第 7 讲　统计量与抽样分布

知识梳理

· 事先说明：数理统计部分中的 $X_1,\cdots,X_n$ 都是独立同分布的随机变量

一　常见统计量

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

样本标准差	样本 k 阶矩	样本 k 阶中心矩
$S=\sqrt{S^2}$	$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$	$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$

· 比较重要的是样本均值和样本方差

1.样本均值的性质

样本均值的期望	样本均值的方差
$E(\overline{X})=E(X)$	$D(\overline{X})=\frac{D(X)}{n}$

2.样本方差的性质

样本方差的期望
$E(S^2)=D(X)$

二　三个抽样分布

1. χ^2 分布

χ^2 分布
$X_i\sim N(0,1)\rightarrow \boxed{Y=\sum_{i=1}^{n}X_i^2\sim\chi^2(n)}$

① 期望和方差

χ^2 分布的期望	χ^2 分布的方差
$E(Y)=n$	$D(Y)=2n$

② 性质

· 若 $Y_1\sim\chi^2(m)$， $Y_2\sim\chi^2(n)$ 且相互独立，则 $Y_1+Y_2\sim\chi^2(m+n)$

2. t 分布

t 分布

$X \sim N(0,1)$，$Y \sim \chi^2(n)$ 且相互独立 → $\boxed{t = \dfrac{X}{\sqrt{Y/n}} \sim t(n)}$

2. F 分布

t 分布

$Y_1 \sim \chi^2(n_1)$，$Y_2 \sim \chi^2(n_2)$ 且相互独立 → $\boxed{F = \dfrac{Y_1/m}{Y_2/n} \sim F(m,n)}$

· $\dfrac{1}{F} \sim F(n,m)$，若 $X \sim t(n)$，则 $X^2 \sim F(1,n)$

三 正态总体下抽样分布的特有性质

1.均值和方差满足分布

· 若 $X_i \sim N(\mu,\sigma^2)$，则 $\overline{X}$ 和 S^2 相互独立，且

正态总体的样本均值	正态总体的样本方差
$\boxed{\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n})}$	$\boxed{\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$

2.正态总体的样本方差的方差

正态总体的样本方差的方差

$$\boxed{D(S^2) = \frac{2\sigma^4}{n-1}}$$

题型解析

十二　求解统计量的性质

1.题型简述与解法

- 通常以填空题的形式考查由样本、样本均值、样本方差等构成的统计量的事件概率、期望、方差等
- 经常用到的知识点：
  - ① 期望、方差、协方差的运算性质
  - ② 正态分布的性质，以及正态变量的线性组合依然是正态变量
  - ③ 样本均值、样本方差的期望和方差，以及正态总体下的特殊性质
- 题干经常会有一大串的式子，其中一些实际上是样本方差的展开式，要能看出来
- 需要注意：不是所有的题目都是正态分布，要看清楚题干再使用正态总体下的结论

2.历年考试典型例题

① **计算协方差或相关系数**

例 1　（15−16 秋冬）设总体 $X\sim N(0,4)$， $X_1,\cdots,X_{100}$ 为来自 X 的简单随机样本，则 $\sum_{i=1}^{60}X_i$ 与 $\sum_{i=51}^{100}|X_i|$ 的相关系数为__________。

解　根据协方差的性质：

$$\mathrm{Cov}(\sum_{i=1}^{60}X_i,\sum_{i=51}^{100}|X_i|)=\sum_{i=1}^{60}\sum_{j=51}^{100}\mathrm{Cov}(X_i,|X_j|)$$

由于 X_i 之间相互独立，因此

$$\sum_{j=51}^{60}\mathrm{Cov}(X_i,|X_i|)$$

由于

$$\mathrm{Cov}(X_i,|X_i|)=E(X_i|X_i|)-E(X_i)E(|X_i|)$$

$$E(X_i|X_i|)=\int_{-\infty}^{+\infty}x\,|\,x\,|\,f(x)\,\mathrm{d}x=0\text{（}x\,|\,x\,|\,f(x)\text{ 为奇函数）}$$

$$E(X_i)=0$$

因此 $\mathrm{Cov}(X_i,|X_i|)=0\ \rightarrow\ \mathrm{Cov}(\sum_{i=1}^{60}X_i,\sum_{i=51}^{100}|X_i|)=0\ \rightarrow$ 相关系数为 0

② **计算事件概率**

例 2　（19−20 秋冬）设总体 $X\sim N(\mu,1)$， $X_1,\cdots,X_{16}$ 是 X 的简单随机样本， $\overline{Y_1}=(X_1+\cdots+X_4)/4$ ， $\overline{Y_2}=(X_5+\cdots+X_{16})/12$ ，则 $P(|\overline{Y_1}-\mu|<1)=$__________.

解　根据正态总体的性质， $\overline{Y_1}\sim N(\mu,\frac{1}{4})$，则由正态分布的性质， $\overline{Y_1}-\mu\sim N(0,\frac{1}{4})$

因此 $P(\left|\overline{Y_1}-\mu\right|<1)=P(|\frac{\overline{Y_1}-\mu}{1/2}|<\frac{1}{1/2})=2\Phi(2)-1$

例 3 （16—17 春夏）设总体 $X\sim N(\mu,\sigma^2)$，μ,σ^2 未知，$X_1,\cdots,X_{16}$ 为来自 X 的简单随机样本，$\overline{X}$ 和 S^2 分别为样本均值与样本方差，则 $P\left(\frac{X_1+X_2}{2}-\overline{X}>\sigma\right)=$_________；

解 由于 $\overline{X}$ 中包含了 X_1 和 X_2，需要注意减号左右两边不是相互独立的

$$\because\ \frac{X_1+X_2}{2}\sim N(\mu,\frac{\sigma^2}{2}),\ \overline{X}\sim N(\mu,\frac{\sigma^2}{16}),\ 记\ Y=\frac{X_1+X_2}{2}-\overline{X}$$

$$\therefore\ E(Y)=\mu-\mu=0,\ D(Y)=\frac{\sigma^2}{2}+\frac{\sigma^2}{16}-2\mathrm{Cov}(\frac{X_1+X_2}{2},\overline{X})$$

$$\because\ \mathrm{Cov}(\frac{X_1+X_2}{2},\overline{X})=\mathrm{Cov}(\sum_{i=1}^{2}\frac{1}{2}X_i,\sum_{j=1}^{16}\frac{1}{16}X_j)=\frac{1}{32}\mathrm{Cov}(\sum_{i=1}^{2}X_j,\sum_{i=1}^{16}X_j)=\frac{D(X_1)+D(X_2)}{32}=\frac{\sigma^2}{16}$$

$$\therefore\ D(Y)=\frac{\sigma^2}{2}+\frac{\sigma^2}{16}-2\frac{\sigma^2}{16}=\frac{7}{16}\sigma^2\quad\therefore\ Y\sim N(0,\frac{7}{16}\sigma^2)$$

$$\therefore\ P(Y>\sigma)=1-\Phi(4/\sqrt{7})=1-\Phi(1.51)=0.07$$

② **计算均值和方差**

例 4 （20—21 秋冬）设总体 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 是 X 的简单随机样本，$\overline{X}$，S^2 分别是样本均值和样本方差，若 $\mu=0$，则 $\mathrm{Var}(S^2-\overline{X}^2)=$__________.

解 根据正态总体的性质，S^2 与 $\overline{X}$ 相互独立

$$\because\ \overline{X}\sim N(0,\frac{\sigma^2}{n})\ \to\ \frac{\overline{X}}{\sigma/\sqrt{n}}\sim N(0,1)\ \to\ \frac{n}{\sigma^2}\overline{X}^2\sim\chi^2(1)$$

$$\therefore\ \mathrm{Var}(\frac{n}{\sigma^2}\overline{X}^2)=2\ \to\ \frac{n^2}{\sigma^4}\mathrm{Var}(\overline{X}^2)=2\ \to\ \mathrm{Var}(\overline{X}^2)=\frac{2\sigma^4}{n^2}$$

$$\because\ \mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1}\ \to\ \mathrm{Var}(S^2-\overline{X}^2)=\frac{2\sigma^4}{n-1}-\frac{2\sigma^4}{n^2}$$

例 5 （19—20 秋冬）设总体 $X\sim N(\mu,1)$，$X_1,\cdots,X_{16}$ 是 X 的简单随机样本，$\overline{Y_1}=(X_1+\cdots+X_4)/4$，$\overline{Y_2}=(X_5+\cdots+X_{16})/12$，则 $\mathrm{Var}\left[\sum_{i=1}^{4}(X_i-\overline{Y_1})^2+\sum_{i=5}^{16}(X_i-\overline{Y_2})^2\right]=$_________.

解 注意到 $\sum_{i=1}^{4}(X_i-\overline{Y_1})^2=(4-1)S_1^2$，$\sum_{i=5}^{16}(X_i-\overline{Y_2})^2=11S_2^2$，且两者相互独立（没有重叠的 X_i）

由正态总体 $\mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1}$，得

$$\mathrm{Var}\left[\sum_{i=1}^{4}(X_i-\overline{Y_1})^2+\sum_{i=5}^{16}(X_i-\overline{Y_2})^2\right]=\mathrm{Var}\left[\sum_{i=1}^{4}(X_i-\overline{Y_1})^2\right]+\mathrm{Var}\left[\sum_{i=5}^{16}(X_i-\overline{Y_2})^2\right]$$

$$=\mathrm{Var}(3S_1^2)+\mathrm{Var}(11S_2^2)=3^2\times\frac{2}{3}+11^2\times\frac{2}{11}=28$$

十三　判断统计量函数的分布

1.题型简述与解法

- 填空题，判断统计量函数属于哪种抽样分布，或已知含未知数的统计量函数满足某分布，求该未知数
- 既然是判断抽样分布，则总体一定是正态分布，以下结论都是适用的：

常用结论 1	常用结论 2	常用结论 3
$\boxed{\dfrac{n(\overline{X}-\mu)^2}{\sigma^2}\sim\chi^2(1)}$	$\boxed{\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)}$	$\boxed{\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2\sim\chi^2(n-1)}$

常用结论 3 实际上是 S^2 的展开式，因为题目通常给出方差的展开式而不是 S^2

- 如果 $\mu=0$，此时要意识到 X_i^2 可以除以 σ^2 变成 χ^2 分布（或者用常用结论 2）
- 大多数题目会考察 F 分布，即会把以上分布作比值处理，此时 σ^2、n 等参数都会被合并或约掉，使我们一眼看不出来，这时就要根据以上结论将给的式子拆成两个 χ^2 分布

2.历年考试典型例题

例 1　（15−16 **春夏**）设总体 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 为来自 X 的简单随机样本，则

$$\sum_{i=1}^{3}\left(X_i-\frac{X_1+X_2+X_3}{3}\right)^2\Big/\sum_{i=4}^{5}(X_i-\mu)^2\sim\_\_\_\_\_\_\_\_\_\_\text{分布}.$$

解　$\sum\limits_{i=1}^{3}\left(X_i-\dfrac{X_1+X_2+X_3}{3}\right)^2$ 具有 3 个样本的方差形式，根据常用结论 3：

$$\frac{1}{\sigma^2}\sum_{i=1}^{3}\left(X_i-\frac{X_1+X_2+X_3}{3}\right)^2\sim\chi^2(2)$$

根据常用结论 2：

$$\frac{1}{\sigma^2}\sum_{i=4}^{5}(X_i-\mu)^2\sim\chi^2(2)$$

两者相互独立，因此

$$\sum_{i=1}^{3}\left(X_i-\frac{X_1+X_2+X_3}{3}\right)^2\Big/\sum_{i=4}^{5}(X_i-\mu)^2=\frac{\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{3}\left(X_i-\dfrac{X_1+X_2+X_3}{3}\right)^2/2}{\dfrac{1}{\sigma^2}\sum\limits_{i=4}^{5}(X_i-\mu)^2/2}\sim F(2,2)$$

例 2　（15−16 **秋冬**）设总体 $X\sim N(0,4)$，$X_1,\cdots,X_{100}$ 为来自 X 的简单随机样本，则 $\sum\limits_{i=1}^{100}|X_i|$ 近似服从________分布，$\left(\sum\limits_{i=1}^{50}X_i\right)^2/\sum\limits_{i=51}^{100}X_i^2\sim$________分布.

解　$\sum\limits_{i=1}^{50}X_i$ 具有前 50 个样本的均值形式，根据常用结论 1：

$$\frac{50(\frac{1}{50}\sum_{i=1}^{50}X_i)^2}{\sigma^2}=\frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}\sim\chi^2(1)$$

注意到$\mu=0$，因此$\sum_{i=51}^{100}X_i^2=\sum_{i=51}^{100}(X_i-0)^2$，根据常用结论 2：

$$\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2\sim\chi^2(50)$$

∴ 两者相互独立：$\left(\sum_{i=1}^{50}X_i\right)^2/\sum_{i=51}^{100}X_i^2=\frac{50\cdot\frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}}{\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2}=\frac{\frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}}{\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2/50}\sim F(1,50)$

例 3 （15－16 **秋冬**）设总体$X\sim N(\mu,1)$，μ未知，$X_1,\cdots,X_{25}$为来自X的简单随机样本，$\overline{X}$是样本均值，若$a(\overline{X}-X_1)^2\sim\chi^2(1)$，则$a=$______________.

解 ∵ $\overline{X}\sim N(\mu,\frac{1}{25})$ $X_1\sim N(\mu,1)$ $\mathrm{Cov}(\overline{X},X_1)=\frac{1}{25}D(X_1)=\frac{1}{25}$

∴ $D(\overline{X}-X_1)=\frac{1}{25}+1-2\times\frac{1}{25}=\frac{24}{25}$ → $\overline{X}-X_1\sim N(0,\frac{24}{25})$

∴ 由常用结论 2，$\frac{1\times(\overline{X}-X_1)^2}{24/25}\sim\chi^2(1)$，比对得到$a=\frac{25}{24}$