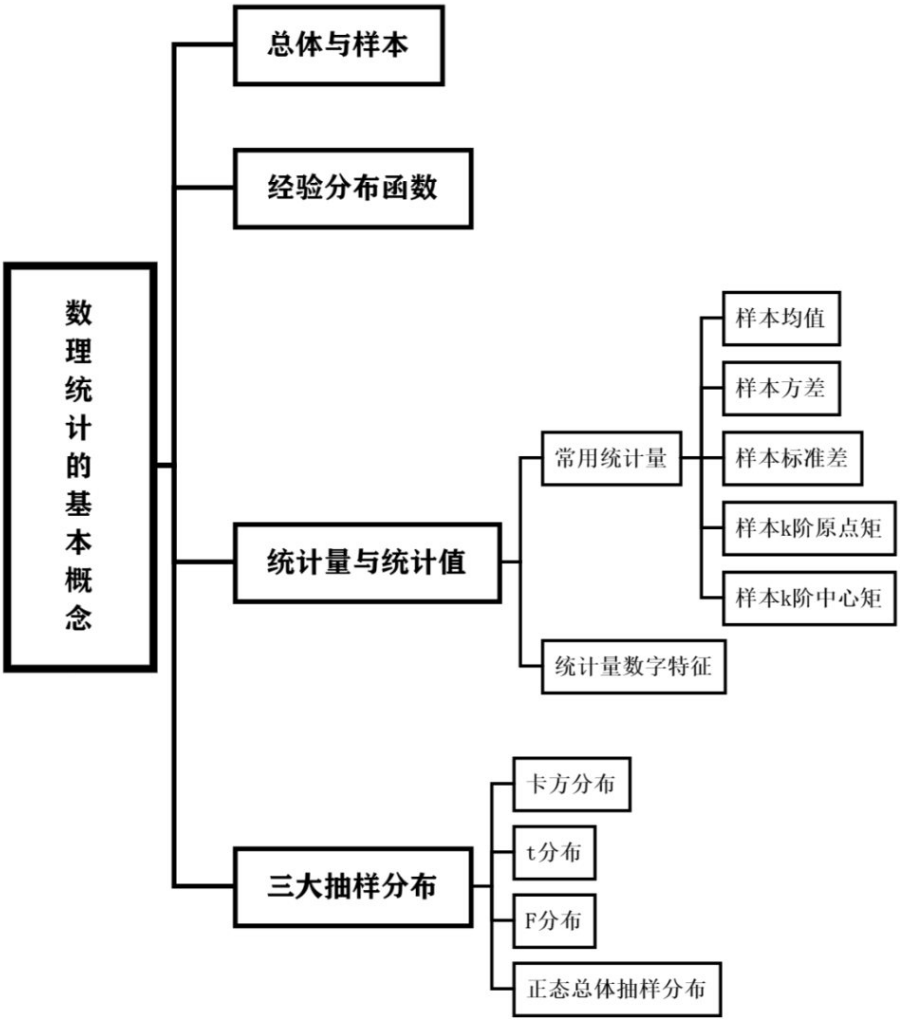
第六章 数理统计的基本概念

知识图解



知识讲解

一、 总体与样本

1. 总体：研究对象某项数量指标的全体称为**总体**. 构成总体的每个成员称为**个体**.

例如，研究一批机器的寿命，则全部机器的寿命构成问题的总体，每一台机器的寿命是一个个体，总体是寿命 X 服从的分布.

注：总体分布的类型有时是明确的，统计的任务是确定未知参数，即参数估计；有时是不明确的，需要先对分布进行假定，再估计未知参数.

2. 样本：在相同条件下对总体 X 进行 n 次简单随机抽样，得到的 n 个观察结果. $X_1, X_2, \cdots, X_n$ **相互独立且同分布于总体** X，称 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个**简单随机样本**，简称**样本**，其中 n 称为**样本容量**. 抽样得到的一组实数记为 $x_1, x_2, \cdots, x_n$，称为**样本观察值**，简称**样本值.**

例如，从该批机器中随机抽 20 台测定其寿命，即得容量为 20 的样本观测值 $x_1, x_2, \cdots, x_{20}$. 抽取前无法预知每台样品的寿命，因此样本 $X_1, X_2, \cdots, X_{20}$ 是随机变量.

注：1）有些教材中为简单化处理，统一用 $x_1, x_2, \cdots, x_n$ 表示样本和样本值.

2）从总体中抽样的方法是多种多样的，本课程仅研究简单随机样本，即 $X_1, X_2, \cdots, X_n$ 相互独立且同服从于总体分布.

二、 经验分布函数

定义 设 $X_1, X_2, \cdots, X_n$ 为总体 X 的一个样本，其样本值为 $x_1, x_2, \cdots, x_n$，则称函数

$$F_n(x) = \frac{\{x_1, x_2, \cdots, x_n\text{中小于或等于}x\text{的个数}\}}{n} \qquad (-\infty < x < +\infty)$$

为样本值 $x_1, x_2, \cdots, x_n$ 的经验分布函数.

若已知样本值 $x_1, x_2, \cdots, x_n$ 的频数、频率分布表为

指标 X	x_1^*	x_2^*	…	x_l^*
频数 n_i	n_1	n_2	…	n_l
频率 f_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_l}{n}$

则经验分布函数

$$F_n(x)=\begin{cases}0, & x<x_1^*,\\ \dfrac{n_1+\cdots+n_i}{n}, & x_i^*\le x<x_{i+1}^*,\quad (i=1,2,\cdots,l-1)\\ 1 & x\ge x_l^*.\end{cases}$$

格里汶科定理：对于任意实数 x，当 $n\to\infty$ 时 $F_n(x)$ 以概率 1 一致收敛于分布函数 $F(x)$. 即当 n 充分大时，经验分布函数的任一个观察值 $F_n(x)$ 是总体分布函数 $F(x)$ 的良好近似，从而实际上可当作 $F(x)$ 来使用.

【例6.1】 设总体 X 的容量为10的一组样本值是 $1,2,4,3,4,2,3,4,4,2$，试求样本的经验分布函数.

注：统计研究中，通常先对样本值进行分组，确定各组的组距组限，再列出频数频率分布表，并借助直方图、茎叶图等图形展示样本数据.

三、 统计量与统计值

定义 不含任何未知参数的样本函数 $g(X_1,X_2,\cdots,X_n)$ 称为**统计量**. 设 $x_1,x_2,\cdots x_n$ 是对应于样本 $X_1,X_2,\cdots,X_n$ 的样本值，则称 $g(x_1,x_2,\cdots x_n)$ 是 $g(X_1,X_2,\cdots,X_n)$ 的观测值，称为**统计值.**

【例6.2】 设总体 $X\sim N(\mu,\sigma^2)$，其中 μ 已知，σ^2 未知，X_1,X_2 是取自总体 X 的样本，则下列各量为统计量的是（　　）

（A）$X_1+\mu X_2$　（B）$2X_1+\sigma\mu$　（C）$X_1+\mu+\sigma^2$　（D）$\dfrac{X_1-\mu}{\sigma}$

四、 常用统计量

设$X_1,X_2,\cdots,X_n$为总体X的一个样本，n为样本容量

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

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

注：推导过程

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

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

4）样本k阶原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$，$k=1,2,\cdots$，

5）样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$，$k=1,2,\cdots$，

【例6.3】 从一批零件中抽取6个样本，测得其直径为1.5, 2, 2.3, 1.7, 2.5, 1.8，求$\overline{x},s^2$.

五、 常用统计量的数字特征

已知$X_1,X_2,\cdots,X_n$为来自总体X的样本，$EX=\mu,DX=\sigma^2$

1. $EX_i=\mu,\ DX_i=\sigma^2$;
2. $E\overline{X}=\mu,\ D\overline{X}=\frac{\sigma^2}{n}$;
3. $ES^2=\sigma^2$

【例6.4】 设总体 $X \sim B(m,\theta)$， $X_1, X_2 \dots, X_n$ 为来自该总体的简单随机样本，

$\overline{X}$ 为样本均值，则 $E\left[\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]=$（ ）

(A) $(m-1)n\theta(1-\theta)$ (B) $m(n-1)\theta(1-\theta)$

(C) $(m-1)(n-1)\theta(1-\theta)$ (D) $mn\theta(1-\theta)$

【例6.5】 设总体 X 服从参数为 $\lambda(\lambda>0)$ 的泊松分布， $X_1, X_2 \dots, X_n$ 为来自该总体的简单随机样本，则对于统计量 $T_1=\frac{1}{n}\sum_{i=1}^{n}X_i, T_2=\frac{1}{n-1}\sum_{i=1}^{n-1}X_i+\frac{1}{n}X_n$, 有（ ）.

（A） $ET_1>ET_2, DT_1>DT_2$ （B） $ET_1>ET_2, DT_1<DT_2$

（C） $ET_1<ET_2, DT_1>DT_2$ （D） $ET_1<ET_2, DT_1<DT_2$

六、 三大抽样分布与分位点

统计量的分布称为**抽样分布**，下面介绍来自正态总体的三大常用统计量的分布.

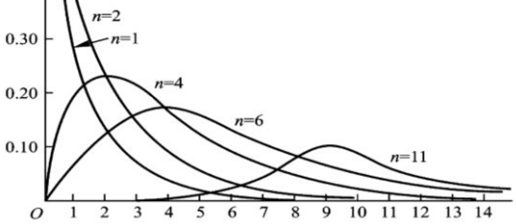
1. χ^2 分布

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

$$\chi^2=X_1^2+X_2^2+\cdots+X_n^2$$

服从自由度为 n 的 χ^2 分布，记为 $\chi^2 \sim \chi^2(n)$.

图像如下：



性质

1）可加性： 设 $\chi_1^2 \sim \chi^2(n_1)$，$\chi_2^2 \sim \chi^2(n_2)$，并且 χ_1^2，χ_2^2 独立，则 $\chi_1^2+\chi_2^2 \sim \chi^2(n_1+n_2)$

2）数学期望和方差（背）： 若 $\chi^2 \sim \chi^2(n)$，则有 $E(\chi^2)=n$，$D(\chi^2)=2n$

3）分位点

分位点定义 设有分布函数 $F(x)$，对给定的 $\alpha(0<\alpha<1)$，若有

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

则称点 x_α 为 $F(x)$ 的上 α 分位点.

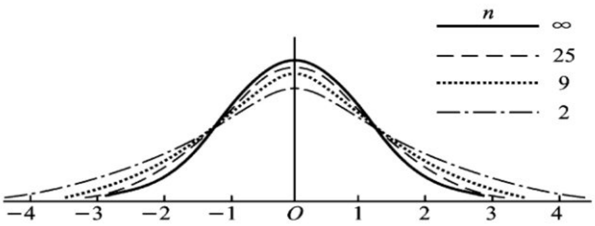
$\chi^2(n)$ 的分位点 设 $\chi^2 \sim \chi^2(n)$，对给定的 $\alpha(0<\alpha<1)$，若有 $\chi_\alpha^2(n)>0$ 满足 $P\{\chi^2>\chi_\alpha^2(n)\}=\alpha$，则称 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的上 α 分位点.

2. t 分布

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

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

服从自由度为 n 的 t 分布，记为 $t \sim t(n)$



性质

1）奇偶性： t 分布的概率密度 $f(t)$ 图像为**偶函数**.

2）分位点 对于给定的 $\alpha(0<\alpha<1)$，称满足条件 $P\{t>t_\alpha(n)\}=\alpha$ 的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位点. 易得 $t_{1-\alpha}(n)=-t_\alpha(n)$.

结论

1）自由度$n=1$的t分布是标准柯西分布，它的期望不存在.

2）自由度$n>1$的t分布期望存在且为0.

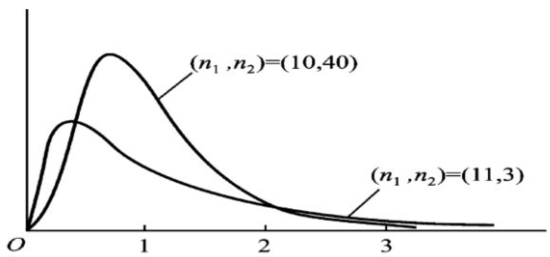
3）当自由度较大（$n\geq 30$）时，t分布与标准正态分布$N(0,1)$近似.

3. F分布

定义 设$X\sim\chi^2(n_1)$，$Y\sim\chi^2(n_2)$，且X,Y独立，则称随机变量

$$F=\frac{X/n_1}{Y/n_2}$$

服从自由度为(n_1,n_2)的F分布，记为$F\sim F(n_1,n_2)$，n_1称为**第一自由度**，n_2称为**第二自由度.**



性质 若$F\sim F(n_1,n_2)$，则$\dfrac{1}{F}\sim F(n_2,n_1)$

分位点 对于给定的$\alpha(0<\alpha<1)$，称满足条件$P\{F>F_\alpha(n_1,n_2)\}=\alpha$的点$F_\alpha(n_1,n_2)$为$F(n_1,n_2)$分布的上$\alpha$分位点. 易得$F_{1-\alpha}(n_1,n_2)=\dfrac{1}{F_\alpha(n_2,n_1)}$

【例6.6】 设X_1,X_2,X_3,X_4为来自总体$X\sim N(0,1)$的简单随机样本，若统计量$Y=\dfrac{X_1^2+X_2^2}{X_3^2+X_4^2}$，且对给定的$\alpha(0<\alpha<1)$，数$y_\alpha$满足$P\{Y>y_\alpha\}=\alpha$，则（ ）.

（A）$y_\alpha y_{1-\alpha}=1$ （B）$y_\alpha y_{1-\frac{\alpha}{2}}=1$

（C）$y_{\frac{\alpha}{2}}y_{1-\frac{\alpha}{2}}=1$ （D）$y_{\frac{\alpha}{2}}y_{\frac{1-\alpha}{2}}=1$

【例6.7】 设$X_1,X_2,\cdots,X_5$为总体$X\sim N(0,1)$的一个样本，设

$Y=\left(X_1+X_2+X_3\right)^2+\left(X_4-\sqrt{2}X_5\right)^2$且$cY$服从$\chi^2$分布，则$c=$__________.

【例6.8】 设随机变量X,Y相互独立，均服从$N(0,3^2)$分布且$X_1,X_2,...,X_9$与

$Y_1,Y_2,...,Y_9$分别是来自总体X,Y的简单随机样本，则统计量$U=\dfrac{X_1+...+X_9}{\sqrt{Y_1^2+...+Y_9^2}}$服

从参数为___的____分布.

【例6.9】 设总体 X 服从正态分布$X\sim N(0,2^2)$,而$X_1,X_2,...,X_{15}$是来自总体的

简单随机样本，则随机变量 $Y=\dfrac{X_1^2+...+X_{10}^2}{2(X_{11}^2+...+X_{15}^2)}$服从____分布，参数

为______.

【例6.10】（练习）设X_1,X_2,X_3,X_4为来自总体$N(1,\sigma^2)$的简单随机样本，则统计量$\dfrac{X_1-X_2}{|X_3+X_4-2|}$的分布为（　　）

（A）$N(0,1)$　　（B）$t(1)$　　（C）$\chi^2(1)$　　（D）$F(1,1)$

七、　正态总体抽样分布

一个正态总体 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，样本均值与样本方差分别是$\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$，$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，则

1）$\overline{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$，即$\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim N(0,1)$；

2）样本均值$\overline{X}$与样本方差S^2相互独立；

3）$\dfrac{(n-1)S^2}{\sigma^2}=\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1)$；$\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$；

4）$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$；

注：由$\sum_{i=1}^{n}(X_i-\overline{X})=0$，故$n$个偏差$X_1-\overline{X},X_2-\overline{X},\cdots,X_n-\overline{X}$中有$n-1$个为自由变量，第$n$个不可自由取值．故自由度为$n-1$

【例6.11】在总体$X\sim N(5,16)$中随机地抽取一个容量为 36 的样本$X_1,X_2,\cdots,X_{36}$，则关于样本均值$\overline{X}$的概率$P\left\{4\le\overline{X}\le 6\right\}=$____________.

【例6.12】 设$X_1,X_2,\dots,X_n\,(n\ge 2)$为来自总体$N(0,1)$的简单随机样本，$\overline{X}$为样本均值，$S^2$为样本方差，则（　　）

（A）$n\overline{X}\sim N(0,1)$ （B）$nS^2\sim\chi^2(n)$

（C）$\dfrac{(n-1)\overline{X}}{S}\sim t(n-1)$ （D）$\dfrac{(n-1)X_1^2}{\sum\limits_{i=2}^{n}X_i^2}\sim F(1,n-1)$

【例6.13】 设$X_1,X_2\cdots X_n$是来自总体$N(\mu,\sigma^2)$的简单随机样本,记

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

（I）证明$ET=\mu^2$. （II）当$\mu=0,\sigma=1$时,求DT.

两个正态总体 设$X_1,\cdots,X_{n_1}$是取自总体$X\sim(\mu_1,\sigma_1^2)$的一个样本，$Y_1,\cdots,Y_{n_2}$是取自总体$Y\sim(\mu_2,\sigma_2^2)$的一个样本，且这两个样本相互独立，即假定$X_1,\cdots,X_{n_1}$，$Y_1,\cdots,Y_{n_2}$是n_1+n_2个相互独立的随机变量，则有

1) $\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$

证明：$\overline{X}\sim N\left(\mu_1,\dfrac{\sigma_1^2}{n_1}\right),\overline{Y}\sim N\left(\mu_2,\dfrac{\sigma_2^2}{n_2}\right)$，$\overline{X}$与$\overline{Y}$相互独立，故

$$\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)，\text{标准化有}\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1).$$

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

证明：$\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1),\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$，且它们相互独立，故

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

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

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\cdot\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中

$$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2,\quad S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2,\quad S_W^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

证明：由$\dfrac{(n_1-1)S_1^2}{\sigma^2}\sim\chi^2(n_1-1),\dfrac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_2-1)$，且相互独立，根据可加性，

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

由$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N(0,1)$，且$\overline{X}-\overline{Y}$与$S_W^2$相互独立，故

$$\frac{\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}}{\sqrt{\dfrac{(n_1+n_2-2)S_W^2}{\sigma^2}/(n_1+n_2-2)}}=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\cdot\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2).$$