概率论第六章答案

1. 设总体 X 服从正态分布 $N(\mu,\sigma^2)\,(\sigma>0)$, 从总体中抽取简单随机样本 $X_1,\cdots,X_{2n}(n\geqslant 2)$, 其样本均值为 $\bar{X}=\frac{1}{2n}\sum_{i=1}^{2n}X_i$, 求统计量 $Y=\sum_{i=1}^{n}(X_i+X_{n+i}-2\bar{X})^2$ 的数学期望.

   解:
   注意到 X_i+X_{n+i} 相互独立，同分布 $N(2\mu,2\sigma^2)$, 则它们可认为是取自同一正态总体 $N(2\mu,2\sigma^2)$ 的样本，其样本均值为
   $$\frac{1}{n}\sum_{i=1}^{n}(X_i+X_{n+i})=\frac{1}{n}\sum_{i=1}^{2n}X_i=2\bar{X}$$
   如果记 $Z_i=X_i+X_{n+i},i=1,\cdots,n$, 即 $Z_i(i=1,\cdots,n)$ 是取自 $N(2\mu,2\sigma^2)$ 的样本，且
   $$\frac{Y}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i+X_{n+i}-2\bar{X}\right)^2=S^2(Z)$$
   则有　$E\left(S^2(Z)\right)=\frac{1}{n-1}E(Y)=2\sigma^2$,
   所以　$E(Y)=2(n-1)\sigma^2$.

2. 设总体 X 服从参数为 λ 的指数分布，$X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个样本，求:
   (1)$\bar{X}$ 的密度;
   (2)$\bar{X}$ 当 n$\to\infty$ 时的极限分布.

   解:
   (1) 利用 $2n\lambda\bar{X}\sim\chi^2(2n)$ 得
   $$f_{\bar{X}}(x)=2n\lambda f_{\chi^2(2n)}(2n\lambda x)=\begin{cases}0 & x\le 0\\ \frac{n^n\lambda^n}{\Gamma(n)}x^{n-1}e^{-n\lambda x} & x>0\end{cases}$$
   (2)$\bar{X}$ 的极限分布为正态分布 $N(\frac{1}{\lambda},(\frac{1}{\sqrt{n}\lambda})^2)$

3. 总体 $X\sim N(\mu,2^2)$,$X_1,X_2,\cdots,X_n$ 为 X 的一个样本，问样本容量 n 应取多大才能使 $E[|\bar{X}-\mu|]\le 0.1$.

   解:
   由 $\bar{X}\sim N(\mu,\frac{2^2}{n})$ 得 $U=\frac{\bar{X}-\mu}{\frac{2}{\sqrt{n}}}\sim N(0,1)$，于是
   $$\begin{aligned}E[|\bar{X}-\mu|]&=\frac{2}{\sqrt{n}}E[|U|]=\frac{2}{\sqrt{n}}\int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\\&=\frac{2}{\sqrt{n}}\times\frac{2}{\sqrt{2\pi}}\le 0.1.\end{aligned}$$
   得
   $$n\ge(\frac{4}{0.1\sqrt{2\pi}})^2\approx 254.8$$
   即 n 至少取 255.

4. 假设 $X_1,X_2,\cdots,X_9$ 是来自总体 $X\sim N(0,2^2)$ 的简单随机样本, 求系数 a,b,c, 使 $Q=a(X_1+X_2)^2+b(X_3+X_4+X_5)^2+c(X_6+X_7+X_8+X_9)^2$ 服从 χ^2 分布，并求其自由度.

   解:

由于 $X_1, X_2, \cdots, X_9$ 相互独立且取自总体 $X \sim N\left(0, 2^2\right)$, 则由正态分布的线性运算性质有

$$X_1+X_2 \sim N(0,8), X_3+X_4+X_5 \sim N(0,12)$$
$$X_6+X_7+X_8+X_9 \sim N(0,16)$$

于是，由 $\chi^2=\chi_1^2+\cdots+\chi_k^2$ 有

$$Q=\frac{\left(X_1+X_2\right)^2}{8}+\frac{\left(X_3+X_4+X_5\right)^2}{12}+\frac{\left(X_6+X_7+X_8+X_9\right)^2}{16} \sim \chi^2(3)$$

故 $a=1/8, b=1/12, c=1/16$, 自由度为 3 .

5. 设 $X_1, \cdots, X_n, X_{n+1}$ 是取自正态总体 $X \sim N\left(\mu, \sigma^2\right)$ 的样本,

$$\bar{X}_n=\frac{1}{n} \sum_{i=1}^n X_i, S_n=\frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}_n\right)^2$$

试确定统计量 $\sqrt{\frac{n}{n+1}} \cdot \frac{X_{n+1}-\bar{X}_n}{S_n}$ 的分布.

解:

将统计量改写成下列形式:

$$\sqrt{\frac{n}{n+1}} \cdot \frac{X_{n+1}-\bar{X}_n}{S_n}=\frac{\left(X_{n+1}-\bar{X}_n\right) / \sqrt{\left(1+\frac{1}{n}\right) \sigma^2}}{\sqrt{\frac{(n-1) S_n^2}{\sigma^2} /(n-1)}}$$

由于 X_{n+1} 与 $X_i(i=1, \cdots, n)$ 相互独立,

$$\bar{X}_n=\frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right), X_{n+1} \sim N\left(\mu, \sigma^2\right),$$

所以 $\quad X_{n+1}-\bar{X}_n \sim N\left(0,\left(1+\frac{1}{n}\right) \sigma^2\right) \quad,$

从而 $\quad\left(X_{n+1}-\bar{X}_n\right) /\left(\sqrt{\left(1+\frac{1}{n}\right) \sigma^2}\right) \sim N(0,1)$.

注意到 $\bar{X}_n$ 与 S_n^2 相互独立, X_{n+1} 也与 S_n^2 相互独立，且

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

故由统计量改写后形式得

$$\sqrt{\frac{n}{n+1}} \cdot \frac{X_{n+1}-\bar{X}_n}{S_n} \sim t(n-1)$$

6. 设 $X_1, X_2, \cdots, X_m$ 是来自 $N(\mu_1, \sigma^2)$ 的样本，$Y_1, Y_2, \cdots, Y_n$ 是来自 $N(\mu_2, \sigma^2)$ 的样本，且这两个样本相互独立，记 $S_x^2=\frac{1}{m-1} \sum_{i=1}^m(X_i-\bar{X})^2$，$S_y^2=\frac{1}{n-1} \sum_{i=1}^n(Y_i-\bar{Y})^2$，其中 $\bar{X}=\frac{1}{m} \sum_{i=1}^m X_i$，$\bar{Y}=\frac{1}{n} \sum_{i=1}^n Y_i$，记 $S_w^2=\frac{(m-1) S_x^2+(n-1) S_y^2}{m+n-2}$，$c$，$d$ 是任意两个不为零的常数，证明:

$$\frac{c(\bar{X}-\mu_1)+d(\bar{Y}-\mu_2))}{S_w \sqrt{\dfrac{c^2}{m}+\dfrac{d^2}{n}}} \sim t(m+n-2)$$

.

解: 根据题意, 有 $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{m})$, $\bar{Y} \sim N(0, \frac{\sigma_2^2}{n})$, 则 $c(\bar{X}-\mu_1) \sim N(0, \frac{c^2 \sigma^2}{m})$, $d(\bar{Y}-\mu_2) \sim N(0, \frac{d^2 \sigma^2}{n})$

由于 $\bar{X}$ 与 $\bar{Y}$ 相互独立，则有 $c(\bar{X}-\mu_1)+d(\bar{Y}-\mu_2)\sim N(0,\frac{c^2\sigma^2}{m}+\frac{d^2\sigma^2}{n})$，于是

$$U=\frac{c(\bar{X}-\mu_1)+d(\bar{Y}-\mu_2}{\sigma\sqrt{\frac{c^2}{m}+\frac{d^2}{m^2}}}\sim N(0,1)$$

.

根据题意，有 $\frac{(m-1)S_x^2}{\sigma^2}\sim\chi^2(m-1)$，$\frac{(n-1)S_y^2}{\sigma^2}\sim\chi^2(n-1)$，且他们相互独立，根据 χ^2 分布的可加性，有

$$V=\frac{(m+n-2)S_w^2}{\sigma^2}=\frac{(m-1)S_x^2}{\sigma^2}+\frac{(n-1)S_y^2}{\sigma^2}\sim\chi^2(m+n-2)$$

.

由于 $\bar{X}$，$\bar{Y}$，S_x^2，S_y^2 相互独立，于是 U,V 独立，根据 t 分布的定义，有

$$\frac{U}{\sqrt{\frac{V}{m+n-2}}}=\frac{c(\bar{X}-\mu_1)+d(\bar{Y}-\mu_2))}{S_w\sqrt{\frac{c^2}{m}+\frac{d^2}{n}}}\sim t(m+n-2)$$

.

7. 设 X_1,X_2,X_3,X_4 为来自总体 $N(1,\sigma^2)(\sigma>0)$ 的随机样本，则统计量 $\frac{X_1-X_2}{|X_3+X_4-2|}$ 的分布为（B）.

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

解:

由于

$$\frac{X_1-X_2}{|X_3+X_4-2|}=\frac{\frac{X_1-X_2}{\sqrt{2}\sigma}}{\sqrt{\left(\frac{X_3+X_4-2}{\sqrt{2}\sigma}\right)^2}}$$

,

根据正态分布的性质可知，$\frac{X_1-X_2}{\sqrt{2}\sigma}$ 和 $\frac{X_3+X_4-2}{\sqrt{2}\sigma}$ 服从标准正态分布且相互独立，则有

$$\frac{\frac{X_1-X_2}{\sqrt{2}\sigma}}{\sqrt{\left(\frac{X_3+X_4-2}{\sqrt{2}\sigma}\right)^2}}\sim t(1)$$

.

8. 随机变量 $X\sim t(n)$,$Y\sim F(1,n)$, 给定 $\alpha(0<\alpha<0.5)$, 常数 c 满足 P(X>c)=α, 则 $P(Y>c^2)$=(c)

a) α　　b) $1-\alpha$　　c) 2α　　d) $1-2\alpha$

9. 分别从方差为 20 和 35 的正态总体中抽取容量为 8 和 10 的两个样本，求第一个样本方差不小于第二个样本方差的两倍的概率.

解：用 S_1^2 和 S_2^2 分别表示两个样本方差，由定理知

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{S_1^2/20}{S_2^2/35}\sim F(8-1,10-1)=F(7,9)$$

.

又设事件 $A=\{S_1^2\geqslant 2S_2^2\}$，下面求 $P\{S_1^2\geqslant 2S_2^2\}$，因

$$P\{S_1^2\geqslant 2S_2^2\}=P\left\{\frac{S_1^2}{S_2^2}\geqslant 2\right\}$$

故有

$$P\{\frac{S_1^2/20}{S_2^2/35}\geqslant 2\times\frac{35}{20}\}=P\{F\geq 3.5\}$$

查 F 分布表得到自由度为 $n_1=7;n_2=9$ 的 F 分布上 α 分位点 $F_\alpha(7,9)$ 有如下数值:

$$F_{0.05}(7,9)=3.29,F_0.025(7,9)=4.20,$$

即事件 A 的概率介于 0.025 和 0.05 之间.

第 10 题

对$T \sim t(5)$,

1）求满足$P(-k_1 \le T \le k_2) = 0.90$，且$k_1 = k_2 = k$时的 k 值（自行查表）

2）说明其基于密度曲线的几何意义

3）请尝试查表找到一组$k_1 \ne k_2$的(k_1, k_2)值，并讨论这组值构成的区间和区间$(-k, k)$的长度关系

答

1）由 t 分布对称性

$$P(-k \le T \le k) = 0.90 \Rightarrow \ P(T > k) = P(T < k) = 0.05$$

$$k = t_{0.05}(5) = 2.015$$

2）几何意义：t 分布密度曲线下，区间$[-2.015, 2.015]$覆盖 90%的面积

3）对于任意区间，由于 T 分布表没有其他的小数概率分位点（例如$t_{0.03}(5)$，$t_{0.07}(5)$），所以要借助其他的现代计算机软件进行计算

例如 Excel

$$= t.inv(0.025,5)$$

给出 0.025 左侧单尾分位点

$$= t.inv.2T(0.05,5)$$

给出 0.05 的双尾分位点，所以 $t.inv.2T(0.05,5) =- t.inv(0.025,5)$

又例如，Python 中使用语句

$$t.ppf(p, dof); \quad t.ppf(1-p, dof); \quad t,ppf(1-\frac{p}{2}, dof)$$

计算左侧，右侧和双尾检验。

最后计算任意组合，例如$[-t_{0.07}(5), t_{0.03}(5)] = [-1.753, 2.421]$

该区间长度 为 $2.421 + 1.753 = 4.174 > 2.015 * 2 = 4.03$

实际上，在单峰，对称分布下，$[-2.015, 2.015]$为最短区间