习题 5.3

1． 在一本书上我们随机的检查了 10 页，发现每页上的错误数为：

4 5 6 0 3 1 4 2 1 4

试计算其样本均值、样本方差和样本标准差.

解：样本均值 $\bar{x}=\dfrac{1}{10}(4+5+6+\cdots+1+4)=3$；

样本方差 $s^2=\dfrac{1}{9}[(4-3)^2+(5-3)^2+(6-3)^2+\cdots+(1-3)^2+(4-3)^2]\approx 3.7778$；

样本标准差 $s=\sqrt{3.7778}\approx 1.9437$.

2． 证明：对任意常数 c, d，有 $\sum\limits_{i=1}^{n}(x_i-c)(y_i-d)=\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+n(\bar{x}-c)(\bar{y}-d)$.

证：$\sum\limits_{i=1}^{n}(x_i-c)(y_i-d)=\sum\limits_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-c)][(y_i-\bar{y})+(\bar{y}-d)]$

$$=\sum_{i=1}^{n}[(x_i-\bar{x})(y_i-\bar{y})+(\bar{x}-c)(y_i-\bar{y})+(x_i-\bar{x})(\bar{y}-d)+(\bar{x}-c)(\bar{y}-d)]$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+(\bar{x}-c)\sum_{i=1}^{n}(y_i-\bar{y})+(\bar{y}-d)\sum_{i=1}^{n}(x_i-\bar{x})+n(\bar{x}-c)(\bar{y}-d)$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+0+0+n(\bar{x}-c)(\bar{y}-d)=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+n(\bar{x}-c)(\bar{y}-d)\,.$$

3． 设 $x_1, \cdots, x_n$ 和 $y_1, \cdots, y_n$ 是两组样本观测值，且有如下关系：$y_i=3x_i-4$，$i=1, \cdots, n$，试求样本均值 $\bar{x}$ 和 $\bar{y}$ 间的关系以及样本方差 s_x^2 和 s_y^2 间的关系.

解：$\bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i=\dfrac{1}{n}\sum\limits_{i=1}^{n}(3x_i-4)=\dfrac{1}{n}\left(3\sum\limits_{i=1}^{n}x_i-4n\right)=\dfrac{3}{n}\sum\limits_{i=1}^{n}x_i-4=3\bar{x}-4$；

$$s_y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2=\frac{1}{n-1}\sum_{i=1}^{n}[(3x_i-4)-(3\bar{x}-4)]^2=\frac{9}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=9s_x^2\,.$$

4． 记 $\bar{x}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$， $s_n^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\bar{x}_n)^2$，$n=1, 2, \cdots$，证明

$$\bar{x}_{n+1}=\bar{x}_n+\frac{1}{n+1}(x_{n+1}-\bar{x}_n)\,,\quad s_{n+1}^2=\frac{n-1}{n}s_n^2+\frac{1}{n+1}(x_{n+1}-\bar{x}_n)^2\,.$$

证：$\bar{x}_{n+1}=\dfrac{1}{n+1}\sum\limits_{i=1}^{n+1}x_i=\dfrac{n}{n+1}\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i+\dfrac{1}{n+1}x_{n+1}=\dfrac{n}{n+1}\bar{x}_n+\dfrac{1}{n+1}x_{n+1}=\bar{x}_n+\dfrac{1}{n+1}(x_{n+1}-\bar{x}_n)$；

$$s_{n+1}^2=\frac{1}{n}\sum_{i=1}^{n+1}(x_i-\bar{x}_{n+1})^2=\frac{1}{n}\sum_{i=1}^{n+1}[(x_i-\bar{x}_n)-(\bar{x}_{n+1}-\bar{x}_n)]^2=\frac{1}{n}\left[\sum_{i=1}^{n+1}(x_i-\bar{x}_n)^2-(n+1)(\bar{x}_n-\bar{x}_{n+1})^2\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\overline{x}_n)^2+(x_{n+1}-\overline{x}_n)^2-(n+1)\cdot\frac{1}{(n+1)^2}(x_{n+1}-\overline{x}_n)^2\right]$$

$$=\frac{1}{n}\left[(n-1)\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_n)^2+\frac{n}{n+1}(x_{n+1}-\overline{x}_n)^2\right]=\frac{n-1}{n}s_n^2+\frac{1}{n+1}(x_{n+1}-\overline{x}_n)^2 .$$

5. 从同一总体中抽取两个容量分别为 n, m 的样本，样本均值分别为 $\overline{x}_1$，$\overline{x}_2$，样本方差分别为 s_1^2，s_2^2，将两组样本合并，其均值、方差分别为 $\overline{x}$, s^2，证明：

$$\overline{x}=\frac{n\overline{x}_1+m\overline{x}_2}{n+m}，\quad s^2=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-1}+\frac{nm(\overline{x}_1-\overline{x}_2)^2}{(n+m)(n+m-1)} .$$

证：$\overline{x}=\dfrac{1}{n+m}\left(\sum\limits_{i=1}^{n}x_{1i}+\sum\limits_{j=1}^{m}x_{2j}\right)=\dfrac{n\overline{x}_1+m\overline{x}_2}{n+m}$；

$$s^2=\frac{1}{n+m-1}\left[\sum_{i=1}^{n}(x_{1i}-\overline{x})^2+\sum_{j=1}^{m}(x_{2j}-\overline{x})^2\right]$$

$$=\frac{1}{n+m-1}\left[\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2+n(\overline{x}_1-\overline{x})^2+\sum_{j=1}^{m}(x_{2j}-\overline{x}_2)^2+m(\overline{x}_2-\overline{x})^2\right]$$

$$=\frac{1}{n+m-1}\left[(n-1)s_1^2+n\left(\overline{x}_1-\frac{n\overline{x}_1+m\overline{x}_2}{n+m}\right)^2+(m-1)s_2^2+m\left(\overline{x}_2-\frac{n\overline{x}_1+m\overline{x}_2}{n+m}\right)^2\right]$$

$$=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-1}+\frac{nm^2(\overline{x}_1-\overline{x}_2)^2+mn^2(\overline{x}_2-\overline{x}_1)^2}{(n+m-1)(n+m)^2}=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-1}+\frac{nm(\overline{x}_1-\overline{x}_2)^2}{(n+m)(n+m-1)} .$$

6. 设有容量为 n 的样本 A，它的样本均值为 $\overline{x}_A$，样本标准差为 s_A，样本极差为 R_A，样本中位数为 m_A．现对样本中每一个观测值施行如下变换：$y=ax+b$，如此得到样本 B，试写出样本 B 的均值、标准差、极差和中位数.

解：$\overline{y}_B=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i=\dfrac{1}{n}\sum\limits_{i=1}^{n}(ax_i+b)=\dfrac{1}{n}(a\sum\limits_{i=1}^{n}x_i+nb)=a\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i+b=a\overline{x}_A+b$；

$$s_B=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y}_B)^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(ax_i+b-a\overline{x}_A-b)^2}=|a|\cdot\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_A)^2}=|a|s_A；$$

$R_B=y_{(n)}-y_{(1)}=a\,x_{(n)}+b-a\,x_{(1)}-b=a\,[x_{(n)}-x_{(1)}]=a\,R_A$；

当 n 为奇数时，$m_{B0.5}=y_{\left(\frac{n+1}{2}\right)}=ax_{\left(\frac{n+1}{2}\right)}+b=am_{A0.5}+b$，

当 n 为偶数时，$m_{B0.5}=\dfrac{1}{2}[y_{\left(\frac{n}{2}\right)}+y_{\left(\frac{n}{2}+1\right)}]=\dfrac{1}{2}[ax_{\left(\frac{n}{2}\right)}+b+ax_{\left(\frac{n}{2}+1\right)}+b]=\dfrac{a}{2}[x_{\left(\frac{n}{2}\right)}+x_{\left(\frac{n}{2}+1\right)}]+b=am_{A0.5}+b$，

故 $m_{B0.5}=a\,m_{A0.5}+b$.

7. 证明：容量为 2 的样本 x_1,x_2 的方差为 $s^2=\dfrac{1}{2}(x_1-x_2)^2$.

证：$s^2=\dfrac{1}{2-1}\left[\left(x_1-\dfrac{x_1+x_2}{2}\right)^2+\left(x_2-\dfrac{x_1+x_2}{2}\right)^2\right]=\dfrac{(x_1-x_2)^2}{4}+\dfrac{(x_2-x_1)^2}{4}=\dfrac{1}{2}(x_1-x_2)^2$.

8．设 $x_1,\cdots,x_n$ 是来自 $U(-1,1)$ 的样本，试求 $E(\overline{X})$ 和 $\mathrm{Var}(\overline{X})$.

解：因 $X_i\sim U(-1,1)$，有 $E(X_i)=\dfrac{-1+1}{2}=0$，$\mathrm{Var}(X_i)=\dfrac{(1+1)^2}{12}=\dfrac{1}{3}$，

故 $E(\overline{X})=E\left(\dfrac{1}{n}\sum_{i=1}^n X_i\right)=\dfrac{1}{n}\sum_{i=1}^n E(X_i)=0$，$\mathrm{Var}(\overline{X})=\mathrm{Var}\left(\dfrac{1}{n}\sum_{i=1}^n X_i\right)=\dfrac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i)=\dfrac{1}{n^2}\cdot n\cdot\dfrac{1}{3}=\dfrac{1}{3n}$.

9．设总体二阶矩存在，$X_1,\cdots,X_n$ 是样本，证明 $X_i-\overline{X}$ 与 $X_j-\overline{X}$ $(i\ne j)$ 的相关系数为 $-(n-1)^{-1}$.

证：因 $X_1,X_2,\cdots,X_n$ 相互独立，有 $\mathrm{Cov}(X_l,X_k)=0$，$(l\ne k)$，

则 $\mathrm{Cov}(X_i-\overline{X},X_j-\overline{X})=\mathrm{Cov}(X_i,X_j)-\mathrm{Cov}(X_i,\overline{X})-\mathrm{Cov}(\overline{X},X_j)+\mathrm{Cov}(\overline{X},\overline{X})$

$$=0-\mathrm{Cov}(X_i,\frac{1}{n}X_i)-\mathrm{Cov}(\frac{1}{n}X_j,X_j)+\mathrm{Var}(\overline{X})$$

$$=-\frac{1}{n}\mathrm{Var}(X_i)-\frac{1}{n}\mathrm{Var}(X_j)+\mathrm{Var}(\overline{X})=-\frac{1}{n}\sigma^2-\frac{1}{n}\sigma^2+\frac{1}{n}\sigma^2=-\frac{1}{n}\sigma^2,$$

且 $\mathrm{Var}(X_i-\overline{X})=\mathrm{Var}(X_i)+\mathrm{Var}(\overline{X})-2\mathrm{Cov}(X_i,\overline{X})=\sigma^2+\dfrac{1}{n}\sigma^2-2\mathrm{Cov}(X_i,\dfrac{1}{n}X_i)$

$$=\sigma^2+\frac{1}{n}\sigma^2-\frac{2}{n}\sigma^2=\frac{n-1}{n}\sigma^2=\mathrm{Var}(X_j-\overline{X}),$$

故 $\mathrm{Corr}(X_i-\overline{X},X_j-\overline{X})=\dfrac{\mathrm{Cov}(X_i-\overline{X},X_j-\overline{X})}{\sqrt{\mathrm{Var}(X_i-\overline{X})}\cdot\sqrt{\mathrm{Var}(X_j-\overline{X})}}=\dfrac{-\frac{1}{n}\sigma^2}{\sqrt{\frac{n-1}{n}\sigma^2}\cdot\sqrt{\frac{n-1}{n}\sigma^2}}=-\dfrac{1}{n-1}$.

10．设 $x_1,x_2,\cdots,x_n$ 为一个样本，$s^2=\dfrac{1}{n-1}\sum_{i=1}^n(x_i-\bar{x})^2$ 是样本方差，试证：

$$\frac{1}{n(n-1)}\sum_{i<j}(x_i-x_j)^2=s^2.$$

证：因 $s^2=\dfrac{1}{n-1}\sum_{i=1}^n(x_i-\bar{x})^2=\dfrac{1}{n-1}\left(\sum_{i=1}^n x_i^2-n\bar{x}^2\right)$，

则 $\sum_{i<j}(x_i-x_j)^2=\dfrac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(x_i-x_j)^2=\dfrac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(x_i^2+x_j^2-2x_ix_j)=\dfrac{1}{2}\left(\sum_{i=1}^n\sum_{j=1}^n x_i^2+\sum_{i=1}^n\sum_{j=1}^n x_j^2-2\sum_{i=1}^n\sum_{j=1}^n x_ix_j\right)$

$$=\frac{1}{2}\left(n\sum_{i=1}^n x_i^2+n\sum_{j=1}^n x_j^2-2\sum_{i=1}^n x_i\sum_{j=1}^n x_j\right)=\frac{1}{2}\left(2n\sum_{i=1}^n x_i^2-2n\bar{x}\cdot n\bar{x}\right)=n\left(\sum_{i=1}^n x_i^2-n\bar{x}^2\right)=n(n-1)s^2,$$

故 $\dfrac{1}{n(n-1)}\sum_{i<j}(x_i-x_j)^2=s^2$.

11．设总体 4 阶中心矩 $\nu_4 = E[X-E(X)]^4$ 存在，试对样本方差 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，有

$$\operatorname{Var}(S^2) = \frac{n(\nu_4-\sigma^4)}{(n-1)^2} - \frac{2(\nu_4-2\sigma^4)}{(n-1)^2} + \frac{\nu_4-3\sigma^4}{n(n-1)^2},$$

其中 σ^2 为总体 X 的方差.

证：因 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}[(X_i-\mu)-(\bar{X}-\mu)]^2 = \dfrac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right]$，其中 $\mu = E(X)$，

则 $\operatorname{Var}(S^2) = \dfrac{1}{(n-1)^2}\operatorname{Var}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right]$

$$= \frac{1}{(n-1)^2}\left\{\operatorname{Var}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] - 2\operatorname{Cov}\left(\sum_{i=1}^{n}(X_i-\mu)^2, n(\bar{X}-\mu)^2\right) + \operatorname{Var}[n(\bar{X}-\mu)^2]\right\}$$

$$= \frac{1}{(n-1)^2}\left\{\sum_{i=1}^{n}\operatorname{Var}(X_i-\mu)^2 - 2n\sum_{i=1}^{n}\operatorname{Cov}((X_i-\mu)^2, (\bar{X}-\mu)^2) + n^2\operatorname{Var}(\bar{X}-\mu)^2\right\},$$

因 $E(X_i-\mu)^2 = \sigma^2$，$E(X_i-\mu)^4 = \nu_4$，

则 $\sum_{i=1}^{n}\operatorname{Var}(X_i-\mu)^2 = \sum_{i=1}^{n}\{E(X_i-\mu)^4 - [E(X_i-\mu)^2]^2\} = \sum_{i=1}^{n}\{\nu_4 - (\sigma^2)^2\} = n(\nu_4-\sigma^4)$，

因 $E(X_i-\mu) = 0$，$E(\bar{X}-\mu)^2 = \operatorname{Var}(\bar{X}) = \dfrac{1}{n}\sigma^2$，且当 $i \ne j$ 时，$X_i-\mu$ 与 $X_j-\mu$ 相互独立，

则 $\sum_{i=1}^{n}\operatorname{Cov}((X_i-\mu)^2, (\bar{X}-\mu)^2) = \sum_{i=1}^{n}\{E[(X_i-\mu)^2(\bar{X}-\mu)^2] - E(X_i-\mu)^2 E(\bar{X}-\mu)^2\}$

$$= \sum_{i=1}^{n}\left\{E\left[(X_i-\mu)^2\cdot\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)\right)^2\right] - \sigma^2\cdot\frac{1}{n}\sigma^2\right\}$$

$$= \sum_{i=1}^{n}\left\{\frac{1}{n^2}\left[E(X_i-\mu)^4 + E(X_i-\mu)^2\cdot\sum_{k\ne i}E(X_k-\mu)^2\right] - \frac{1}{n}\sigma^4\right\}$$

$$= \sum_{i=1}^{n}\left\{\frac{1}{n^2}[\nu_4 + \sigma^2\cdot(n-1)\sigma^2] - \frac{1}{n}\sigma^4\right\} = \frac{1}{n}(\nu_4-\sigma^4),$$

且 $\operatorname{Var}(\bar{X}-\mu)^2 = E(\bar{X}-\mu)^4 - [E(\bar{X}-\mu)^2]^2 = E\left[\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right]^4 - \left[\dfrac{1}{n}\sigma^2\right]^2$

$$= \frac{1}{n^4}E\left[\sum_{i=1}^{n}(X_i-\mu)^4 + \binom{4}{2}\sum_{i<j}(X_i-\mu)^2(X_j-\mu)^2\right] - \frac{1}{n^2}\sigma^4$$

$$= \frac{1}{n^4}\left[\sum_{i=1}^{n}E(X_i-\mu)^4 + 6\sum_{i<j}E(X_i-\mu)^2E(X_j-\mu)^2\right] - \frac{1}{n^2}\sigma^4$$

$$=\frac{1}{n^4}\left[n\nu_4+6\cdot\binom{n}{2}\sigma^2\cdot\sigma^2\right]-\frac{1}{n^2}\sigma^4=\frac{1}{n^4}[n\nu_4+3n(n-1)\sigma^4]-\frac{1}{n^2}\sigma^4=\frac{1}{n^3}(\nu_4-3\sigma^4)+\frac{2}{n^2}\sigma^4,$$

故 $\mathrm{Var}(S^2)=\dfrac{1}{(n-1)^2}\left\{n(\nu_4-\sigma^4)-2n\cdot\dfrac{1}{n}(\nu_4-\sigma^4)+n^2\left[\dfrac{1}{n^3}(\nu_4-3\sigma^4)+\dfrac{2}{n^2}\sigma^4\right]\right\}$

$$=\frac{1}{(n-1)^2}\left\{n(\nu_4-\sigma^4)-2(\nu_4-\sigma^4)+\frac{1}{n}(\nu_4-3\sigma^4)+2\sigma^4\right\}$$

$$=\frac{1}{(n-1)^2}\left\{n(\nu_4-\sigma^4)-2(\nu_4-2\sigma^4)+\frac{1}{n}(\nu_4-3\sigma^4)\right\}=\frac{n(\nu_4-\sigma^4)}{(n-1)^2}-\frac{2(\nu_4-2\sigma^4)}{(n-1)^2}+\frac{\nu_4-3\sigma^4}{n(n-1)^2}.$$

12．设总体 X 的 3 阶矩存在，设 $X_1, X_2, \cdots, X_n$ 是取自该总体的简单随机样本，$\overline{X}$ 为样本均值，S^2 为样本方差，试证：$\mathrm{Cov}(\overline{X}, S^2)=\dfrac{\nu_3}{n}$，其中 $\nu_3=E[X-E(X)]^3$.

证：因 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2=\dfrac{1}{n-1}\left[\sum\limits_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]$，其中 $\mu=E(X)$，

则 $\mathrm{Cov}(\overline{X}, S^2)=\mathrm{Cov}(\overline{X}-\mu, S^2)=\mathrm{Cov}\left(\overline{X}-\mu, \dfrac{1}{n-1}\left[\sum\limits_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]\right)$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\mathrm{Cov}(\overline{X}-\mu,(X_i-\mu)^2)-n\,\mathrm{Cov}(\overline{X}-\mu,(\overline{X}-\mu)^2)\right],$$

因 $E(\overline{X}-\mu)=E(X_i-\mu)=0$，$E(X_i-\mu)^2=\sigma^2$，$E(X_i-\mu)^3=\nu_3$，且当 $i\neq j$ 时，$X_i-\mu$ 与 $X_j-\mu$ 相互独立，

则 $\sum\limits_{i=1}^{n}\mathrm{Cov}(\overline{X}-\mu,(X_i-\mu)^2)=\sum\limits_{i=1}^{n}\mathrm{Cov}\left(\dfrac{1}{n}\sum\limits_{k=1}^{n}(X_k-\mu),(X_i-\mu)^2\right)=\dfrac{1}{n}\sum\limits_{i=1}^{n}\mathrm{Cov}(X_i-\mu,(X_i-\mu)^2)$

$$=\frac{1}{n}\sum_{i=1}^{n}[E(X_i-\mu)^3-E(X_i-\mu)E(X_i-\mu)^2]=\frac{1}{n}\cdot n\nu_3=\nu_3,$$

且 $\mathrm{Cov}(\overline{X}-\mu,(\overline{X}-\mu)^2)=E(\overline{X}-\mu)^3-E(\overline{X}-\mu)E(\overline{X}-\mu)^2=E\left[\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)\right]^3$

$$=\frac{1}{n^3}E\left[\sum_{i=1}^{n}(X_i-\mu)^3\right]=\frac{1}{n^3}\sum_{i=1}^{n}E(X_i-\mu)^3=\frac{1}{n^3}\cdot n\nu_3=\frac{1}{n^2}\nu_3,$$

故 $\mathrm{Cov}(\overline{X}, S^2)=\dfrac{1}{n-1}\left(\nu_3-n\cdot\dfrac{1}{n^2}\nu_3\right)=\dfrac{1}{n-1}\cdot\dfrac{n-1}{n}\nu_3=\dfrac{\nu_3}{n}$.

13．设 $\overline{X}_1$ 与 $\overline{X}_2$ 是从同一正态总体 $N(\mu, \sigma^2)$独立抽取的容量相同的两个样本均值．试确定样本容量 n，使得两样本均值的距离超过 σ 的概率不超过 0.01．

解：因 $E(\overline{X}_1)=E(\overline{X}_2)=\mu$，$\mathrm{Var}(\overline{X}_1)=\mathrm{Var}(\overline{X}_2)=\dfrac{\sigma^2}{n}$，$\overline{X}_1$ 与 $\overline{X}_2$ 相互独立，且总体分布为 $N(\mu, \sigma^2)$，

则 $E(\overline{X}_1-\overline{X}_2)=\mu-\mu=0$， $\text{Var}(\overline{X}_1-\overline{X}_2)=\dfrac{\sigma^2}{n}+\dfrac{\sigma^2}{n}=\dfrac{2\sigma^2}{n}$， 即 $\overline{X}_1-\overline{X}_2\sim N\left(0,\dfrac{2\sigma^2}{n}\right)$，

因 $P\{|\overline{X}_1-\overline{X}_2|>\sigma\}=2\left[1-\Phi\left(\dfrac{\sigma}{\sigma\sqrt{2/n}}\right)\right]=2-2\Phi\left(\sqrt{\dfrac{n}{2}}\right)\le 0.01$， 有 $\Phi\left(\sqrt{\dfrac{n}{2}}\right)\ge 0.995$， $\sqrt{\dfrac{n}{2}}\ge 2.5758$，

故 $n\ge 13.2698$，即 n 至少 14 个.

14. 利用切比雪夫不等式求抛均匀硬币多少次才能使正面朝上的频率落在 (0.4, 0.6) 间的概率至少为 0.9．如何才能更精确的计算这个次数？是多少？

解：设 $X_i=\begin{cases}1, & 第\ i\ 次正面朝上,\\ 0, & 第\ i\ 次反面朝上,\end{cases}$ 有 $X_i\sim B(1, 0.5)$，且正面朝上的频率为 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$，

则 $E(X_i)=0.5$，$\text{Var}(X_i)=0.25$，且 $E(\overline{X})=0.5$， $\text{Var}(\overline{X})=\dfrac{0.25}{n}$，

由切比雪夫不等式得 $P\{0.4<\overline{X}<0.6\}=P\{|\overline{X}-0.5|<0.1\}\ge 1-\dfrac{0.25}{0.1^2n}=1-\dfrac{25}{n}$，

故当 $1-\dfrac{25}{n}\ge 0.9$ 时，即 $n\ge 250$ 时， $P\{0.4<\overline{X}<0.6\}\ge 0.9$；

利用中心极限定理更精确地计算，当 n 很大时 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 的渐近分布为正态分布 $N(0.5,\dfrac{0.25}{n})$，

则 $P\{0.4<\overline{X}<0.6\}=F(0.6)-F(0.4)=\Phi(\dfrac{0.6-0.5}{\sqrt{\dfrac{0.25}{n}}})-\Phi(\dfrac{0.4-0.5}{\sqrt{\dfrac{0.25}{n}}})=\Phi(0.2\sqrt{n})-\Phi(-0.2\sqrt{n})$

$=2\Phi(0.2\sqrt{n})-1\ge 0.9$，

即 $\Phi(0.2\sqrt{n})\ge 0.95$， $0.2\sqrt{n}\ge 1.64$，

故当 $n\ge 67.24$ 时，即 $n\ge 68$ 时， $P\{0.4<\overline{X}<0.6\}\ge 0.9$.

15. 从指数总体 $Exp(1/\theta)$ 抽取了 40 个样品，试求 $\overline{X}$ 的渐近分布.

解：因 $E(\overline{X})=E(X)=\theta$， $\text{Var}(\overline{X})=\dfrac{\text{Var}(X)}{n}=\dfrac{1}{40}\theta^2$， 故 $\overline{X}$ 的渐近分布为 $N(\theta,\dfrac{1}{40}\theta^2)$.

16. 设 $X_1,\cdots,X_{25}$ 是从均匀分布 $U(0, 5)$ 抽取的样本，试求样本均值 $\overline{X}$ 的渐近分布.

解：因 $E(\overline{X})=E(X)=\dfrac{5}{2}$， $\text{Var}(\overline{X})=\dfrac{\text{Var}(X)}{n}=\dfrac{(5-0)^2}{25\times 12}=\dfrac{1}{12}$， 故 $\overline{X}$ 的渐近分布为 $N(\dfrac{5}{2},\dfrac{1}{12})$.

17. 设 $X_1,\cdots,X_{20}$ 是从二点分布 $b(1,p)$ 抽取的样本，试求样本均值 $\overline{X}$ 的渐近分布.

解：因 $E(\overline{X})=E(X)=p$， $\text{Var}(\overline{X})=\dfrac{\text{Var}(X)}{n}=\dfrac{p(1-p)}{20}$， 故 $\overline{X}$ 的渐近分布为 $N(p,\dfrac{p(1-p)}{20})$.

18. 设 $X_1,\cdots,X_8$ 是从正态分布 $N(10, 9)$ 中抽取的样本，试求样本均值 $\overline{X}$ 的标准差.

解：因 $\text{Var}(\overline{X})=\dfrac{\text{Var}(X)}{n}=\dfrac{9}{8}$， 故 $\overline{X}$ 的标准差为 $\sqrt{\text{Var}(\overline{X})}=\dfrac{3\sqrt{2}}{4}$.

19. 切尾均值也是一个常用的反映样本数据的特征量，其想法是将数据的两端的值舍去，而用剩下的当中的值为计算样本均值，其计算公式是

$$\bar{X}_\alpha = \frac{X_{([n\alpha]+1)} + X_{([n\alpha]+2)} + \cdots + X_{(n-[n\alpha])}}{n-2[n\alpha]},$$

其中 $0 < \alpha < 1/2$ 是切尾系数，$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ 是有序样本．现我们在高校采访了 16 名大学生，了解他们平时的学习情况，以下数据是大学生每周用于看电视的时间：

15 14 12 9 20 4 17 26 15 18 6 10 16 15 5 8

取 $\alpha = 1/16$，试计算其切尾均值．

解：因 $n\alpha = 1$，且有序样本为 4, 5, 6, 8, 9, 10, 12, 14, 15, 15, 15, 16, 17, 18, 20, 26，

故切尾均值 $\bar{x}_{1/16} = \dfrac{1}{16-2}(5+6+8+\cdots+20) = 12.8571$.

20．有一个分组样本如下：

区间	组中值	频数
(145,155)	150	4
(155,165)	160	8
(165,175)	170	6
(175,185)	180	2

试求该分组样本的样本均值、样本标准差、样本偏度和样本峰度．

解：$\bar{x} = \dfrac{1}{20}(150\times4 + 160\times8 + 170\times6 + 180\times2) = 163$；

$$s = \sqrt{\frac{1}{19}[(150-163)^2\times4 + (160-163)^2\times8 + (170-163)^2\times6 + (180-163)^2\times2]} = 9.2338;$$

因 $b_2 = \dfrac{1}{20}[(150-163)^2\times4 + (160-163)^2\times8 + (170-163)^2\times6 + (180-163)^2\times2] = 81$，

$b_3 = \dfrac{1}{20}[(150-163)^3\times4 + (160-163)^3\times8 + (170-163)^3\times6 + (180-163)^3\times2] = 144$，

$b_4 = \dfrac{1}{20}[(150-163)^4\times4 + (160-163)^4\times8 + (170-163)^4\times6 + (180-163)^4\times2] = 14817$，

故样本偏度 $\gamma_1 = \dfrac{b_3}{b_2^{3/2}} = 0.1975$，样本峰度 $\gamma_2 = \dfrac{b_4}{b_2^2} - 3 = -0.7417$.

21．检查四批产品，其批次与不合格品率如下：

批号	批量	不合格品率
1	100	0.05
2	300	0.06
3	250	0.04
4	150	0.03

试求这四批产品的总不合格品率．

解：$\bar{p} = \dfrac{1}{800}(100\times0.05 + 300\times0.06 + 250\times0.04 + 150\times0.03) = 0.046875$.

22．设总体以等概率取 1, 2, 3, 4, 5，现从中抽取一个容量为 4 的样本，试分别求 $X_{(1)}$ 和 $X_{(4)}$ 的分布．

解：因总体分布函数为

$$F(x)=\begin{cases}0, & x<1,\\ \frac{1}{5}, & 1\le x<2,\\ \frac{2}{5}, & 2\le x<3,\\ \frac{3}{5}, & 3\le x<4,\\ \frac{4}{5}, & 4\le x<5,\\ 1, & x\ge 5,\end{cases}$$

则 $F_{(1)}(x)=P\{X_{(1)}\le x\}=1-P\{X_{(1)}>x\}=1-P\{X_1>x, X_2>x, X_3>x, X_4>x\}=1-[1-F(x)]^4$

$$=\begin{cases}0, & x<1,\\ \frac{369}{625}, & 1\le x<2,\\ \frac{544}{625}, & 2\le x<3,\\ \frac{609}{625}, & 3\le x<4,\\ \frac{624}{625}, & 4\le x<5,\\ 1, & x\ge 5,\end{cases}$$

且 $F_{(4)}(x)=P\{X_{(4)}\le x\}=P\{X_1\le x, X_2\le x, X_3\le x, X_4\le x\}=[F(x)]^4$

$$=\begin{cases}0, & x<1,\\ \frac{1}{625}, & 1\le x<2,\\ \frac{16}{625}, & 2\le x<3,\\ \frac{81}{625}, & 3\le x<4,\\ \frac{256}{625}, & 4\le x<5,\\ 1, & x\ge 5,\end{cases}$$

故 $X_{(1)}$ 和 $X_{(4)}$ 的分布为

$X_{(1)}$	1	2	3	4	5
P	$\frac{369}{625}$	$\frac{175}{625}$	$\frac{65}{625}$	$\frac{15}{625}$	$\frac{1}{625}$

；

$X_{(4)}$	1	2	3	4	5
P	$\frac{1}{625}$	$\frac{15}{625}$	$\frac{65}{625}$	$\frac{175}{625}$	$\frac{369}{625}$

.

23. 设总体 X 服从几何分布，即 $P\{X=k\}=pq^{k-1}$，$k=1, 2, \cdots$，其中 $0<p<1$，$q=1-p$，$X_1, X_2, \cdots, X_n$ 为该总体的样本．求 $X_{(n)}, X_{(1)}$的概率分布．

解：因 $P\{X\le k\}=\sum_{j=1}^{k}pq^{j-1}=\frac{p(1-q^k)}{1-q}=1-q^k$，$k=1, 2, \cdots$,

故 $P\{X_{(n)}=k\}=P\{X_{(n)}\le k\}-P\{X_{(n)}\le k-1\}=\prod_{i=1}^{n}P\{X_i\le k\}-\prod_{i=1}^{n}P\{X_i\le k-1\}=(1-q^k)^n-(1-q^{k-1})^n$；

且 $P\{X_{(1)}=k\}=P\{X_{(1)}>k-1\}-P\{X_{(1)}>k\}=\prod_{i=1}^{n}P\{X_i>k-1\}-\prod_{i=1}^{n}P\{X_i>k\}=q^{n(k-1)}-q^{nk}$.

24. 设 $X_1, \cdots, X_{16}$ 是来自 $N(8, 4)$ 的样本，试求下列概率

（1）$P\{X_{(16)}>10\}$；

（2）$P\{X_{(1)} > 5\}$.

解：（1）$P\{X_{(16)} > 10\} = 1 - P\{X_{(16)} \le 10\} = 1 - \prod_{i=1}^{16} P\{X_i \le 10\} = 1 - [F(10)]^{16} = 1 - [\Phi(\frac{10-8}{2})]^{16}$

$= 1 - [\Phi(1)]^{16} = 1 - 0.8413^{16} = 0.9370$；

（2）$P\{X_{(1)} > 5\} = \prod_{i=1}^{16} P\{X_i > 5\} = [1 - F(5)]^{16} = [1 - \Phi(\frac{5-8}{2})]^{16} = [\Phi(1.5)]^{16} = 0.9332^{16} = 0.3308$.

25．设总体为韦布尔分布，其密度函数为

$$p(x; m, \eta) = \frac{mx^{m-1}}{\eta^m} \exp\left\{-\left(\frac{x}{\eta}\right)^m\right\},\ x > 0,\ m > 0,\ \eta > 0.$$

现从中得到样本 $X_1, \cdots, X_n$，证明 $X_{(1)}$ 仍服从韦布尔分布，并指出其参数.

解：总体分布函数 $F(x) = \int_0^x p(t)\mathrm{d}t = \int_0^x \frac{mt^{m-1}}{\eta^m} \mathrm{e}^{-\left(\frac{t}{\eta}\right)^m} \mathrm{d}t = \int_0^x \mathrm{e}^{-\left(\frac{t}{\eta}\right)^m} \mathrm{d}\left(\frac{t}{\eta}\right)^m = -\mathrm{e}^{-\left(\frac{t}{\eta}\right)^m}\Big|_0^x = 1 - \mathrm{e}^{-\left(\frac{x}{\eta}\right)^m}$，$x > 0$，

则 $X_{(1)}$ 的密度函数为

$$p_1(x) = n[1 - F(x)]^{n-1} p(x) = n\mathrm{e}^{-(n-1)\left(\frac{x}{\eta}\right)^m} \cdot \frac{mx^{m-1}}{\eta^m} \mathrm{e}^{-\left(\frac{x}{\eta}\right)^m} = \frac{mnx^{m-1}}{\eta^m} \mathrm{e}^{-n\left(\frac{x}{\eta}\right)^m} = \frac{mx^{m-1}}{(\eta/\sqrt[m]{n})^m} \mathrm{e}^{-\left(\frac{x}{\eta/\sqrt[m]{n}}\right)^m},$$

故 $X_{(1)}$ 服从参数为 $\left(m, \frac{\eta}{\sqrt[m]{n}}\right)$ 的韦布尔分布.

26．设总体密度函数为 $p(x) = 6x(1-x),\ 0 < x < 1, X_1, \cdots, X_9$ 是来自该总体的样本，试求样本中位数的分布.

解：总体分布函数 $F(x) = \int_0^x p(t)\mathrm{d}t = \int_0^x 6t(1-t)\mathrm{d}t = (3t^2 - 2t^3)\Big|_0^x = 3x^2 - 2x^3$，$0 < x < 1$，

因样本容量 $n = 9$，有样本中位数 $m_{0.5} = x_{\left(\frac{n+1}{2}\right)} = x_{(5)}$，其密度函数为

$$p_5(x) = \frac{9!}{4!\cdot 4!}[F(x)]^4[1 - F(x)]^4 p(x) = \frac{9!}{4!\cdot 4!}(3x^2 - 2x^3)^4(1 - 3x^2 + 2x^3)^4 \cdot 6x(1-x).$$

27．证明公式

$$\sum_{k=0}^{r} \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{r!(n-r-1)!}\int_p^1 x^r (1-x)^{n-r-1} dx,\ 其中\ 0 \le p \le 1.$$

证：设总体 X 服从区间(0, 1)上的均匀分布，$X_1, X_2, \cdots, X_n$ 为样本，$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$是顺序统计量，则样本观测值中不超过 p 的样品个数服从二项分布 $b(n, p)$，即最多有 r 个样品不超过 p 的概率为

$$P\{X_{(r+1)} > p\} = \sum_{k=0}^{r} \binom{n}{k} p^k (1-p)^{n-k},$$

因总体 X 的密度函数与分布函数分别为

$$p(x) = \begin{cases} 1, & 0 < x < 1; \\ 0, & 其他. \end{cases} \quad F(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \le x < 1; \\ 1, & x \ge 1. \end{cases}$$

则 $X_{(r+1)}$的密度函数为

$$p_{r+1}(x)=\frac{n!}{r!(n-r-1)!}[F(x)]^r[1-F(x)]^{n-r-1}p(x)=\begin{cases}\dfrac{n!}{r!(n-r-1)!}x^r(1-x)^{n-r-1}, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

故 $\sum_{k=0}^{r}\binom{n}{k}p^k(1-p)^{n-k}=P\{X_{(r+1)}>p\}=\dfrac{n!}{r!(n-r-1)!}\int_p^1 x^r(1-x)^{n-r-1}dx$.

28．设总体 X 的分布函数 $F(x)$是连续的，$X_{(1)},\cdots,X_{(n)}$为取自此总体的次序统计量，设 $\eta_i=F(X_{(i)})$，试证：

（1）$\eta_1\le\eta_2\le\cdots\le\eta_n$，且 η_i 是来自均匀分布 $U(0, 1)$总体的次序统计量；

（2）$E(\eta_i)=\dfrac{i}{n+1}$， $\mathrm{Var}(\eta_i)=\dfrac{i(n+1-i)}{(n+1)^2(n+2)}$， $1\le i\le n$；

（3）η_i 和 η_j 的协方差矩阵为

$$\begin{pmatrix}\dfrac{a_1(1-a_1)}{n+2} & \dfrac{a_1(1-a_2)}{n+2}\\ \dfrac{a_1(1-a_2)}{n+2} & \dfrac{a_2(1-a_2)}{n+2}\end{pmatrix}$$

其中 $a_1=\dfrac{i}{n+1}$， $a_2=\dfrac{j}{n+1}$.

注：第（3）问应要求 $i<j$.

解：（1）首先证明 $Y=F(X)$的分布是均匀分布 $U(0, 1)$，

因分布函数 $F(x)$连续，对于任意的 $y\in(0, 1)$，存在 x，使得 $F(x)=y$，

则 $F_Y(y)=P\{Y=F(X)\le y\}=P\{F(X)\le F(x)\}=P\{X\le x\}=F(x)=y$，

即 $Y=F(X)$的分布函数是

$$F_Y(y)=\begin{cases}0, & y<0;\\ y, & 0\le y<1;\\ 1, & y\ge 1.\end{cases}$$

可得 $Y=F(X)$的分布是均匀分布 $U(0, 1)$，即 $F(X_1), F(X_2),\cdots, F(X_n)$是均匀分布总体 $U(0, 1)$的样本，

因分布函数 $F(x)$单调不减，$\eta_i=F(X_{(i)})$，且 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$是总体 X 的次序统计量，

故 $\eta_1\le\eta_2\le\cdots\le\eta_n$，且 η_i 是来自均匀分布 $U(0, 1)$总体的次序统计量；

（2）因均匀分布 $U(0, 1)$ 的密度函数与分布函数分别为

$$p_Y(y)=\begin{cases}1, & 0<y<1;\\ 0, & \text{其他}.\end{cases}\quad F_Y(y)=\begin{cases}0, & y<0;\\ y, & 0\le y<1;\\ 1, & y\ge 1.\end{cases}$$

则 $\eta_i=F(X_{(i)})$的密度函数为

$$p_i(y)=\frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}p_Y(y)=\begin{cases}\dfrac{n!}{(i-1)!(n-i)!}y^{i-1}(1-y)^{n-i}, & 0<y<1,\\ 0, & \text{其他}.\end{cases}$$

即 η_i 服从贝塔分布 $Be(i, n-i+1)$，即 $Be(a, b)$，其中 $a=i$，$b=n-i+1$，

故 $E(\eta_i)=\dfrac{a}{a+b}=\dfrac{i}{n+1}$， $\mathrm{Var}(\eta_i)=\dfrac{ab}{(a+b)^2(a+b+1)}=\dfrac{i(n+1-i)}{(n+1)^2(n+2)}$， $1\le i\le n$；

（3）当 $i<j$ 时，(η_i, η_j)的联合密度函数为

$$p_{ij}(y,z)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F_Y(y)]^{i-1}[F_Y(z)-F_Y(y)]^{j-i-1}[1-F_Y(z)]^{n-j}p_Y(y)p_Y(z)\mathrm{I}_{y<z}$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} y^{i-1}(z-y)^{j-i-1}(1-z)^{n-j}\ \mathrm{I}_{0<y<z<1},$$

则 $E(\eta_i\eta_j) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yz\cdot p_{ij}(y,z)dydz = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\int_0^1 dz\int_0^z y^i(z-y)^{j-i-1}\cdot z(1-z)^{n-j}dy$,

令 $y = zu$，有 $dy = zdu$，且当 $y = 0$ 时，$u = 0$；当 $y = z$ 时，$u = 1$，

则 $\int_0^z y^i(z-y)^{j-i-1}\cdot z(1-z)^{n-j}dy = z(1-z)^{n-j}\int_0^1 (zu)^i(z-zu)^{j-i-1}\cdot zdu$

$$= z(1-z)^{n-j}\cdot z^j\int_0^1 u^i(1-u)^{j-i-1}du = z^{j+1}(1-z)^{n-j}\cdot B(i+1, j-i) = \frac{i!(j-i-1)!}{j!}z^{j+1}(1-z)^{n-j},$$

即 $E(\eta_i\eta_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\int_0^1 \frac{i!(j-i-1)!}{j!}z^{j+1}(1-z)^{n-j}dz$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\cdot\frac{i!(j-i-1)!}{j!}B(j+2, n-j+1)$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}\cdot\frac{i!(j-i-1)!}{j!}\cdot\frac{(j+1)!(n-j)!}{(n+2)!} = \frac{i(j+1)}{(n+1)(n+2)},$$

可得 $\mathrm{Cov}(\eta_i,\eta_j) = E(\eta_i\eta_j) - E(\eta_i)E(\eta_j) = \frac{i(j+1)}{(n+1)(n+2)} - \frac{i}{n+1}\cdot\frac{j}{n+1} = \frac{i(n+1-j)}{(n+1)^2(n+2)}$,

因 $a_1 = \frac{i}{n+1}$，$a_2 = \frac{j}{n+1}$，

则 $\mathrm{Cov}(\eta_i,\eta_j) = \frac{i(n+1-j)}{(n+1)^2(n+2)} = \frac{a_1(1-a_2)}{n+2}$,

且 $\mathrm{Var}(\eta_i) = \frac{i(n+1-i)}{(n+1)^2(n+2)} = \frac{a_1(1-a_1)}{n+2}$，$\mathrm{Var}(\eta_j) = \frac{j(n+1-j)}{(n+1)^2(n+2)} = \frac{a_2(1-a_2)}{n+2}$,

故 η_i 和 η_j 的协方差矩阵为

$$\begin{pmatrix} \mathrm{Var}(\eta_i) & \mathrm{Cov}(\eta_i,\eta_j) \\ \mathrm{Cov}(\eta_i,\eta_j) & \mathrm{Var}(\eta_j) \end{pmatrix} = \begin{pmatrix} \frac{a_1(1-a_1)}{n+2} & \frac{a_1(1-a_2)}{n+2} \\ \frac{a_1(1-a_2)}{n+2} & \frac{a_2(1-a_2)}{n+2} \end{pmatrix}.$$

29．设总体 X 服从 $N(0, 1)$，从此总体获得一组样本观测值

$$x_1 = 0, x_2 = 0.2, x_3 = 0.25, x_4 = -0.3, x_5 = -0.1, x_6 = 2, x_7 = 0.15, x_8 = 1, x_9 = -0.7, x_{10} = -1.$$

（1）计算 $x = 0.15$（即 $x_{(6)}$）处的 $E[F(X_{(6)})]$，$\mathrm{Var}[F(X_{(6)})]$；

（2）计算 $F(X_{(6)})$在 $x = 0.15$ 的分布函数值．

解：（1）根据第 28 题的结论知 $E[F(X_{(i)})] = \frac{i}{n+1}$，$\mathrm{Var}[F(X_{(i)})] = \frac{i(n+1-i)}{(n+1)^2(n+2)}$，且 $n = 10$，

故 $E[F(X_{(6)})] = \frac{6}{11}$，$\mathrm{Var}[F(X_{(6)})] = \frac{6\times 5}{11^2\times 12} = \frac{5}{242}$；

（2）因 $F(X_{(i)})$服从贝塔分布 $Be(i, n-i+1)$，即这里的 $F(X_{(6)})$服从贝塔分布 $Be(6, 5)$，

则 $F(X_{(6)})$在 $x = 0.15$ 的分布函数值为 $F_6(0.15) = \frac{10!}{5!\cdot 4!}\int_0^{0.15} x^5(1-x)^4dx$,

故根据第 27 题的结论知

$$F_6(0.15) = \frac{10!}{5!\cdot 4!}\int_0^{0.15} x^5(1-x)^4dx = 1 - \sum_{k=0}^{5}\binom{10}{k}\times 0.15^k\times 0.85^{10-k} = 0.0014.$$

30．在下列密度函数下分别寻求容量为 n 的样本中位数 $m_{0.5}$ 的渐近分布．

（1）$p(x)=6x(1-x)$，$0<x<1$；

（2）$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$；

（3）$p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & 其他.\end{cases}$

（4）$p(x)=\dfrac{\lambda}{2}\mathrm{e}^{-\lambda|x|}$.

解：样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(x_{0.5},\dfrac{1}{4n\cdot p^2(x_{0.5})}\right)$，其中 $p(x)$是总体密度函数，$x_{0.5}$ 是总体中位数，

（1）因 $p(x)=6x(1-x)$，$0<x<1$，有 $0.5=F(x_{0.5})=\int_0^{x_{0.5}}6x(1-x)dx=(3x^2-2x^3)\Big|_0^{x_{0.5}}=3x_{0.5}^2-2x_{0.5}^3$，

则 $x_{0.5}=0.5$，有 $\dfrac{1}{4n\cdot p^2(0.5)}=\dfrac{1}{4n\times(6\times0.5\times0.5)^2}=\dfrac{1}{9n}$，

故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(0.5,\dfrac{1}{9n}\right)$；

（2）因 $p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$，有 $0.5=F(x_{0.5})=F(\mu)$，

则 $x_{0.5}=\mu$，有 $\dfrac{1}{4n\cdot p^2(\mu)}=\dfrac{1}{4n\times\left(\dfrac{1}{\sqrt{2\pi}\sigma}\right)^2}=\dfrac{\pi\sigma^2}{2n}$，

故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(\mu,\dfrac{\pi\sigma^2}{2n}\right)$；

（3）因 $p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & 其他.\end{cases}$ 有 $0.5=F(x_{0.5})=\int_0^{x_{0.5}}2xdx=x^2\Big|_0^{x_{0.5}}=x_{0.5}^2$，

则 $x_{0.5}=\dfrac{1}{\sqrt2}$，有 $\dfrac{1}{4n\cdot p^2\left(\dfrac{1}{\sqrt2}\right)}=\dfrac{1}{4n\times\left(2\times\dfrac{1}{\sqrt2}\right)^2}=\dfrac{1}{8n}$，

故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(\dfrac{1}{\sqrt2},\dfrac{1}{8n}\right)$；

（4）因 $p(x)=\dfrac{\lambda}{2}\mathrm{e}^{-\lambda|x|}$，有 $0.5=F(x_{0.5})=F(0)$，

则 $x_{0.5}=0$，有 $\dfrac{1}{4n\cdot p^2(0)}=\dfrac{1}{4n\times\left(\dfrac{\lambda}{2}\right)^2}=\dfrac{1}{n\lambda^2}$，

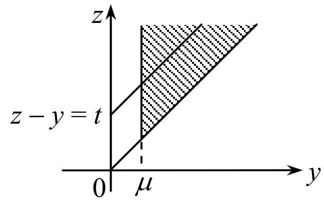
故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(0,\dfrac{1}{n\lambda^2}\right)$.

31．设总体 X 服从双参数指数分布，其分布函数为

$$F(x)=\begin{cases}1-\exp\left\{-\dfrac{x-\mu}{\sigma}\right\}, & x>\mu;\\ 0, & x\le\mu.\end{cases}$$

其中，$-\infty<\mu<+\infty$，$\sigma>0$，$X_{(1)}\le\cdots\le X_{(n)}$为样本的次序统计量．试证明 $(n-i-1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})$ 服从自由度为 2 的χ^2 分布（$i=2,\cdots,n$）.

注：此题有误，讨论的随机变量应为 $(n-i+1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})$.



证：因$(X_{(i-1)}, X_{(i)})$的联合密度函数为

$$p_{(i-1)i}(y,z)=\frac{n!}{(i-2)!(n-i)!}[F(y)]^{i-2}[1-F(z)]^{n-i}p(y)p(z)\mathrm{I}_{y<z}$$

$$=\frac{n!}{(i-2)!(n-i)!}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}\left[\exp\left\{-\frac{z-\mu}{\sigma}\right\}\right]^{n-i}\cdot\frac{1}{\sigma}\exp\left\{-\frac{y-\mu}{\sigma}\right\}\cdot\frac{1}{\sigma}\exp\left\{-\frac{z-\mu}{\sigma}\right\}\mathrm{I}_{\mu<y<z}$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\exp\left\{-\frac{y-\mu}{\sigma}\right\}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}\left[\exp\left\{-\frac{z-\mu}{\sigma}\right\}\right]^{n-i+1}\mathrm{I}_{\mu<y<z},$$

则 $T=X_{(i)}-X_{(i-1)}$的密度函数为

$$p_T(t)=\int_{-\infty}^{+\infty}p_{(i-1)i}(y,y+t)\cdot 1\cdot dy$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\int_{\mu}^{+\infty}\exp\left\{-\frac{y-\mu}{\sigma}\right\}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}\left[\exp\left\{-\frac{y+t-\mu}{\sigma}\right\}\right]^{n-i+1}dy$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\left[\exp\left\{-\frac{t}{\sigma}\right\}\right]^{n-i+1}\int_{\mu}^{+\infty}\left[\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{n-i+1}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}(-\sigma)d\left[\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\left[\exp\left\{-\frac{t}{\sigma}\right\}\right]^{n-i+1}\int_{1}^{0}u^{n-i+1}(1-u)^{i-2}(-\sigma)du$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\}\int_{0}^{1}u^{n-i+1}(1-u)^{i-2}du$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\}B(n-i+2,i-1)$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\}\cdot\frac{(n-i+1)!(i-2)!}{n!}=\frac{n-i+1}{\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\},\quad t>0,$$

可得 $S=(n-i+1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})=(n-i+1)\dfrac{2}{\sigma}T$ 的密度函数为

$$p_S(s)=p_T\left(\frac{\sigma}{2(n-i+1)}s\right)\cdot\frac{\sigma}{2(n-i+1)}=\frac{n-i+1}{\sigma}\exp\left\{-\frac{s}{2}\right\}\cdot\frac{\sigma}{2(n-i+1)}=\frac{1}{2}\exp\left\{-\frac{s}{2}\right\},\quad s>0,$$

故 $S=(n-i+1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})$ 服从参数为 $\dfrac{1}{2}$ 的指数分布，也就是服从自由度为 2 的 χ^2 分布．

32．设总体 X 的密度函数为

$$p(x)=\begin{cases}3x^2, & 0<x<1;\\ 0, & 其他.\end{cases}$$

$X_{(1)}\le X_{(2)}\le\cdots\le X_{(5)}$为容量为 5 的取自此总体的次序统计量，试证 $\dfrac{X_{(2)}}{X_{(4)}}$ 与 $X_{(4)}$相互独立．

证：因总体 X 的密度函数和分布函数分别为

$$p(x)=\begin{cases}3x^2, & 0<x<1;\\ 0, & 其他.\end{cases}\qquad F(x)=\begin{cases}0, & x<0;\\ x^3, & 0\le x<1;\\ 1, & x\ge 1.\end{cases}$$

则$(X_{(2)}, X_{(4)})$的联合密度函数为

$$p_{24}(x_{(2)},x_{(4)})=\frac{5!}{1!\cdot 1!\cdot 1!}[F(x_{(2)})]^1[F(x_{(4)})-F(x_{(2)})]^1[1-F(x_{(4)})]^1\,p(x_{(2)})p(x_{(4)})\,\mathrm{I}_{x_{(2)}<x_{(4)}}$$

$$=120x_{(2)}^3(x_{(4)}^3-x_{(2)}^3)(1-x_{(4)}^3)\cdot 3x_{(2)}^2\cdot 3x_{(4)}^2\,\mathrm{I}_{0<x_{(2)}<x_{(4)}<1}=1080x_{(2)}^5x_{(4)}^2(x_{(4)}^3-x_{(2)}^3)(1-x_{(4)}^3)\,\mathrm{I}_{0<x_{(2)}<x_{(4)}<1},$$

设 $Y_1=\dfrac{X_{(2)}}{X_{(4)}}$，$Y_2=X_{(4)}$，有 $X_{(2)}=Y_1Y_2$，$X_{(4)}=Y_2$，

则$(X_{(2)}, X_{(4)})$关于(Y_1, Y_2)的雅可比行列式为

$$J=\frac{\partial(x_{(2)},x_{(4)})}{\partial(y_1,y_2)}=\begin{vmatrix}y_2 & y_1\\ 0 & 1\end{vmatrix}=y_2,$$

且 $0<X_{(2)}\le X_{(4)}<1$ 对应于 $0<Y_1<1, 0<Y_2<1$，
可得(Y_1, Y_2)的联合密度函数为

$$p(y_1,y_2)=p_{24}(y_1y_2,y_2)\cdot|J|=1080(y_1y_2)^5y_2^2[y_2^3-(y_1y_2)^3](1-y_2^3)\,\mathrm{I}_{0<y_1<1,\,0<y_2<1}\cdot y_2$$

$$=1080y_1^5(1-y_1^3)\,\mathrm{I}_{0<y_1<1}\cdot y_2^{11}(1-y_2^3)\,\mathrm{I}_{0<y_2<1},$$

由于$(Y_1, Y_2, \cdots, Y_n)$的联合密度函数 $p(y_1, y_2)$可分离变量，

故 $Y_1=\dfrac{X_{(2)}}{X_{(4)}}$ 与 $Y_2=X_{(4)}$相互独立．

33．（1）设 $X_{(1)}$和 $X_{(n)}$分别为容量 n 的最小和最大次序统计量，证明极差 $R_n=X_{(n)}-X_{(1)}$的分布函数

$$F_{R_n}(x)=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}p(y)dy$$

其中 $F(y)$与 $p(y)$分别为总体的分布函数与密度函数；

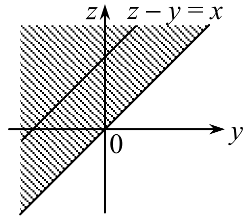
（2）利用（1）的结论，求总体为指数分布 $Exp(\lambda)$时，样本极差 R_n 的分布．

注：第（1）问应添上 $x>0$ 的要求．

解：（1）分布函数法，因$(X_{(1)}, X_{(n)})$的联合密度函数为

$$p_{1n}(y,z)=\frac{n!}{(n-2)!}[F(z)-F(y)]^{n-2}p(y)p(z)\mathrm{I}_{y<z}=n(n-1)[F(z)-F(y)]^{n-2}p(y)p(z)\mathrm{I}_{y<z}，$$

故 $R_n=X_{(n)}-X_{(1)}$的分布函数为



$$F_{R_n}(x)=P\{R_n=X_{(n)}-X_{(1)}\le x\}=\int_{-\infty}^{+\infty}dy\int_{-\infty}^{y+x}p_{1n}(y,z)dz$$

$$=n(n-1)\int_{-\infty}^{+\infty}dy\int_{y}^{y+x}[F(z)-F(y)]^{n-2}p(y)p(z)dz$$

$$=n(n-1)\int_{-\infty}^{+\infty}dy\cdot p(y)\int_{y}^{y+x}[F(z)-F(y)]^{n-2}d[F(z)]$$

$$=n(n-1)\int_{-\infty}^{+\infty}dy\cdot p(y)\cdot\frac{1}{n-1}[F(z)-F(y)]^{n-1}\Big|_{y}^{y+x}=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}p(y)dy，\quad x>0;$$

（2）因指数分布 $Exp(\lambda)$的密度函数与分布函数分别为

$$p(x)=\begin{cases}\lambda\,\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}\qquad F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}$$

故 $R_n=X_{(n)}-X_{(1)}$的分布函数为

$$F_{R_n}(x)=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}p(y)dy=n\int_{0}^{+\infty}[(1-\mathrm{e}^{-\lambda(y+x)})-(1-\mathrm{e}^{-\lambda y})]^{n-1}\cdot\lambda\,\mathrm{e}^{-\lambda y}\,dy$$

$$=n\int_{0}^{+\infty}(\mathrm{e}^{-\lambda y})^{n-1}(1-\mathrm{e}^{-\lambda x})^{n-1}\cdot(-1)d\,\mathrm{e}^{-\lambda y}=n(1-\mathrm{e}^{-\lambda x})^{n-1}\cdot\left(-\frac{1}{n}\right)(\mathrm{e}^{-\lambda y})^{n}\Big|_{0}^{+\infty}=(1-\mathrm{e}^{-\lambda x})^{n-1}，\quad x>0.$$

34．设 $X_1,\cdots,X_n$ 是来自 $U(0,\theta)$ 的样本，$X_{(1)}\le\cdots\le X_{(n)}$ 为次序统计量，令

$$Y_i=\frac{X_{(i)}}{X_{(i+1)}}，\quad i=1,\cdots,n-1，\quad Y_n=X_{(n)}，$$

证明 $Y_1,\cdots,Y_n$ 相互独立.

解：总体密度函数 $p(x)=\dfrac{1}{\theta}I_{0<x<\theta}$，

且 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 联合密度函数为 $p(x_{(1)},x_{(2)},\cdots,x_{(n)})=n!\dfrac{1}{\theta^n}I_{0<x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}<\theta}$，

由于 $Y_i=\dfrac{X_{(i)}}{X_{(i+1)}}$，$i=1,2,\ldots,n-1$，$Y_n=X_{(n)}$，

有 $X_{(1)}=Y_1Y_2\cdots Y_n$，$X_{(2)}=Y_2\cdots Y_n$，$\cdots$，$X_{(n-1)}=Y_{n-1}Y_n$，$X_{(n)}=Y_n$，
则 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 关于 $(Y_1,Y_2,\cdots,Y_n)$ 的雅可比行列式为

$$\frac{\partial(x_{(1)},x_{(2)},\cdots,x_{(n)})}{\partial(y_1,y_2,\cdots,y_n)}=\begin{vmatrix}y_2\cdots y_n & y_1y_3\cdots y_n & \cdots & y_1y_2\cdots y_{n-1}\\ 0 & y_3\cdots y_n & \cdots & y_2y_3\cdots y_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & 1\end{vmatrix}=y_2y_3^2\cdots y_n^{n-1}，$$

且 $0<X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}<\theta$ 对应于 $0<Y_1\le 1, 0<Y_2\le 1,\cdots, 0<Y_{n-1}\le 1, 0<Y_n<\theta$，
可得 $(Y_1,Y_2,\cdots,Y_n)$ 的联合密度函数为

$$p(y_1,y_2,\cdots,y_n)=n!\frac{1}{\theta^n}y_2y_3^2\cdots y_n^{n-1}I_{0<y_1\le 1}I_{0<y_2\le 1}\cdots I_{0<y_{n-1}\le 1}I_{0<y_n<\theta}，$$

由于 $(Y_1,Y_2,\cdots,Y_n)$ 的联合密度函数 $p(y_1,y_2,\cdots,y_n)$可分离变量，

故 $Y_1, Y_2, \cdots, Y_n$ 相互独立.

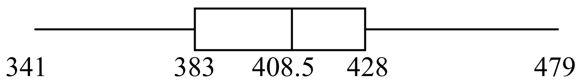
35．对下列数据构造箱线图

472	425	447	377	341	369	412	419
400	382	366	425	399	398	423	384
418	392	372	418	374	385	439	428
429	428	430	413	405	381	403	479
381	443	441	433	419	379	386	387

解：$x_{(1)} = 341$，$m_{0.25} = \dfrac{1}{2}(x_{(10)} + x_{(11)}) = 383$，$m_{0.5} = \dfrac{1}{2}(x_{(20)} + x_{(21)}) = 408.5$，$m_{0.75} = \dfrac{1}{2}(x_{(30)} + x_{(31)}) = 428$，

$x_{(n)} = 479$，

箱线图



36．根据调查，某集团公司的中层管理人员的年薪数据如下（单位：千元）

40.6	39.6	43.8	36.2	40.8	37.3	39.2	42.9
38.6	39.6	40.0	34.7	41.7	45.4	36.9	37.8
44.9	45.4	37.0	35.1	36.7	41.3	38.1	37.9
37.1	37.7	39.2	36.9	44.5	40.4	38.4	38.9
39.9	42.2	43.5	44.8	37.7	34.7	36.3	39.7
42.1	41.5	40.6	38.9	42.2	40.3	35.8	39.2

试画出箱线图.

解：$x_{(1)} = 34.7$，$m_{0.25} = \dfrac{1}{2}(x_{(12)} + x_{(13)}) = 37.5$，$m_{0.5} = \dfrac{1}{2}(x_{(24)} + x_{(25)}) = 39.4$，$m_{0.75} = \dfrac{1}{2}(x_{(36)} + x_{(37)}) = 41.6$，

$x_{(n)} = 45.4$，

箱线图