样品及抽样分布

February 21, 2017

1 随机样本

总体：

试验的全部可能观察值；

个体：

每一个可能的观察值；

容量：

总体中所包含的个体的数量；

总体中的每一个个体是随机试验的一个观察值，它是某一随机变量 X 的值；一个总体对应于一个随机变量；

通过从总体中抽取一部分个体，根据获得的数据对总体分布得出推断；被抽出的部分个体，叫总体的一个样本；

从总体中抽取一个个体，即对总体 X 进行一次观察并记录结果；在相同条件下，对总体 X 进行 n 次重复的、独立的观察；将 n 次观察结果按试验的次序记为 $X_1, X_2, \cdots, X_n$。由于 $X_1, X_2, \cdots, X_n$ 是对随机变量 X 观察的结果，且各次观察是在相同条件下独

立进行的，则 $X_1, X_2, \cdots, X_n$ 是相互独立的，且都是与 X 具有相同分布的随机变量 $\Longrightarrow X_1, X_2, \cdots, X_n$ 称为来自总体 X 的一个简单随机样本，n：样本的容量；n 次观察得到一组实数 $x_1, x_2, \cdots, x_n$，依次是随机变量 $X_1, X_2, \cdots, X_n$ 的观察值，称为样本值；

简单随机样本：

设 X 是具有分布函数 F 的随机变量，若 $X_1, X_2, \cdots, X_n$ 是具有同一分布函数 F 的、相互独立的随机变量，称 $X_1, X_2, \cdots, X_n$ 为从分布函数 F(或总体 F、总体 X) 得到的容量为 n 的简单随机样本，简称样本；

样本值：

观察值 $x_1, x_2, \cdots, x_n$；X 的 n 个独立的观察值；

若 $X_1, X_2, \cdots, X_n$ 为 F 的一个样本，则 $X_1, X_2, \cdots, X_n$ 相互独立，且它们的分布函数都是 F，则 $(X_1, X_2, \cdots, X_n)$ 的分布函数：

$$F^*(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i) \tag{1}$$

若 X 具有概率密度 f，则 $(X_1, X_2, \cdots, X_n)$ 的概率密度：

$$f^*(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i) \tag{2}$$

2 抽样分布

统计量：

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本，$g(X_1, X_2, \cdots, X_n)$是 $X_1, X_2, \cdots, X_n$ 的函数，若 g 中不含未知参数，则称 $g(X_1, X_2, \cdots, X_n)$ 是一统计量；

因为 $X_1, X_2, \cdots, X_n$ 是随机变量，而统计量 $g(X_1, X_2, \cdots, X_n)$ 是随机变量的函数，因此也是一随机变量。设 $x_1, x_2, \cdots, x_n$ 是相应于样本 $X_1, X_2, \cdots, X_n$ 的样本值，

称$g(x_1,x_2,\cdots,x_n)$是 $g(X_1,X_2,\cdots,X_n)$ 的观察值。

样本平均值：

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{3}$$

样本方差：

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right) \tag{4}$$

样本标准差：

$$S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2} \tag{5}$$

样本 k 阶 (原点) 矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\quad k=1,2,\cdots \tag{6}$$

样本 k 阶中心矩

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,\quad k=1,2,\cdots \tag{7}$$

其观察值：

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{8}$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right) \tag{9}$$

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2} \tag{10}$$

$$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\quad k=1,2,\cdots \tag{11}$$

$$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\quad k=1,2,\cdots \tag{12}$$

若总体 X 的 k 阶矩 $E(X^k) \xrightarrow{\text{记成}} \mu_k$ 存在，则当 $n \to \infty$ 时，$A_k \xrightarrow{P} \mu_k, k=1,2,\cdots$。由于 $X_1, X_2, \cdots, X_n$ 独立且与 X 同分布，所以 $X_1, X_2, \cdots, X_n$ 独立且与 X^k 同分布。故有

$$E(X_1^k) = E(X_2^k) = \cdots = E(X_n^k) = \mu_k$$

由辛钦定理可知

$$A_k = \frac{1}{n}\sum_{i=1}^{k} X_i^k \xrightarrow{P} \mu_k, k=1,2,\cdots .$$

由依概率收敛的序列的性质，

$$g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \cdots, \mu_k) . \tag{13}$$

其中 g 为连续函数。这是矩估计法的理论依据。

经验分布函数

与总体分布函数 $F(x)$ 相应的统计量；

设 $X_1, X_2, \cdots, X_n$ 是总体 F 的一个样本，用 $S(x), -\infty < x < \infty$ 表示 $X_1, X_2, \cdots, X_n$ 中不大于 x 的随机变量的个数。定义经验分布函数 $F_n(x)$ 为

$$F_n(x) = \frac{1}{n}S(x), \quad -\infty < x < \infty . \tag{14}$$

经验分布函数 $F_n(x)$ 的观察值

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} , \\ \frac{k}{n}, & x_{(k)} \leqslant x < x_{(k+1)} , \\ 1, & x \geqslant x_{(n)} . \end{cases}$$

对于任一实数 x，当 $x \to \infty$ 时，$F_n(x)$ 以概率 1 一致收敛于分布函数 $F(x)$，即

$$P\{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0\} = 1 .$$

因此对于任一实数 x，当 n 充分大时，经验分布函数的任一观察值 $F_n(x)$ 与总体分布函数 $F(x)$ 只有微小的差别，可当作 $F(x)$ 来使用。

抽样分布：

统计量的分布；

使用统计量进行统计推断时需要知道它的分布；

当总体的分布函数已知，抽样分布是确定的；

以下是来自正态总体的统计量分布

2.1 χ^2 分布

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(0,1)$ 的样本，称统计量

$$\chi^2 = X_1^2 + X_1^2 + \cdots + X_n^2 \tag{15}$$

服从自由度为 n 的 χ^2 分布，记为$\chi^2 \sim \chi^2(n)$。自由度：独立变量的个数；

$\chi^2(n)$ 分布的概率密度函数：

$$f(y) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, & y > 0 \\ 0, & \end{cases} \tag{16}$$

2.2 t 分布

也称学生氏分布

设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X，Y 独立，称随机变量

$$t = \frac{X}{\sqrt{Y/n}} \tag{17}$$

服从自由度为 n 的 t 分布，记为$t \sim t(n)$。

$t(n)$ 分布的概率密度函数：

$$h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < \infty \tag{18}$$

2.3 F 分布

设 $U \sim \chi^2(n_1)$，$V \sim \chi^2(n_2)$，且 U、V 独立，称随机变量

$$F = \frac{U/n_1}{V/n_2} \tag{19}$$

服从自由度为 (n_1, n_2) 的 F 分布，记为$F \sim F(n_1, n_2)$。

F 分布的概率密度函数：

$$\psi(y) = \begin{cases} \dfrac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n_1/2} y^{n_1/2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1+(n_1 y/n_2)]^{(n_1+n_2)/2}}, & y > 0 \\ 0, & \end{cases} \tag{20}$$

3 直方图和箱线图