第 6 章　数理统计的基本概念

6.1　知 识 要 点

6.1.1　总体与样本

研究对象的全体称为**总体**，总体中的每个成员（或元素）称为**个体**；一个总体对应一个随机变量 X， X 的分布函数和数字特征也称为总体的分布函数和数字特征. 从总体中抽取若干个成员，这些成员称为总体的一个**样本**；成员个数称为**样本容量**. 设总体 X 的分布函数为 $F(x)$，若样本 $X_1,X_2,\cdots,X_n$ 相互独立，且它们具有相同的分布函数 $F(x)$，则称 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的**简单随机样本，**简称**样本**.

6.1.2　统计量与抽样分布

设 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个样本，其观测值为 $x_1,x_2,\cdots,x_n$， $g(x_1,x_2,\cdots,x_n)$ 为 $x_1,x_2,\cdots,x_n$ 的函数，若 $g(x_1,x_2,\cdots,x_n)$ 不含有任何未知参数，则 $g(X_1,X_2,\cdots,X_n)$ 称为样本 $X_1,X_2,\cdots,X_n$ 的**统计量**，统计量的分布称为**抽样分布**.

6.1.3　一些常用的统计量

（1）样本均值： $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ ；

（2）样本方差： $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2=\dfrac{1}{n-1}\left(\sum\limits_{i=1}^{n}X_i^2-n\overline{X}^2\right)$ ；

（3）样本标准差： $S=\sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$ ；

（4）样本 k 阶原点矩： $A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$， $k=1,2,\ldots$ ；

（5）样本 k 阶中心矩： $B_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k$， $k=2,3,\cdots$.

6.1.4　经验分布函数

设 $X_1,X_2\cdots,X_n$ 是总体 X 的一个样本，经验分布函数 $F_n(x)$ 定义为

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}I(X_i\leqslant x)，\quad -\infty<x<+\infty .$$

若已知样本的观测值，则可以很容易地得到经验分布函数 $F_n(x)$ 的**观测值**.例如，设

$x_1, x_2 \cdots, x_n$ 是总体 X 的一个样本值，从小到大排列后，记为 $x_{(1)}, \cdots, x_{(n)}$，则经验分布函数 $F_n(x)$ 的**观测值**为

$$F_n(x)=\begin{cases}0, & x<x_{(1)}, \\ \dfrac{k}{n}, & x_{(k)} \leqslant x<x_{(k+1)}, \ k=1,2,\cdots,n-1, \\ 1, & x \geqslant x_{(n)} .\end{cases}$$

格里汶科定理：当 $n \to \infty$ 时，$F_n(x)$ 以概率 1 一致收敛于分布函数 $F(x)$，即

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

*6.1.5　顺序统计量

设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个样本，记

$$X_{(1)}=\min\{X_1, X_2, \cdots, X_n\}, \quad X_{(n)}=\max\{X_1, X_2, \cdots, X_n\}$$

则 $X_{(1)}$ 和 $X_{(n)}$ 分别称为样本的**最小顺序统计量**和**最大顺序统计量**．设总体 X 的分布函数为 $F(x)$，则 $X_{(1)}$ 和 $X_{(n)}$ 的分布函数分别为

$$F_1(x)=1-[1-F(x)]^n, \quad F_n(x)=F^n(x) .$$

若总体 X 的密度函数为 $f(x)$，则 $X_{(1)}$ 和 $X_{(n)}$ 的概率密度函数分别为

$$f_1(x)=n[1-F(x)]^{n-1} f(x), \quad f_n(x)=nF^{n-1}(x) f(x) .$$

6.1.6　χ^2 分布

设 $X_1, X_2, \ldots, X_n$ 相互独立且都服从标准正态分布 $N(0,1)$，称 $\chi^2=\sum\limits_{i=1}^{n} X_i^2$ 服从自由度为 n 的 $\boldsymbol{\chi^2}$ **分布**，记为 $\chi^2 \sim \chi^2(n)$．$\chi^2(n)$ 分布的密度函数 $f(y)$ 的图像如图 6.1 所示．

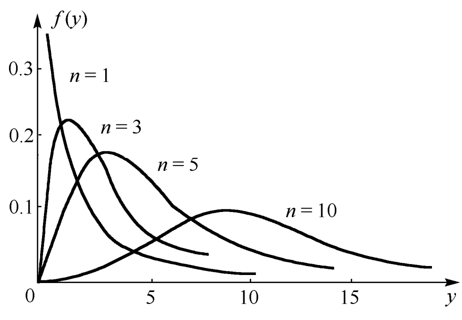

图 6.1　密度函数

χ^2 分布的性质：

（1）若 $\chi_1^2 \sim \chi^2(m)$，$\chi_2^2 \sim \chi^2(n)$，且 χ_1^2 与 χ_2^2 相互独立，则 $\chi_1^2+\chi_2^2 \sim \chi^2(m+n)$；

（2）若 $\chi^2 \sim \chi^2(n)$，则 $E(\chi^2)=n$，$D(\chi^2)=2n$．

6.1.7　t 分布

设 X 与 Y 相互独立，$X \sim N(0,1)$，$Y \sim \chi^2(n)$，称 $t = \dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 **t 分布**，记为 $t \sim t(n)$． $t(n)$ 分布的密度函数 $h(t)$ 如图 6.2 所示．

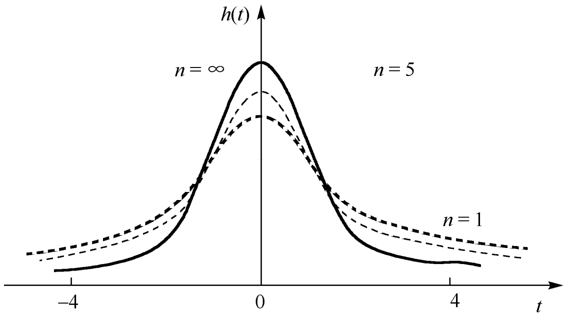

图 6.2　密度函数

t 分布的性质：

（1）t 分布的概率密度曲线 $h(t)$ 的图像关于直线 $t=0$ 对称，即 $h(t)$ 为偶函数；

（2）$\lim\limits_{n\to\infty} h(t) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$，故当 n 足够大时，t 分布近似于标准正态分布 $N(0,1)$；

（3）$E[t(n)] = 0$，$D[t(n)] = \dfrac{n}{n-2}(n>2)$．

6.1.8　F 分布

设 X 与 Y 相互独立，且 $X \sim \chi^2(m)$，$Y \sim \chi^2(n)$，称 $F = \dfrac{X/m}{Y/n}$ 服从自由度为 (m,n) 的 **F 分布**，记为 $F \sim F(m,n)$． $F(m,n)$ 分布的密度函数 $f(y)$ 如图 6.3 所示．

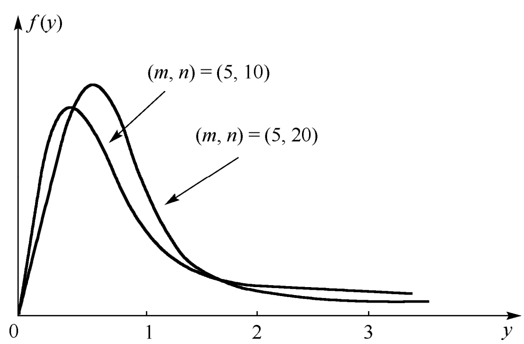

图 6.3　密度函数

F 分布的性质：

（1）若 $F \sim F(m,n)$，则 $\dfrac{1}{F} \sim F(n,m)$；

（2）若 $T \sim t(n)$，则 $T^2 \sim F(1,n)$．

6.1.9　上 α 分位点

我们曾在 2.1.6 节给出了连续型随机变量分布的上 α 分位点的概念，这里我们给出几个具体连续型分布的上 α 分位点.

若 $X \sim N(0,1)$，z_α 满足

$$P\{X > z_\alpha\} = 1 - \Phi(z_\alpha) = \int_{z_\alpha}^{+\infty} \phi(x)\mathrm{d}x,$$

则称 z_α 为 $N(0,1)$ 的上 α 分位点.

类似地，可以定义 $\chi^2(n)$ 分布的上 α 分位点 $\chi_\alpha^2(n)$，t 分布的上 α 分位点 $t_\alpha(n)$，$F(m,n)$ 分布的上 α 分位点 $F_\alpha(m,n)$ 等. 可以证明

$$z_{1-\alpha} = -z_\alpha,\quad t_{1-\alpha}(n) = -t_\alpha(n),\quad F_{1-\alpha}(n,m) = \frac{1}{F_\alpha(m,n)}.$$

6.1.10　正态总体的样本均值与样本方差的分布

性质 1（单总体情形）：设 $X_1, X_2 \cdots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ X 的一个样本，$\overline{X}$ 和 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 分别是样本均值和样本方差，则有

（1）$E(\overline{X}) = \mu$，$D(\overline{X}) = \frac{\sigma^2}{n}$，且 $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$，即 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$；

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

（3）$\overline{X}$ 与 S^2 相互独立，且 $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

性质 2（双总体情形）：设 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，且 X 与 Y 独立，$X_1, X_2, \cdots, X_m$ 和 $Y_1, Y_2, \cdots, Y_n$ 分别为来自 X 与 Y 的样本，$\overline{X}, \overline{Y}$ 分别是这两个样本的样本均值，$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2$，$S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ 分别是这两个样本的样本方差，则有

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

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

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

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_\omega\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} \sim t(m+n-2),$$

其中 $S_\omega = \sqrt{\dfrac{(m-1)S_1^2+(n-1)S_2^2}{(m+n-2)}}$.

6.1.11 几个常用的结论

（1）记总体 X 的 k 阶矩 $E(X^k)=\mu_k$， g 为连续函数，则当 $n\to\infty$ 时，有

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

$$g(A_1,A_2,\cdots A_k)\xrightarrow{P}g(\mu_1,\mu_2,\cdots\mu_k),\ k=1,2,\ldots.$$

（2）设总体 X 的均值为 μ，方差为 σ^2， $X_1,X_2\cdots,X_n$ 是来自总体 X 的一个样本， $\overline{X}$ ， $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 以及 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 分别是样本均值、样本方差和修正的样本方差，则有

$$E(\overline{X})=\mu,\quad D(\overline{X})=\frac{\sigma^2}{n},\quad E(S_n^2)=\frac{n-1}{n}\sigma^2,\quad E(S^2)=\sigma^2.$$

6.2 典型例题分析

6.2.1 题型一：统计量与抽样分布问题

例 6.2.1 设 $X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，其中 μ,σ^2 为未知参数，$\overline{X}$ 和 S^2 分别为样本均值和样本方差，则下列表达式为统计量的是（ ）.

（A） $T_1=\max\{X_1,X_2,\cdots,X_n\}$ （B） $T_2=\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$

（C） $T_3=\dfrac{(n-1)S^2}{\sigma^2}$ （D） $T_4=\overline{X}-E(\overline{X})$

解 注意到 $E(\overline{X})=\mu$，即 T_2,T_3,T_4 中都含有未知参数，因此 T_2,T_3,T_4 均不是统计量．故答案选 A.

例 6.2.2 【2002（3）】设随机变量 X 和 Y 都服从标准正态分布，则（ ）.

（A） $X+Y$ 服从正态分布 （B） X^2+Y^2 服从 χ^2 分布

（C） X^2 和 Y^2 服从 χ^2 分布 （D） X^2/Y^2 服从 F 分布

解 答案选 C．由于随机变量 X 和 Y 不一定相互独立，因此选项 B 和 D 错误；选项 A 具有一定的迷惑性，需要注意的是，有限个相互独立的正态随机变量的线性组合仍服从正态分布．这里 X 和 Y 不一定相互独立，因此 $X+Y$ 不一定服从正态分布．例如取 $Y=-X$，则 $P\{X+Y=0\}=1$，此时 $X+Y$ 服从退化分布.

例 6.2.3 设 $X_1,X_2,\cdots,X_n,X_{n+1}$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，且 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$， $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，试讨论统计量 $T=\dfrac{X_{n+1}-\overline{X}}{S}\sqrt{\dfrac{n}{n+1}}$ 的分布.

解 由题意，$a\overline{X}+bX_{n+1}$ 服从期望为 $(a+b)\mu$，方差为 $\left(\dfrac{a^2}{n}+b^2\right)\sigma^2$ 的正态分布．取 $a=-1$，

$b=1$，则 $-\overline{X}+X_{n+1}\sim N\left(0,\dfrac{n}{n+1}\sigma^2\right)$，即有

$$\frac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\frac{n}{n+1}}\sim N(0,1)$$

而

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

且 $\overline{X}$，X_{n+1} 均与 S^2 相互独立，因此 $\dfrac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\dfrac{n}{n+1}}$ 与 $\dfrac{(n-1)S^2}{\sigma^2}$ 相互独立，故

$$T=\frac{X_{n+1}-\overline{X}}{S}\sqrt{\frac{n}{n+1}}=\frac{\dfrac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\dfrac{n}{n+1}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}}\sim t(n-1)$$

例 6.2.4　设 $X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(0,1)$ 的一个样本，且 $\overline{X}$ 和 S^2 分别为样本均值和样本方差，试讨论统计量 $n(\overline{X})^2+(n-1)S^2$ 的分布.

解　由于 $\overline{X}\sim N\left(0,\dfrac{1}{n}\right)$，因此 $n(\overline{X})^2\sim\chi^2(1)$. 又因为

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

且 $\overline{X}$ 与 S^2 相互独立，因此 $n(\overline{X})^2$ 与 $\dfrac{(n-1)S^2}{\sigma^2}$ 相互独立，由 χ^2 分布的可加性知

$$n(\overline{X})^2+(n-1)S^2\sim\chi^2(n).$$

例 6.2.5　设 $X_1,X_2,\cdots,X_m,X_{m+1},\cdots,X_{m+n}$ 为来自总体 $X\sim N(0,\sigma^2)$ 的一个容量为 $m+n$ 的样本，试讨论统计量 $\dfrac{n}{m}\cdot\dfrac{X_1^2+X_2^2+\cdots+X_m^2}{X_{m+1}^2+X_{m+2}^2+\cdots+X_{m+n}^2}$ 的分布.

解　由于 $X_1,X_2,\cdots,X_m,X_{m+1},\cdots,X_{m+n}$ 均服从正态分布 $N(0,\sigma^2)$，因此

$$\frac{X_i}{\sigma}\sim N(0,1),\ i=1,2,3,\cdots,m+n.$$

又因为 $X_1,X_2,\cdots,X_m,X_{m+1},\cdots,X_{m+n}$ 相互独立，故

$$\frac{X_1^2+X_2^2+\cdots+X_m^2}{\sigma^2}\sim\chi^2(m),\quad \frac{X_{m+1}^2+X_{m+2}^2+\cdots+X_{m+n}^2}{\sigma^2}\sim\chi^2(n),$$

且 $\dfrac{X_1^2+X_2^2+\cdots+X_m^2}{\sigma^2}$ 与 $\dfrac{X_{m+1}^2+X_{m+2}^2+\cdots+X_{m+n}^2}{\sigma^2}$ 相互独立．根据 F 分布的定义，

$$\frac{n}{m}\cdot\frac{X_1^2+X_2^2+\cdots+X_m^2}{X_{m+1}^2+X_{m+2}^2+\cdots+X_{m+n}^2}=\frac{\dfrac{X_1^2+X_2^2+\cdots+X_m^2}{m\sigma^2}}{\dfrac{X_{m+1}^2+X_{m+2}^2+\cdots+X_{m+n}^2}{n\sigma^2}}\sim F(m,n).$$

6.2.2　题型二：概率的计算问题

例 6.2.6　设 $X_1,X_2,\cdots,X_5$ 为来自总体 X 的一个样本，且 X 服从 $(0,1)$ 上的均匀分布，试求：（1）$P\{X_{(5)}>0.5\}$；（2）$P\{X_{(1)}>0.5\}$.

解　（1）
$$\begin{aligned}P\{X_{(5)}>0.5\}&=1-P\{X_{(5)}\leqslant 0.5\}\\&=1-P\{X_1\leqslant 0.5\}P\{X_2\leqslant 0.5\}\cdots P\{X_5\leqslant 0.5\}\\&=1-(P\{X\leqslant 0.5\})^5=1-\left(\int_0^{0.5}1\mathrm{d}x\right)^5\approx 1-0.0313\approx 0.9687.\end{aligned}$$

（2）
$$\begin{aligned}P\{X_{(1)}>0.5\}&=P(X_1>0.5)P(X_2>0.5)\cdots P(X_5>0.5)\\&=(P\{X>0.5\})^5=[1-P\{X\leqslant 0.5\}]^5\approx 0.0313.\end{aligned}$$

例 6.2.7　现从总体 $N(10,4)$ 中分别抽取样本容量为 16 和 20 的两个样本，试求两样本均值差的绝对值大于 0.5 的概率.

解　令 $\overline{X}$ 和 $\overline{Y}$ 分别表示容量为 16 和 20 的两个样本的样本均值，则
$$\overline{X}\sim N\left(10,\frac{4}{16}\right),\quad \overline{Y}\sim N\left(10,\frac{4}{20}\right).$$
从而 $\overline{X}-\overline{Y}\sim N\left(0,\dfrac{4}{16}+\dfrac{4}{20}\right)$，即 $\overline{X}-\overline{Y}\sim N(0,0.45)$．因此
$$\begin{aligned}P\{|\overline{X}-\overline{Y}|>0.5\}&=P\left\{\frac{|\overline{X}-\overline{Y}|}{\sqrt{0.45}}>\frac{0.5}{\sqrt{0.45}}\right\}=2-2\Phi\left(\frac{0.5}{\sqrt{0.45}}\right)\\&\approx 2-2\Phi(0.745)\approx 0.453.\end{aligned}$$

6.2.3　题型三：随机变量的数字特征问题

例 6.2.8　【2003（3）】设总体 $X\sim N(\mu_1,\sigma^2)$，$Y\sim N(\mu_2,\sigma^2)$，$X_1,X_2,\cdots,X_m$ 和 $Y_1,Y_2,\cdots,Y_n$ 分别为来自总体 X 和 Y 的简单随机样本，则 $E\left[\dfrac{1}{m+n-2}\left(\sum\limits_{i=1}^m(X_i-\overline{X})^2+\sum\limits_{j=1}^n(Y_j-\overline{Y})^2\right)\right]=$
________________.

解　由于
$$\sum_{i=1}^m\frac{(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(m-1),\quad \sum_{j=1}^n\frac{(Y_j-\overline{Y})^2}{\sigma^2}\sim\chi^2(n-1),$$
因此
$$E\left(\sum_{i=1}^m\frac{(X_i-\overline{X})^2}{\sigma^2}\right)=m-1,\quad E\left(\sum_{j=1}^n\frac{(Y_j-\overline{Y})^2}{\sigma^2}\right)=n-1.$$
从而
$$E\left(\sum_{i=1}^m(X_i-\overline{X})^2\right)=(m-1)\sigma^2,\quad E\left(\sum_{j=1}^n(Y_j-\overline{Y})^2\right)=(n-1)\sigma^2,$$

故

$$E\left[\frac{1}{m+n-2}\left(\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{j=1}^{n}(Y_j-\overline{Y})^2\right)\right]=\sigma^2 .$$

例 6.2.9　设总体 X 的分布函数为 $F(x)$，经验分布函数为 $F_n(x)$，试证明

$$E[F_n(x)]=F(x)\text{，}\quad D[F_n(x)]=\frac{1}{n}F(x)[1-F(x)] .$$

证　记 $Y_i=I(X_i\leqslant x)$，由于

$$E(Y_i)=1\times P\{X_i\leqslant x\}+0\times P\{X_i>x\}=F(x)\text{，}$$

因此

$$E[F_n(x)]=\frac{1}{n}\sum_{i=1}^{n}E(Y_i)=\frac{1}{n}\sum_{i=1}^{n}F(x)=F(x) .$$

又因为

$$E(Y_i^2)=1^2\times P\{X_i\leqslant x\}+0^2\times P\{X_i>x\}=F(x)\text{，}$$

$$D(Y_i)=E(Y_i^2)-[E(Y_i)]^2=F(x)-F^2(x)=F(x)[1-F(x)]\text{，}$$

因此

$$D[F_n(x)]=\frac{1}{n^2}\sum_{i=1}^{n}D(Y_i)=\frac{1}{n}F(x)[1-F(x)] .$$

6.2.4　题型四：常数的求解问题

例 6.2.10　设 X_1,X_2,X_3,X_4 为来自总体 $X\sim N(1,4)$ 的样本，已知 $a(X_1-X_2)^2+b(X_3-2X_4+c)^2$ 服从自由度为 n 的 χ^2 分布，其中 a,b 均为不等于 0 的常数，试求常数 a,b,c 以及自由度 n 的值.

解　由于 X_1,X_2,X_3,X_4 相互独立且服从正态分布 $N(1,4)$，因此 X_1-X_2 和 X_3-2X_4+c 服从正态分布．其期望和方差分别为

$$E(X_1-X_2)=0\text{，}\quad D(X_1-X_2)=D(X_1)+D(X_2)=8\text{，}$$

$$E(X_3-2X_4+c)=c-1\text{，}\quad D(X_3-2X_4+c)=D(X_3)+4D(X_4)=20\text{，}$$

故

$$\frac{X_1-X_2}{2\sqrt{2}}\sim N(0,1)\text{，}\qquad \frac{X_3-2X_4+c}{2\sqrt{5}}\sim N(c-1,1) .$$

又因为 $\dfrac{X_1-X_2}{2\sqrt{2}}$ 和 $\dfrac{X_3-2X_4+c}{2\sqrt{5}}$ 相互独立，根据 χ^2 分布的定义可知 $c=1$，且

$$\left(\frac{X_1-X_2}{2\sqrt{2}}\right)^2+\left(\frac{X_3-2X_4+c}{2\sqrt{5}}\right)^2=\frac{1}{8}(X_1-X_2)^2+\frac{1}{20}(X_3-2X_4+1)^2\sim\chi^2(2) .$$

而由题设，$a(X_1-X_2)^2+b(X_3-2X_4+c)^2$ 服从 $\chi^2(n)$ 分布，因此 $a=\dfrac{1}{8}$，$b=\dfrac{1}{20}$，c=1，$n=2$.

6.2.5　题型五：经验分布函数的求解

例 6.2.11　设 $4,2,8,6,8,4,8,9$ 为来自总体 X 的样本观测值，试求 X 的经验分布函数.

解　将各个观测值从小到大排列，得 $2,4,4,6,8,8,8,9$，因此经验分布函数为

$$F_8(x)=\begin{cases}0, & x<2,\\ \dfrac{1}{8}, & 2\leqslant x<4,\\ \dfrac{3}{8}, & 4\leqslant x<6,\\ \dfrac{1}{2}, & 6\leqslant x<8,\\ \dfrac{7}{8}, & 8\leqslant x<9,\\ 1, & x\geqslant 9,\end{cases}$$

6.2.6　题型六：样本容量问题

例 6.2.12　设 $X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(\mu,4)$ 的一个样本，$\overline{X}$ 为样本均值，若已知 $P\{|\overline{X}-\mu|<0.5\}\geqslant 0.95$，试求最小的样本容量 n.

解　由于 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，其中 $\sigma=2$，因此

$$P\{|\overline{X}-\mu|<0.5\}=P\left(\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right|<\frac{0.5}{\sigma/\sqrt{n}}\right)=\Phi\left(\frac{0.5}{\sigma/\sqrt{n}}\right)-\Phi\left(-\frac{0.5}{\sigma/\sqrt{n}}\right)$$

$$=2\Phi\left(\frac{0.5}{\sigma/\sqrt{n}}\right)-1\geqslant 0.95.$$

于是 $\Phi\left(\dfrac{0.5}{\sigma/\sqrt{n}}\right)\geqslant 0.975$，经查表得 $\dfrac{0.5}{\sigma/\sqrt{n}}\geqslant 1.96$，即 $n\geqslant 61.46$，故最小的样本容量 n 应取 65.

6.3　深 化 训 练

6.3.1　填空题

（1）设 $X_1,X_2,\cdots,X_n$ 为来自正态分布 $N(\mu,0.3^2)$ 的一个样本，$\overline{X}$ 为样本均值，为使 $P\{|\overline{X}-\mu|<0.1\}\geqslant 0.95$，则样本容量 n 应至少为________.

（2）设 $X_1,X_2,\cdots,X_5$ 为来自总体 $X\sim N(1,4)$ 的样本，记 $Y=a(X_1-3X_2+b)^2+c(X_3+X_4-2X_5)^2$，则当 $a=$________，$b=$________，$c=$________时，Y 服从自由度为________的 χ^2 分布.

（3）【2001（3）】设 $X\sim N(0,2^2)$，$X_1,X_2,\cdots,X_{15}$ 为来自总体 X 的简单随机样本，则随机变量 $Y=\dfrac{X_1^2+X_2^2+\cdots+X_{10}^2}{2(X_{11}^2+X_{12}^2+\cdots+X_{15}^2)}$ 服从________分布，参数为________.

（4）【2006（3）】设总体 X 的概率密度函数为 $f(x)=\dfrac{1}{2}\mathrm{e}^{-|x|}$，$X_1,X_2,\cdots,X_n$ 是来自总体 X 的一个样本，样本方差为 S^2，则 $E(S^2)=$________.

6.3.2　单项选择题

（1）【2013（1）】设 $X\sim t(n)$，$Y\sim F(1,n)$，给定 α $(0<\alpha<0.5)$，常数 $C>0$ 满足 $P\{X>C\}=\alpha$，则 $P\{Y>C^2\}=$（　　）.

（A）α　　（B）$1-\alpha$　　（C）2α　　（D）$1-2\alpha$

（2）【2005（1）】设 $X_1,X_2,\cdots,X_n$ $(n\geqslant 2)$ 为来自 $N(0,1)$ 的简单随机样本，$\overline{X}$ 为样本均值，S^2 为样本方差，则下列结论正确的是（　　）

（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）$\dfrac{(n-1)X_1^2}{\sum\limits_{i=2}^{n}X_i^2}\sim F(1,n-1)$

（3）设 $X_1,X_2,\cdots,X_n$ $(n\geqslant 2)$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，对于任意的 $\varepsilon>0$，下列结论正确的是（　　）.

（A）$P\{|X-\mu|<\varepsilon\}=P\{|\overline{X}-\mu|<\varepsilon\}$　　（B）$P\{|X-\mu|<\varepsilon\}>P\{|\overline{X}-\mu|<\varepsilon\}$

（C）$P\{|X-\mu|<\varepsilon\}<P\{|\overline{X}-\mu|<\varepsilon\}$　　（D）以上结论都不对

（4）【2003（1）】设随机变量 $X\sim t(n)$，其中 $n>1$，若 $Y=\dfrac{1}{X^2}$，则下列结论正确的是（　　）.

（A）$Y\sim\chi^2(n)$　　（B）$Y\sim\chi^2(n-1)$　　（C）$Y\sim F(n,1)$　　（D）$Y\sim F(1,n)$

（5）若随机变量 $X_1\sim\chi^2(n_1)$，$X_2\sim\chi^2(n_2)$，且 X_1 和 X_2 相互独立，则下列结论正确的是（　　）.

（A）$D(X_1+X_2)=n_1+n_2$　　（B）$D(X_1+X_2)=2(n_1+n_2)$

（C）$D(X_1+X_2)=n_1^2+n_2^2$　　（D）$D(X_1+X_2)=2(n_1^2+n_2^2)$

（6）设 $X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(2,1)$ 的一个样本，且 $a\left(\sum\limits_{i=1}^{n}X_i-b\right)^2$ 服从自 $\chi^2(k)$ 分布，则 a，b 和 k 的值分别为（　　）.

（A）$a=\dfrac{1}{\sqrt{n}},b=2n,k=1$　　（B）$a=\dfrac{1}{n},b=2n,k=n$

（C）$a=\dfrac{1}{n},b=n,k=1$　　（D）$a=\dfrac{1}{n},b=2n,k=1$

6.3.3　设 $X_1,X_2,\cdots,X_{2n}$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，记

$$Y=\frac{X_1^2+X_3^2+\cdots+X_{2n-1}^2}{X_2^2+X_4^2+\cdots+X_{2n}^2},$$

试求统计量 Y 的抽样分布.

6.3.4　【1999】设 $X_1,X_2,\cdots,X_9$ 为来自总体 $N(\mu,\sigma^2)$ 的一个样本，记

$$Y_1=\frac{1}{6}(X_1+X_2+\cdots+X_6)，\quad Y_2=\frac{1}{3}(X_7+X_8+X_9)，\quad S^2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2，$$

试证明 $\dfrac{\sqrt{2}(Y_1-Y_2)}{S}\sim t(2)$.

6.3.5　设 X_1,X_2 为来自总体 $X\sim N(0,\sigma^2)$ 的一个样本，试求 $Y=\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2}$ 的分布.

6.3.6　设随机变量 $X\sim F(n,n)$，试证明 $P\{X<1\}=0.5$.

6.3.7　设 $X_1,X_2,\cdots,X_n$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，记 $Y=\dfrac{1}{n}\sum_{i=1}^{n}|X_i-\mu|$，试求 $E(Y)$ 和 $D(Y)$.

6.4　深化训练详解

6.3.1　填空题

（1）35；

（2）$\dfrac{1}{40}$，2，$\dfrac{1}{24}$，2；

（3）F，$(10,5)$；

（4）2；**提示**　由于样本方差 S^2 是总体方差的无偏估计，因此 $E(S^2)=D(X)$. 而

$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{2}\mathrm{e}^{-|x|}\mathrm{d}x=0，$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{2}\mathrm{e}^{-|x|}\mathrm{d}x=\int_{0}^{+\infty}x^2\cdot\mathrm{e}^{-x}\mathrm{d}x=\int_{0}^{+\infty}x^{3-1}\cdot\mathrm{e}^{-x}\mathrm{d}x=\Gamma(3)=2!=2，$$

因此

$$E(S^2)=D(X)=E(X^2)=2.$$

6.3.2　单项选择题

（1）C；**提示**　由 $X\sim t(n),Y\sim F(1,n)$，根据 t 分布与 F 的关系可知

$$P\{Y>C^2\}=P\{X^2>C^2\}=P\{X>C\}+P\{X<-C\}=2\alpha，$$

故选项 C 正确.

（2）D；**提示**　由于 $\overline{X}\sim N\left(0,\dfrac{1}{n}\right)$，因此 $n\overline{X}\sim N(0,n)$，故选项 A 错误；由于 $(n-1)S^2\sim\chi^2(n-1)$，因此选项 B 错误；由于

$$\frac{\sqrt{n}\cdot\overline{X}}{\sqrt{\dfrac{(n-1)S^2}{(n-1)}}}=\frac{\sqrt{n}\cdot\overline{X}}{S}\sim t(n-1)$$

因此选项 C 错误.

（3）C；**提示**

由于 $\frac{X-\mu}{\sigma} \sim N(0,1)$，$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，故有

$$P\{|X-\mu|<\varepsilon\}=P\left(\frac{|X-\mu|}{\sigma}<\frac{\varepsilon}{\sigma}\right)=2\Phi\left(\frac{\varepsilon}{\sigma}\right)-1,$$

$$P\{|\overline{X}-\mu|<\varepsilon\}=P\left(\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}}<\frac{\varepsilon}{\sigma/\sqrt{n}}\right)=2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)-1,$$

又因为标准正态分布的分布函数 $\Phi(x)$ 单调递增，$\frac{\varepsilon}{\sigma}<\frac{\sqrt{n}\varepsilon}{\sigma}$，因此

$$P\{|X-\mu|<\varepsilon\}<P\{|\overline{X}-\mu|<\varepsilon\},$$

故选项 C 正确.

（4）C；　　（5）B；　　（6）D.

6.3.3　由于 $X_1,X_2,\cdots,X_{2n}$ 相互独立，因此 $X_1^2+X_3^2+\cdots+X_{2n-1}^2$ 与 $X_2^2+X_4^2+\cdots+X_{2n}^2$ 相互独立．又因为

$$\frac{X_1^2+X_3^2+\cdots+X_{2n-1}^2}{\sigma^2} \sim \chi^2(n),\quad \frac{X_2^2+X_4^2+\cdots+X_{2n}^2}{\sigma^2} \sim \chi^2(n),$$

故

$$Y=\frac{X_1^2+X_3^2+\cdots+X_{2n-1}^2}{X_2^2+X_4^2+\cdots+X_{2n}^2}=\frac{(X_1^2+X_3^2+\cdots+X_{2n-1}^2)/(n\sigma^2)}{(X_2^2+X_4^2+\cdots+X_{2n}^2)/(n\sigma^2)} \sim F(n,n).$$

6.3.4　由于 $X_1,X_2,\cdots,X_9$ 相互独立且服从正态分布，因此 Y_1 和 Y_2 相互独立，且 Y_1-Y_2 服从正态分布. 而

$$E(Y_1-Y_2)=\mu-\mu=0,\quad D(Y_1-Y_2)=\frac{\sigma^2}{6}+\frac{\sigma^2}{3}=\frac{\sigma^2}{2},$$

因此 $Y_1-Y_2 \sim N\left(0,\frac{\sigma^2}{2}\right)$. 又因为 $\frac{(3-1)S^2}{\sigma^2}=\frac{2S^2}{\sigma^2} \sim \chi^2(2)$，且 S^2 与 Y_1-Y_2 相互独立，于是

$$\frac{\sqrt{2}(Y_1-Y_2)}{S}=\frac{\dfrac{Y_1-Y_2}{\sigma/\sqrt{2}}}{\sqrt{\dfrac{2S^2}{2\sigma^2}}} \sim t(2).$$

6.3.5　记 $U=X_1+X_2$，$V=X_1-X_2$，则

$$U \sim N(0,2\sigma^2),\quad V \sim N(0,2\sigma^2),$$

因此

$$\frac{U^2}{2\sigma^2} \sim \chi^2(1),\quad \frac{V^2}{2\sigma^2} \sim \chi^2(1).$$

又因为

$$\begin{aligned}\mathrm{Cov}(U,V)&=\mathrm{Cov}(X_1+X_2,X_1-X_2)\\&=D(X_1)+\mathrm{Cov}(X_2,X_1)-\mathrm{Cov}(X_1,X_2)-D(X_2)=0,\end{aligned}$$

因此统计量 U 和 V 相互独立，故根据 F 分布的定义，

$$Y=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}=\frac{U^2/(2\sigma^2)}{V^2/(2\sigma^2)}\sim F(1,1)\ .$$

6.3.6　记 $Y=\dfrac{1}{X}$，由于 $X\sim F(n,n)$，则 $Y\sim F(n,n)$，因此

$$P\{X<1\}=P\{Y<1\}=P\left\{\frac{1}{X}<1\right\}=P\{X>1\}\ .$$

又因为 X 为连续型随机变量，因此 $P\{X=1\}=0$，而

$$P\{X<1\}+P\{X=1\}+P\{X>1\}=1,$$

故 $P\{X<1\}=0.5$.

6.3.7　$E(Y)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}E(|X_i-\mu|)=E(|X-\mu|)=\sigma E\left(\dfrac{|X-\mu|}{\sigma}\right)$

$$=\sigma\int_{-\infty}^{+\infty}|t|\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}t=\sqrt{\frac{2}{\pi}}\sigma\int_0^{+\infty}t\mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}t$$

$$=-\sqrt{\frac{2}{\pi}}\sigma\int_0^{+\infty}\mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}\left(-\frac{1}{2}t^2\right)=\sqrt{\frac{2}{\pi}}\sigma\ .$$

$$D(Y)=\frac{1}{n^2}\sum_{i=1}^{n}D(|X_i-\mu|)=\frac{1}{n}D(|X-\mu|)=\frac{\sigma^2}{n}D\left(\frac{|X-\mu|}{\sigma}\right)$$

$$=\frac{\sigma^2}{n}\left\{E\left(\frac{X-\mu}{\sigma}\right)^2-\left[E\left(\frac{|X-\mu|}{\sigma}\right)\right]^2\right\}$$

$$=\frac{\sigma^2}{n}\left[D\left(\frac{X-\mu}{\sigma}\right)-\left(\sqrt{\frac{2}{\pi}}\right)^2\right]=\frac{\sigma^2}{n}\left(1-\frac{2}{\pi}\right).$$

6.5　综合提高训练

例 6.5.1　设总体 X 服从参数为 λ 的指数分布，X_1,X_2 为来自 X 的样本，记 $Y=\sqrt{X_1X_2}$，试求 $E(Y)$.

解　由题意 X 的概率密度为

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & \text{其他},\end{cases}$$

因此

$$E(\sqrt{X})=\int_0^{+\infty}\sqrt{x}\cdot\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x\overset{t=\lambda x}{=\!=\!=}\frac{1}{\sqrt{\lambda}}\int_0^{+\infty}\sqrt{t}\cdot\mathrm{e}^{-t}\mathrm{d}t=\frac{1}{\sqrt{\lambda}}\int_0^{+\infty}t^{\frac{3}{2}-1}\cdot\mathrm{e}^{-t}\mathrm{d}t$$

$$=\frac{1}{\sqrt{\lambda}}\Gamma\left(\frac{3}{2}\right)=\frac{1}{2\sqrt{\lambda}}\Gamma\left(\frac{1}{2}\right)=\frac{\sqrt{\pi}}{2\sqrt{\lambda}}\ .$$

因为 X_1 与 X_2 相互独立，故 $\sqrt{X_1}$ 与 $\sqrt{X_2}$ 相互独立，且有

$$E(\sqrt{X_1})=E(\sqrt{X_2})=E(\sqrt{X}),$$

故

$$E(Y)=E(\sqrt{X_1X_2})=E(\sqrt{X_1})\cdot E(\sqrt{X_2})=\frac{\sqrt{\pi}}{2\sqrt{\lambda}}\cdot\frac{\sqrt{\pi}}{2\sqrt{\lambda}}=\frac{\pi}{4\lambda}.$$

例 6.5.2　设 $X_1,X_2,\cdots,X_{2n}$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，样本均值 $\overline{X}=\dfrac{1}{2n}\sum\limits_{i=1}^{2n}X_i$，记 $\xi=\sum\limits_{i=1}^{n}(X_i+X_{n+i}-2\overline{X})^2$，试求 $E(\xi)$ 和 $D(\xi)$.

解　记 $Y_i=X_i+X_{n+i},\ i=1,2,\cdots,n$，由于 $X_1,X_2,\cdots,X_{2n}$ 相互独立且均服从 $N(\mu,\sigma^2)$，因此 $Y_1,Y_2,\cdots,Y_n$ 相互独立且均服从 $N(2\mu,2\sigma^2)$，因此 $Y_1,Y_2,\cdots,Y_n$ 可以理解为来自总体 $Y\sim N(2\mu,2\sigma^2)$ 的一个样本，样本均值为

$$\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i=\frac{1}{n}\sum_{i=1}^{n}(X_i+X_{n+i})=\frac{1}{n}\sum_{i=1}^{2n}X_i=2\overline{X}.$$

因此

$$\frac{\xi}{2\sigma^2}=\sum_{i=1}^{n}\frac{(X_i+X_{n+i}-2\overline{X})^2}{2\sigma^2}=\sum_{i=1}^{n}\frac{(Y_i-\overline{Y})^2}{2\sigma^2}\sim\chi^2(n-1),$$

根据 χ^2 分布的性质可知，

$$E\left(\frac{\xi}{2\sigma^2}\right)=n-1,\quad D\left(\frac{\xi}{2\sigma^2}\right)=2(n-1),$$

故

$$E(\xi)=2(n-1)\sigma^2,\quad D(\xi)=8(n-1)\sigma^4.$$

例 6.5.3　设 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个样本，$x_1,x_2,\cdots,x_n$ 为样本观测值，试证明当 $c=\overline{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$ 时，函数 $f(c)=\sum\limits_{i=1}^{n}(x_i-c)^2$ 取得最小值.

证法 1　由于

$$\begin{aligned}f(c)&=\sum_{i=1}^{n}(x_i-c)^2=\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-c)^2\\&=\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\overline{x}-c)^2+2\sum_{i=1}^{n}(x_i-\overline{x})(\overline{x}-c)\\&=\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\overline{x}-c)^2+2(\overline{x}-c)\cdot\sum_{i=1}^{n}(x_i-\overline{x})\\&=\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\overline{x}-c)^2+2(\overline{x}-c)\cdot(n\overline{x}-n\overline{x})\\&=\sum_{i=1}^{n}(x_i-\overline{x})^2+\sum_{i=1}^{n}(\overline{x}-c)^2,\end{aligned}$$

因此当 $c=\overline{x}$ 时，函数 $f(c)=\sum_{i=1}^{n}(x_i-c)^2$ 取得最小值.

证法 2　由题意，$f'(c)=-2\sum_{i=1}^{n}(x_i-c)$，令 $f'(c)=0$，解得 $c=\overline{x}$．又因为

$$f''(c)=-2\sum_{i=1}^{n}(x_i-c)'=2n>0,$$

故 $f''(\overline{x})>0$，因此函数 $f(c)=\sum_{i=1}^{n}(x_i-c)^2$ 在 $c=\overline{x}$ 处取得最小值.

例 6.5.4　设 X_1,X_2,X_3 为来自总体 $X\sim N(0,\sigma^2)$ 的一个样本，试证明 $Y=\sqrt{\dfrac{2}{3}}\cdot\dfrac{X_1+X_2+X_3}{|X_1-X_2|}$ 服从自由为1的 t 分布.

证　记 $U=X_1+X_2+X_3$，$V=X_1-X_2$，则

$$U\sim N(0,3\sigma^2),\quad V\sim N(0,2\sigma^2),\quad \frac{V^2}{2\sigma^2}\sim\chi^2(1).$$

又因为

$$\mathrm{Cov}(U,V)=\mathrm{Cov}(X_1+X_2+X_3,X_1-X_2)=0,$$

因此统计量 U 和 V 相互独立，故根据 t 分布的定义，

$$Y=\sqrt{\frac{2}{3}}\cdot\frac{U}{|V|}=\frac{U/(\sqrt{3}\sigma)}{\sqrt{V^2/2\sigma^2}}\sim t(1).$$

例 6.5.5　【2005（3）】设 $X_1,X_2\cdots,X_n$ $(n>2)$ 为来自总体 $X\sim N(0,\sigma^2)$ 的样本，$\overline{X}$ 为样本均值，记 $Y_i=X_i-\overline{X}$，$i=1,2,\cdots,n$，求：

（1）Y_i 的方差 $D(Y_i)$；　（2）$\mathrm{Cov}(Y_1,Y_n)$；　（3）常数 C，使得 $E[C(Y_1+Y_n)^2]=\sigma^2$.

解　（1）$D(Y_i)=D(X_i-\overline{X})=D\left[\left(1-\dfrac{1}{n}\right)X_i-\dfrac{1}{n}\sum\limits_{\substack{k=1\\k\neq i}}^{n}X_k\right]$

$$=\left(1-\frac{1}{n}\right)^2\sigma^2+\frac{n-1}{n^2}\sigma^2=\frac{n-1}{n}\sigma^2;$$

（2）
$$\begin{aligned}\mathrm{Cov}(Y_1,Y_n)&=\mathrm{Cov}(X_1-\overline{X},X_n-\overline{X})\\&=\mathrm{Cov}(X_1,X_n)-\mathrm{Cov}(X_1,\overline{X})-\mathrm{Cov}(\overline{X},X_n)+D(\overline{X})\\&=0-2\mathrm{Cov}(X_1,\overline{X})+\frac{\sigma^2}{n}\\&=-2\mathrm{Cov}\left(X_1,\frac{1}{n}X_1\right)+\frac{\sigma^2}{n}\\&=-\frac{2}{n}\sigma^2+\frac{\sigma^2}{n}=-\frac{1}{n}\sigma^2;\end{aligned}$$

（3）由于 $E(Y_1)=E(Y_n)=0$，因此

$$
\begin{aligned}
E[C(Y_1+Y_n)^2] &= CE[(Y_1+Y_n)^2] = CD(Y_1+Y_n) \\
&= C[D(Y_1)+D(Y_n)+2\mathrm{Cov}(Y_1,Y_n)\} \\
&= C\left[\frac{2(n-1)}{n}\sigma^2 - \frac{2}{n}\sigma^2\right] \\
&= C\cdot\frac{2n-4}{n}\sigma^2 = \sigma^2,
\end{aligned}
$$

故 $C=\dfrac{n}{2n-4}$.