 X 且容量为 n 的样本, 则:

(1) 随机变量 (X_1, \dots, X_n) 的联合分布函数 $F^*(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) = \prod_{i=1}^n F(x_i)$.

(2) 离散型随机变量的样本 X_1, \dots, X_n 的联合分布律

$$P\{X_1 = x_1, \dots, X_n = x_n\} = P\{X_1 = x_1\} \cdots P\{X_n = x_n\} = \prod_{i=1}^n P\{X = x_i\}.$$

(3) 连续型随机变量的样本 X_1, \dots, X_n 的联合概率密度

$$f^*(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \prod_{i=1}^n f(x_i).$$

[例6.1.1] 设总体 $X \sim \text{Exp}(\theta)$, X_1, \dots, X_n 是取自 X 的样本.

(1) 求随机变量 (X_1, \dots, X_n) 的概率密度.

(2) 求 (X_1, \dots, X_n) 的分布函数.

[解] X 的概率密度 $f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$, 分布函数 $F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$.

$$(1) f(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i) = \begin{cases} \frac{1}{\theta^n} e^{-\frac{x_1 + \dots + x_n}{\theta}}, & x_1, \dots, x_n > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$(2) F(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i) = \begin{cases} \prod_{i=1}^n (1 - e^{-\frac{x_i}{\theta}}), & x_1, \dots, x_n \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

[例6.1.2] 设总体 $X \sim b(n, p)$, X_1, \dots, X_n 是取自 X 的样本. 求随机变量 (X_1, \dots, X_n) 的分布律.

[解] X 的分布律 $P\{X = k\} = C_n^k \cdot p^k \cdot (1-p)^{n-k}$ ($k = 0, 1, \dots, n$).

$$P\{X_1 = x_1, \dots, X_n = x_n\} = \prod_{i=1}^n P\{X_i = x_i\} = \prod_{i=1}^n C_n^{x_i} \cdot p^{x_i} \cdot (1-p)^{n-x_i} \quad (x_i = 0, 1, \dots, n).$$

6.2 统计量与经验分布函数

6.2.1 统计量

[定义6.2.1] 设 X_1, \dots, X_n 是取自总体 X 的一个样本, $g(X_1, \dots, X_n)$ 是 X_1, \dots, X_n 的一个函数. 若 g 中不含未知参数, 则称 $g(X_1, \dots, X_n)$ 是一个**统计量**. 若 x_1, \dots, x_n 是样本 X_1, \dots, X_n 的样本值, 则称 $g(x_1, \dots, x_n)$ 是统计量 $g(X_1, \dots, X_n)$ 的**观察值**.

[例6.2.1] 设总体 $X \sim N(\mu, \sigma^2)$, 其中 μ 已知, σ^2 未知. X_1, \dots, X_5 是取自 X 的样本.

(1) $\frac{X_1 + \dots + X_5}{5}$ 是统计量.

(2) $\mu \cdot X_1 + 2X_2^2$ 是统计量.

(3) $\frac{X_1 + X_2 + X_5}{\sigma}$ 不是统计量, 因包含未知参数 σ .

[定义6.2.2] 设 X_1, \dots, X_n 的取自总体 X 的一个样本, x_1, \dots, x_n 是样本的观察值.

定义	统计量	观察值
样本均值	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
样本方差	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ $= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ $= \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$
样本标准差	$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$	$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
样本 k 阶(原点)矩	$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (k = 1, 2, \dots)$	$a_k = \frac{1}{n} \sum_{i=1}^n x_i^k \quad (k = 1, 2, \dots)$
样本 k 阶中心矩	$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad (k = 2, 3, \dots)$	$b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad (k = 2, 3, \dots)$

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

[注1] 注意方差和标准差中的系数为 $\frac{1}{n-1}$ 而非 $\frac{1}{n}$.

[注2] 样本的均值隐含了总体的均值的信息, 样本的方差隐含了总体的方差的信息.

[定理6.2.1] 设 X_1, \dots, X_n 的取自总体 X 的一个样本. 若 X 的 k ($k = 1, 2, \dots$) 阶矩 $E(X^k) = \mu_k$ 存在, 则 $n \rightarrow +\infty$ 时, 样本的 k 阶矩 $A_k \xrightarrow{P} \mu_k$.

[证] 因 X_1, \dots, X_n 相互独立且与 X 同分布, 则 X_1^k, \dots, X_n^k 相互独立且与 X^k 同分布,

进而 $E(X_i^k) = E(X^k) = \mu_k \quad (i = 1, \dots, n)$.

由**弱大数定理**: $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mu_k$.

[注] 由依概率收敛的序列的性质: 对连续函数 g , 有 $g(A_1, \dots, A_k) \xrightarrow{P} g(\mu_1, \dots, \mu_k)$, 这是矩估计法的理论依据.

6.2.2 经验分布函数

[定义6.2.3] 设 X_1, \dots, X_n 是总体 F 的一个样本. 设 $S(x)$ 为 X_1, \dots, X_n 中 $\leq x$ 的随机变量的个数, 定义**经验分布函数** $F_n(x) = \frac{1}{n}S(x)$. 对一个样本值, 易得 $F_n(x)$ 的观察值, 仍用 $F_n(x)$ 表示.

[例6.2.2]

(1) 设总体 F 有一个样本值 1, 2, 3, 求其经验分布函数.

(2) 设总体 F 有一个样本值 1, 1, 2, 求其经验分布函数.

[解]

$$(1) \text{ 经验分布函数 } F_3(x) \text{ 的观察值 } F_3(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{3}, & 1 \leq x < 2 \\ \frac{2}{3}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}.$$

$$(2) \text{ 经验分布函数 } F_3(x) \text{ 的观察值 } F_3(x) = \begin{cases} 0, & x < 1 \\ \frac{2}{3}, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}.$$

[注1] 可将随机变量的取值的频率近似为其概率, 用求离散型随机变量的分布函数的方式求经验分布函数.

[注2] 设 x_1, \dots, x_n 是总体 F 的一个容量为 n 的样本值. 求经验分布函数的方法:

① 将 x_1, \dots, x_n 非降序排列, 重新编号为 $x_{(1)} \leq \dots \leq x_{(n)}$.

$$\text{② 经验分布函数 } F_n(x) \text{ 的观察值 } F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{n}, & x_{(k)} \leq x < x_{(k+1)}, k = 1, \dots, n-1 \\ 1, & x \geq x_{(n)} \end{cases}.$$

[定理6.2.2] [Glivenko定理] 设 X_1, \dots, X_n 是总体 X 的一个样本, 其经验分布函数为 $F_n(x)$. 设 X 的分布函数为 $F(x)$, 则 $n \rightarrow +\infty$ 时, $F_n(x)$ 以概率 1 一致收敛于 $F(x)$, 即 $P \left\{ \lim_{n \rightarrow +\infty} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)| = 0 \right\} = 1$.

[注1] 本定理表明: n 充分大时, 经验分布的观察值 $F_n(x)$ 可近似为总体的分布函数 $F(x)$.

[注2] 对固定的 $x \in \mathbb{R}$, 有 $S(x) \sim b(n, F(x))$, 则 $E[F_n(x)] = \frac{1}{n}E[S(x)] = \frac{1}{n}E[n \cdot F(x)] = F(x)$.

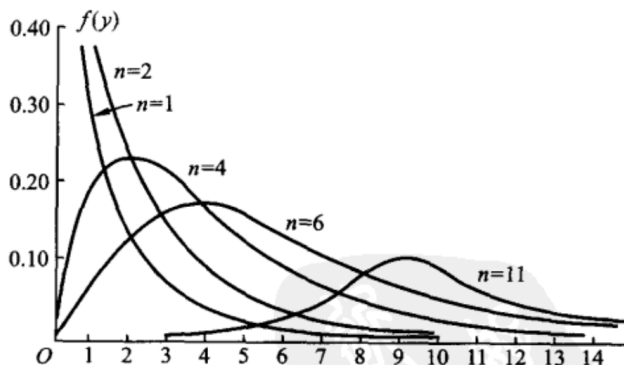
6.3 抽样分布

[定义6.3.1] 统计量的分布称为**抽样分布**. 总体的分布函数已知时, 抽样分布确定.

6.3.1 χ^2 分布

[定义6.3.2] X_1, \dots, X_n 是取自正态总体 $N(0, 1)$ 的样本, 称统计量 $\chi^2 = X_1^2 + \dots + X_n^2$ 服从自由度为 n 的 χ^2 分布或卡方分布, 记作 $\chi^2 \sim \chi^2(n)$, 其中自由度指上式右端包含的独立的随机变量的个数.

[注1] $\chi^2(n)$ 分布的概率密度 $f(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot y^{\frac{n}{2}-1} \cdot e^{-\frac{y}{2}}, y > 0 \\ 0, otherwise \end{cases}$, 其图象如下图所示:



[注2] 若随机变量 $X \sim N(0, 1)$, 则 $X \sim \chi^2(1)$.

[例6.3.1] 设 X_1, \dots, X_4 是取自总体 $X \sim N(0, 1)$ 的样本, 则:

(1) $X_1^2 + X_2^2 + X_3^2 \sim \chi^2(3)$.

(2) $X_1^2 + \frac{1}{2}(X_2 + X_3)^2 \sim \chi^2(2)$, 而非 $\chi^2(3)$,

因为 $X_2, X_3 \sim N(0, 1)$, 则 $X_2 + X_3 \sim N(0, 2)$, 进而 $\frac{(X_2 + X_3) - 0}{\sqrt{2}} \sim N(0, 1)$.

故 $X_1^2 + \left(\frac{X_2 + X_3}{\sqrt{2}}\right)^2 \sim \chi^2(2)$.

[定理6.3.1] [χ^2 分布的性质]

(1) [χ^2 分布的可加性] 若随机变量 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$ 且相互独立, 则 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.

(2) 若随机变量 $\chi^2 \sim \chi^2(n)$, 则期望 $E(\chi^2) = n$, 方差 $D(\chi^2) = 2n$.

[证]

(2) $E(\chi^2) = \sum_{i=1}^n E(X_i^2) = n \cdot E(X^2) = n \cdot \{D(X) + [E(X)]^2\} = n$, 其中随机变量 $X \sim N(0, 1)$.

$$\begin{aligned} D(\chi^2) &= \sum_{i=1}^n D(X_i^2) = n \cdot D(X^2) = n \cdot \{E(X^4) - [E(X^2)]^2\} \\ &= n \cdot \left(\int_{-\infty}^{+\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - 1 \right) = n \cdot 2 = 2n. \end{aligned}$$

[例6.3.1] 设 X_1, \dots, X_4 是总体 $X \sim N(0, 1)$ 的样本.

(1) 随机变量 $X^2 \sim \chi^2(1)$, 则期望 $E(X^2) = 1$, 方差 $D(X^2) = 2$.

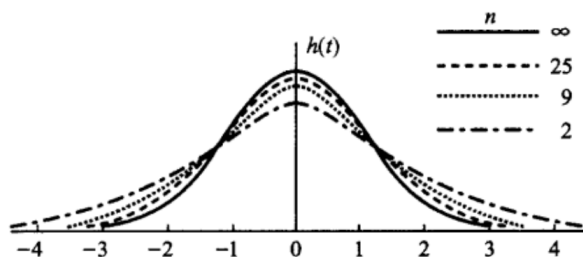
(2) 随机变量 $X_1^2 + \dots + X_4^2 \sim \chi^2(4)$, 则期望 $E(X_1^2 + \dots + X_4^2) = 4$, 方差 $D(X_1^2 + \dots + X_4^2) = 8$.

6.3.2 t 分布

[定义6.3.3] 设随机变量 $X \sim N(0, 1)$, $Y \sim \chi^2(n)$ 且相互独立, 则称统计量 $t = \frac{X}{\sqrt{\frac{Y}{n}}}$ 服从自由度为 n 的 t 分布

或Student分布, 记作 $t \sim t(n)$.

[注1] $t(n)$ 分布的概率密度 $h(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$, 其图象如下图所示:



$h(t)$ 的图象关于 $t = 0$ 对称.

[注2] 可以证明 $\lim_{n \rightarrow +\infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, 则 n 充分大时, t 分布近似于标准正态分布.

[例6.3.2] 设 X_1, \dots, X_{10} 是总体 $X \sim N(0, 1)$ 的样本,

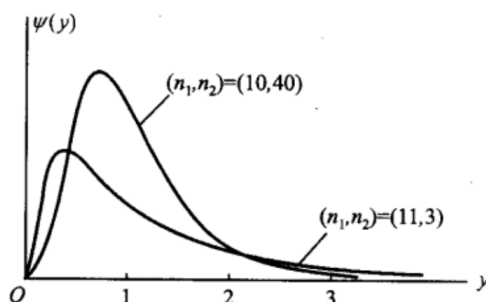
则 $\frac{X_1}{\sqrt{\frac{X_2^2 + \dots + X_{10}^2}{9}}} \sim t(9)$, 因为 $X_2^2 + \dots + X_{10}^2 \sim \chi^2(9)$.

6.3.3 F 分布

[定义6.3.4] 设随机变量 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$ 且相互独立, 则称随机变量 $F = \frac{\frac{U}{n_1}}{\frac{V}{n_2}}$ 服从自由度为 (n_1, n_2)

的 F 分布, 记作 $F \sim F(n_1, n_2)$.

[注] $F(n_1, n_2)$ 分布的概率密度 $\psi(y) = \begin{cases} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right) \cdot \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \cdot y^{\frac{n_1}{2}-1}}{\Gamma\left(\frac{n_1}{2}\right) \cdot \Gamma\left(\frac{n_2}{2}\right) \cdot \left(1 + \frac{n_1}{n_2}y\right)^{\frac{n_1+n_2}{2}}}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$, 其图象如下图所示:



[定理6.3.2] 若随机变量 $F \sim F(n_1, n_2)$, 则随机变量 $\frac{1}{F} \sim F(n_2, n_1)$.

[例6.3.3] 设 X_1, \dots, X_{32} 是总体 $N(0, 1)$ 的样本, 则:

(1) 统计量 $Y = \sum_{k=1}^{10} X_k^2 \sim \chi^2(10)$.

(2) 因随机变量 X_{16} 与 Y 相互独立, 则统计量 $Z = \frac{X_{16}}{\sqrt{\frac{Y}{10}}} \sim t(10)$.

(3) 因 Y 与随机变量 $\sum_{k=11}^{30} X_k^2$ 相互独立, 而 $\sum_{k=11}^{30} X_k^2 \sim \chi^2(20)$,

则统计量 $W = \frac{2Y}{\sum_{k=11}^{30} X_k^2} = \frac{\frac{Y}{10}}{\frac{\sum_{k=11}^{30} X_k^2}{20}} \sim F(10, 20)$.

[例6.3.4] 设 X_1, \dots, X_{20} 是总体 $N(0, 4)$ 的样本, 则随机变量 $\frac{X_k - 0}{2} = \frac{X_k}{2} \sim N(0, 1) \ (k = 1, \dots, 20)$.

(1) 设统计量 $Y = \sum_{k=1}^9 X_k^2$, 则 $\frac{Y}{4} = \sum_{k=1}^9 \left(\frac{X_k}{2}\right)^2 \sim \chi^2(9)$.

(2) 因随机变量 X_{20} 与 Y 相互独立, 则统计量 $Z = \frac{3X_{20}}{\sqrt{Y}} = \frac{\frac{X_{20}}{2}}{\sqrt{\frac{Y}{4}}} \sim t(9)$.

(3) 因 Y 与随机变量 $\sum_{k=11}^{18} X_k^2$ 相互独立, 则统计量 $W = \frac{8}{9} \cdot \frac{Y}{\sum_{k=11}^{18} X_k^2} = \frac{\frac{Y}{9}}{\frac{\sum_{k=11}^{18} X_k^2}{8}} \sim F(9, 8)$.

6.4 分位点

[定义6.4.1] 设随机变量 X 的概率密度为 $f(x)$. 若对固定的 $\alpha \in (0, 1)$, $\exists z_\alpha \in \mathbb{R}$ s. t.

$$P\{X > z_\alpha\} = \int_{z_\alpha}^{+\infty} f(x)dx = \alpha, \text{ 则称点 } z_\alpha \text{ 为 } X \text{ 的上 } \alpha \text{ 分位点.}$$

[定义6.4.2] 设随机变量 $X \sim N(0, 1)$. 对固定的 $\alpha \in (0, 1)$, 称 s. t. $P\{X > z_\alpha\} = \int_{z_\alpha}^{+\infty} \varphi(x)dx = \alpha$ 的点 z_α 为 $N(0, 1)$ 的上 α 分位点, 其中 $\varphi(x)$ 是 $N(0, 1)$ 的概率密度.

[定理6.4.1] 设随机变量 $X \sim N(0, 1)$. 对固定的 $\alpha \in (0, 1)$, 设 z_α 是 X 的上 α 分位点, 则 $z_{1-\alpha} = -z_\alpha$.

[证] 由正态分布的概率密度的图象关于 y 轴对称, 结合几何意义即证.

[定义6.4.3] 设随机变量 $\chi^2 \sim \chi^2(n)$. 对固定的 $\alpha \in (0, 1)$, 称 $s. t. P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{+\infty} f(y)dy = \alpha$ 的点 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 的上 α 分位点, 其中 $f(y)$ 是 $\chi^2(n)$ 的概率密度.

[定义6.4.4] 设随机变量 $t \sim t(n)$. 对固定的 $\alpha \in (0, 1)$, 称 $s. t. P\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^{+\infty} h(t)dt = \alpha$ 的点 $t_\alpha(n)$ 为 $t(n)$ 的上 α 分位点, 其中 $h(t)$ 是 $t(n)$ 的概率密度.

[定理6.4.2] 设随机变量 $X \sim t(n)$. 对固定的 $\alpha \in (0, 1)$, 设 z_α 是 X 的上 α 分位点, 则 $t_{1-\alpha} = -t_\alpha$.

[证] 由 t 分布的概率密度的图象关于 y 轴对称, 结合几何意义即证.

[定义6.4.5] 设随机变量 $F \sim F(n)$. 对固定的 $\alpha \in (0, 1)$, 称 $s. t. P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{+\infty} \psi(y)dy = \alpha$ 的点 $F_\alpha(n_1, n_2)$ 为 $F(n_1, n_2)$ 的上 α 分位点, 其中 $\psi(x)$ 是 $F(n)$ 的概率密度.

[定理6.4.3] 设随机变量 $X \sim F(n_1, n_2)$. 对固定的 $\alpha \in (0, 1)$, 设 F_α 是 X 的上 α 分位点, 则 $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$.

[证] 因 $F \sim F(n_1, n_2)$,

$$\begin{aligned} \text{则 } 1 - \alpha &= P\{F > F_{1-\alpha}(n_1, n_2)\} = P\left\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} \\ &= 1 - P\left\{\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} = 1 - P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\}, \text{ 进而} \\ P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\} &= \alpha. \end{aligned}$$

$$\text{因 } \frac{1}{F} \sim F(n_2, n_1), \text{ 则 } P\left\{\frac{1}{F} > F_\alpha(n_2, n_1)\right\} = \alpha = P\left\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\right\},$$

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

6.5 正态总体的样本均值与样本方差的分布

[定理6.5.1] 设 X_1, \dots, X_n 是取自总体 X 的样本, 期望 $E(X) = \mu$ 和方差 $D(X) = \sigma^2$ 都存在, 则:

(1) 样本均值的期望 $E(\bar{X}) = \mu$.

(2) 样本均值的方差 $D(\bar{X}) = \frac{\sigma^2}{n}$.

(3) 样本方差的期望 $E(S^2) = \sigma^2$.

[证]

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

(2) $D(\bar{X}) = \frac{1}{n^2} D\left(\sum_{i=1}^n X_i\right) \stackrel{\text{独立性}}{=} \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{1}{n^2} \cdot (n\sigma^2) = \frac{\sigma^2}{n}$.

(3) 因 $E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2$,

$$E(\bar{X}^2) = D(\bar{X}) + [E(\bar{X})]^2 = \frac{\sigma^2}{n} + \mu^2,$$

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

[注] 本定理与 X 服从何分布无关, 只需保证期望和方差存在.

[定理6.5.2] 设 X_1, \dots, X_n 是取自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 样本均值为 \bar{X} , 样本方差为 S^2 , 则:

$$(1) \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

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

(3) \bar{X} 与 S^2 相互独立.

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

$$(5) \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1).$$

[证]

(1) 因 $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ 且相互独立, 而 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$,

则 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, 其中 $E(\bar{X}) = E(X) = \mu, D(\bar{X}) = \frac{D(X)}{n} = \frac{\sigma^2}{n}$.

(4) 由(1): $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, 则标准化变量 $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

由(2): $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. 由(3): \bar{X} 与 S^2 相互独立.

$$\text{故 } \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \cdot \frac{1}{n-1}}} \sim t(n-1).$$

[注] (2)中 $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$, 注意非 $\chi^2(n)$,

因为随机变量 $(X_i - \bar{X})$ ($1 \leq i \leq n$) 非独立,

因为 $\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - n \cdot \bar{X} = 0$, 则它们存在约束关系.

[例6.5.1] 设 X_1, \dots, X_5 是取自总体 $X \sim N(2, 7)$ 的样本, 由定理6.5.2:

$$(1) \text{ 样本均值 } \bar{X} \sim N\left(2, \frac{7}{5}\right).$$

$$(2) \text{ 随机变量 } \frac{4}{7}S^2 = \frac{5-1}{7}S^2 \sim \chi^2(4).$$

$$(3) \text{ 随机变量 } \frac{\bar{X} - 2}{\frac{S}{\sqrt{5}}} \sim t(4).$$

[定理6.5.3] 设 X_1, \dots, X_{n_1} 是取自总体 $N(\mu_1, \sigma_1^2)$ 的样本, Y_1, \dots, Y_{n_2} 是取自总体 $N(\mu_2, \sigma_2^2)$ 的样本, 且这两个样本相互独立. 设两样本的样本方差分别为 $S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$, $S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$, 则:

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

$$(2) \text{ 两总体的方差相同, 即 } \sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ 时, 有 } \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \text{ 其中}$$

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

[证]

$$(1) \text{ 由定理6.5.2: } \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1), \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1).$$

$$\text{因 } S_1^2 \text{ 与 } S_2^2 \text{ 相互独立, 则 } \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{\frac{(n_1 - 1)S_1^2}{\sigma_1^2}}{\frac{(n_2 - 1)S_2^2}{\sigma_2^2}} \sim F(n_1 - 1, n_2 - 1).$$