一、随机变量X的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x\ge\mu;\\ 0, & x<\mu.\end{cases}\qquad(\theta>0)$$

计算方差$D(X)$.

解　$E(X)=\int_{\mu}^{+\infty}\frac{x}{\theta}e^{-\frac{x-\mu}{\theta}}dx=\int_{0}^{+\infty}\frac{1}{\theta}(y+\mu)e^{-\frac{y}{\theta}}dy$　（式中$y=\dfrac{x-\mu}{\theta}$）

$$=\int_{0}^{+\infty}\frac{y}{\theta}e^{-\frac{y}{\theta}}dy+\mu\int_{0}^{+\infty}\frac{1}{\theta}e^{-\frac{y}{\theta}}dy=E(Y)+\mu=\theta+\mu$$

$$E(X^2)=\int_{\mu}^{+\infty}\frac{x^2}{\theta}e^{-\frac{x-\mu}{\theta}}dx=\int_{0}^{+\infty}\frac{1}{\theta}(y+\mu)^2e^{-\frac{y}{\theta}}dy$$

$$=\int_{0}^{+\infty}\frac{y^2}{\theta}e^{-\frac{y}{\theta}}dy+\int_{0}^{+\infty}2y\mu\frac{1}{\theta}e^{-\frac{y}{\theta}}dy+\mu^2\int_{0}^{+\infty}\frac{1}{\theta}e^{-\frac{y}{\theta}}dy$$

$$=E(Y^2)+2\mu E(Y)+\mu^2=2\theta^2+2\mu\theta+\mu^2.$$

$$D(X)=\mathrm{E}(X^2)-[\mathrm{E(X)}]^2=2\theta^2+2\mu\theta+\mu^2-(\theta+\mu)^2=\theta^2$$

二、$(X,\ Y)$服从二维正态分布$N(\mu,\sigma^2;\mu,\sigma^2;\rho)$，其中$\sigma^2>0,\rho\ne0$.设随机变量$Z_1=X+Y$与$Z_2=X-Y$. 讨论$Z_1$与$Z_2$的相关性和独立性, 并确定$(Z_1,Z_2)$的联合概率密度.

解　$E(X+Y)=2\mu,\quad E(X-Y)=0$

$$D(Z_1)=D(X)+D(Y)+2Cov(X,Y)=2\sigma^2+2\rho\sigma^2=2(1+\rho)\sigma^2>0$$

$$D(Z_2)=D(X)+D(Y)-2Cov(X,Y)=2\sigma^2-2\rho\sigma^2=2(1-\rho)\sigma^2>0$$

$$Cov(Z_1,Z_2)=Cov(X+Y,X-Y)$$

$$=Cov(X,X)-Cov(Y,Y)+Cov(X,Y)-Cov(X,Y)$$

$$=D(X)-D(Y)=0$$

从而$\rho_{Z_1Z_2}=0$, 即Z_1与Z_2不相关.

又因$(X,\ Y)$服从二维正态分布，Z_1与Z_2是二维正态分布$(X,\ Y)$的线性组合, 所以(Z_1,Z_2)服从二维正态分布, 可知Z_1,Z_2相互独立,其联合概率密度为

$$\varphi(z_1,z_2)=\varphi(z_1)\varphi(z_2)$$

$$=\frac{1}{2\pi}\exp\{\frac{(z_1-2\mu)^2}{4(1+\rho)\sigma^2}+\frac{{z_2}^2}{4(1-\rho)\sigma^2}\},\qquad(z_1,z_2)\in R^2$$

三、某射手每次射中目标的概率为 p，现有 10 发子弹，准备对一目标连续射击（每次打一发），一旦射中或子弹打完了就立刻转移到别的地方。问他在转移前平均射击多少次？

解：设他转移前射击了 X 次，则 X 的可能取值为 1,2,…,10。当$1\le k\le9$时，事件$\{X=k\}$表示前 k-1 次未中，第 k 次才射中；事件$\{X=10\}$表示前 9 次未中，射手需射第 10 次，且射 10 次后不论中还是不中都需转移。于是 X 的分布律为：

$$P\{X=k\}=(1-p)^{k-1}p=q^{k-1}p,(k=1,2,\cdots,9;q=1-p)$$

$$P\{X=10\}=(1-p)^{k-1}(p+q)=q^9$$

则他在转移前射击的平均数为：

$$E(X)=\sum_{k=1}^{10}kp_k=p(1+2q+3q^2+\cdots+9q^8)+10q^9$$
$$=p(1+q+q^2+\cdots+q^9)'+10q^9$$
$$=p(\frac{1-q^{10}}{1-q})'+10q^9$$
$$=p\frac{-10q^9(1-q)+(1-q^{10})}{(1-q)^2}+10q^9$$
$$=\frac{1}{p}(1-q^{10})=\frac{1}{p}[1-(1-p)^{10}]$$

四、设随机变量 X 与 Y 同分布，X 的概率密度为

$$f(x)=\begin{cases}\frac{3}{8}x^2, 0<x<2\\ 0, \quad 其他\end{cases}$$

（1） 设事件$\{X>a\}=A$ 与事件$\{Y>a\}=B$ 相互独立，且 $P\{A+B\}=3/4$，求常数 a；

（2）求 $1/X^2$的数学期望。

解：由题设知 $P(A)=P(B), P(AB)=P(A)P(B)$， $P(A+B)=P(A)+P(B)-P(A)P(B)$

$=2P(A)-[P(A)]^2=3/4$，解得 $P(A)=1/2$，而

$\frac{1}{2}=P(A)=P\{X>a\}=\int_{-\infty}^{+\infty}f(x)dx=\int_a^2\frac{3}{8}x^2dx=\frac{1}{8}(8-a^3)$，由此解得 $a=\sqrt[3]{4}$。

（2） $E(1/X^2)=\int_{-\infty}^{+\infty}\frac{1}{x^2}f(x)dx=\int_0^2\frac{1}{x^2}\cdot\frac{3}{8}x^2dx=\frac{3}{4}$。

五、两个随机变量的相关系数表征了二者间的什么关系? 若 $\boldsymbol{\rho_{XY}=0}$，能否说它们无关系？请举例说明.

答: 两个随机变量的相关系数是刻画二者间线性相关程度的数字特征. 若 $\boldsymbol{\rho_{XY}=0}$ 仅说明二者无明显的线性关系, 但可能有其它关系.

例如: P116.例 4.4.4 随机变量 $\theta\sim U(0,2\pi)$, $X=\cos\theta, Y=\cos(\theta+\frac{\pi}{2})$，有 $\boldsymbol{\rho_{XY}=0}$,即二者不相关,但又有 $X^2+Y^2=1$, 二者有确定的函数关系.

六、为较为精确地测量某种零件的长度 μ，在相同条件下对其进行 n 次独立测量. 记第 k 次的测量结果是随机变量 X_k，将 n 次测量结果的平均 $Y_n=\frac{1}{n}\sum_{k=1}^{n}X_k$ 作为长度的最终测量值. 请你用自己掌握的理论解释这种测量方法的合理性.

解 设 $E(X_k)=\mu, D(X_k)=\sigma^2$,有

(1) $E(\frac{1}{n}\sum_{i=1}^{n}X_k)=\frac{1}{n}\sum_{i=1}^{n}E(X_k)=\frac{1}{n}\cdot n\mu=\mu$

(2) $D(\frac{1}{n}\sum_{i=1}^{n}X_k)=\frac{1}{n^2}D(\sum_{i=1}^{n}X_k)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_k)=\frac{1}{n^2}\cdot n\sigma^2=\sigma^2$

解释一 (1)式说明最终测量值Y_n的取值的集中点是工件的真实长度.

由(2)式知当$n \ge 2$时,有$D(Y_n)=\dfrac{\sigma^2}{n}<\sigma^2$，且随着测量次数的增大，方差减小.说明最终测量结果Y_n相对工件真实长度μ的离散程度在减小(集中程度增高)，精度得以提高.

解释二 设想在相同条件下的无限次重复测量,其结果构成相互独立同分布随机变量序列X_k，$\mathrm{k}=1,2,\cdots$，根据 P129 定理 5.2.3 独立同分布大数定律知，对任意$\varepsilon>0$

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1$$

当重复测量次数足够大时，Y_n相对工件真实长度μ的偏离程度控制在任意小的范围内的概率很大，精度得以提高.

五章

1、 设有一批电子元件，合格品占 1/6。从中任意选择 6000 个，试问把误差限ε定为多少时，才能保证频率与概率之差的绝对值不大于ε的概率为 99%？此时，合格品数落在哪个范围内？

解：设 6000 个电子元件中合格品数为X，$X\sim B(n,p)$,其中 n=6000，p=1/6，由棣莫弗-拉普拉斯中心极限定理可得：

$$P\{|\frac{X}{6000}-p|<\varepsilon\}=P\{-\varepsilon\sqrt{\frac{n}{pq}}<\frac{X-np}{\sqrt{npq}}<\varepsilon\sqrt{\frac{n}{pq}}\}$$

$$\approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}})-1=0.99$$

即：$\Phi(\varepsilon\sqrt{\frac{n}{pq}})=0.995$，查表得$\varepsilon\sqrt{\frac{n}{pq}}=2.58,\varepsilon=0.0124$，把$\varepsilon=0.0124$代入上式得

$$P\{|\frac{X}{6000}-\frac{1}{6}|<0.0124\}=P\{|X-1000|<74.4\}=P\{925.6<X<1074.4\}=0.99$$

即相应合格品的个数落在 926 个与 1027 个之间。

2、 某校有 900 名学生选修 6 名教师主讲的“高等数学”课。假定每名学生完全随意地选择一位老师，且学生之间选择教师是彼此独立的。问每个教师的上课教室应该设有多少座位才能保证因缺少座位而使学生离去的概率小于 1%？其中

（$\Phi(2.33)=0.9901,\Phi(2.4)=0.9918,\Phi(2.43)=0.9925$）

解：只需考虑某个教师甲的上课教室，设教室需设M个座位。定义随机变量如下：

$$X_i=\begin{cases}1,\text{第}i\text{个学生选择教师甲}\\0,\text{其他}\end{cases}，i=1,2,\cdots,900.$$

则$P\{X_i=1/6\},P\{X_i=5/6\}$，且$X_1,X_2,\cdots,X_{900}$相互独立同分布，$E(X_i)=1/6,$

$D(X_i)=5/36$。选择教师甲的学生总人数为$X=\sum_{i=1}^{900}X_i$，为达到要求则需 M 使得

$P\{M \geq X\} \geq 99\%$。应用独立同分布中心极限定理得：

$$P\{X \leq M\} = P\{\sum_{i=1}^{900} X_i \leq M\} = P\{\frac{\sum_{i=1}^{900}(X_i - 1/6)}{30\times\sqrt{5}/6} \leq \frac{M-150}{5\sqrt{5}}\}$$
$$\approx \Phi(\frac{M-150}{5\sqrt{5}}) \geq 99\%$$

查标准正太分布表得 $\frac{M-150}{5\sqrt{5}} \geq 2.33, M \geq 150 + 2.33\times 5\sqrt{5} = 176.05$。因此取 M=176 即可。

3、在计算机模拟试验中，将由 12 个相互独立同在(0，1)上服从均匀分布的随机变量$X_1, X_2, \ldots,$ X_{12}的函数 $Y = \sum_{i=1}^{12} X_i - 6$ 视为标准正态分布的随机变量．请你给出理论解释．

解　将 $X_2, X_2 \cdots, X_{12}$ 视为独立同分布随机变量序列 $X_2, X_2 \cdots, X_n \cdots$ 的前 12 项，因

$$X_n \sim U(0,1), E(X_n) = \frac{1}{2}, D(X_n) = \frac{1}{12}$$

根据独立同分布中心极限定理知 $X_1, X_2, \cdots, X_n, \cdots$ 服从中心极限定律，其前 12 项和的标准化随机变量

$$\frac{[\sum_{i=1}^{12} X_i - 6]}{\sqrt{12/12}} = \sum_{i=1}^{12} X_i - 6$$

近似服从标准正态分布．

4、请用独立同分布中心极限定理解释现实中哪一类随机变量可用正态分布描述？

解 现实中的随机变量呈现为**较大量相互独立的微小随机变量X_k的叠加总和** $\sum_{k=1}^{n} X_k$ 时可用正态分布描述．

根据独立同分布中心极限定理，若随机变量序列{X_k }，k =1,2…满足以下条件：

1.相互独立；　2.具有相同分布；3.数学期望和方差均存在．

则有

$$\lim_{n\to\infty} P\left\{ Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leq x \right\} = \Phi(x)$$

成立．即可认为当 n 足够大时时，Y_n 标准化后的分布函数近似标准正态分布函数 $\Phi(x)$，故可认为

$$\sum_{k=1}^{n} X_k \sim N\left(\sum_{k=1}^{n} E(X_k), \sum_{k=1}^{n} D(X_k) \right)$$

近似成立,

六章答案

习题六

P150-151

1、4、5、6、7、8、10、11、12

补充：

1、总体 X~$N(\mu,\sigma^2)$，$X_1, X_2, \ldots, X_{20}$ 是 X 的一个样本，令 $Y=3\sum_{i=1}^{10}X_i-4\sum_{i=11}^{20}X_i$ ，

则 Y 服从分布（　　　）　　　　　。

2、$X_1, X_2, \ldots, X_5$ 是来自总体 X~$N(0,1)$的一个样本，若 $\dfrac{C(X_1+X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$ **服从 t 分布，**

则常数 C=（　　）。

3 设$X_1, X_2, \ldots, X_8$ 和$Y_1, Y_2, \ldots, Y_{10}$ 分别来自正态总体$N(-1, 2^2)$和$N(2, 5)$的样本，且相互

独立，S_1^2 和S_2^2 分别表示两样本的样本方差，则服从$F(7,9)$的统计量是__________.____

$(A)\ \dfrac{2S_1^2}{5S_2^2};\qquad (B)\ \dfrac{5S_1^2}{4S_2^2};\qquad (C)\ \dfrac{4S_2^2}{5S_1^2};\qquad (D)\ \dfrac{5S_1^2}{2S_2^2};$

4. 样本 $X_1, X_2, \ldots, X_n(n>1)$来自总体 X~$N(0,1)$，$\bar{X}$ 与 S 分别是样本均值和样本标准差，则

有_____.

$(A)\ \bar{X}\sim N(0,1);\qquad (B)\ nX\sim N(0,1);\qquad (C)\ \sum_{i=1}^{n}X_i^2\sim\chi^2(n);\qquad (D)\ \dfrac{\bar{X}}{S}\sim t(n-1);$

5. $X_1, X_2, \ldots, X_n(n>1)$是来自 X~$N(\mu,\sigma^2)$的一个样本，$\bar{X}$ 和 S^2 分别是样本均值和样本方差，则下列结论正确的是____________.

$(A)\quad 2X_1-X_2\sim N(\mu,\sigma^2);\qquad (B)\quad \dfrac{n(\bar{X}-\mu)^2}{S^2}\sim F(1,n-1);$

$(C)\quad \dfrac{S^2}{\sigma^2}\sim\chi^2(n-1);\qquad (D)\quad \dfrac{\bar{X}-\mu}{S}\sqrt{n-1}\sim t(n-1);$

6、$X_1, X_2, \ldots, X_n$ 是来自总体 X~$N(\mu,\sigma^2)$ 的一个样本，S^2 为样本方差，求样本容量的最大值，使其满足不等式 $P\{\dfrac{(n-1)S^2}{\sigma^2}\le 15\}\ge 0.95$

1、$N(-10\mu, 250\sigma^2)$

2、$\boxed{\sqrt{\dfrac{3}{2}}}$

3、B

4、C

5、B

6、

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

$$P\{\frac{(n-1)S^2}{\sigma^2}\le 15\}\ge 0.95 \quad \Rightarrow P\{\frac{(n-1)S^2}{\sigma^2}>15\}<0.05$$

查表得 $\chi^2_{0.05}(7)=14.067, \chi^2_{0.05}(8)=15.507$

因此 $n-1\leqslant 7$，即 $n\leqslant 8$ 时不等式成立，样本容量的最大值为 8．

七章

1、 设总体的概率密度为 $f(x)=\begin{cases}2e^{-2(x-\theta)}, x>\theta \\ 0, \quad x\le\theta\end{cases}$，其中$\theta>0$是未知参数，从总体$X$中抽取随机样本$X_1,X_2,\cdots,X_n$，记$\hat{\theta}=\min\{X_1,X_2,\cdots,X_n\}$。（1）求总体$X$的分布函数；（2）求统计量$\hat{\theta}$的分布函数$F_{\hat{\theta}}(x)$；（3）如果用$\hat{\theta}$作为$\theta$的估计量，讨论它是否具有无偏性。

解：（1）$F(x)=\int_{-\infty}^{x} f(x)dx=\begin{cases}1-e^{-2(x-\theta)}, x>\theta \\ 0, \quad x\le\theta\end{cases}$

（2）
$$\begin{aligned}F_{\hat{\theta}}(x)&=P\{\hat{\theta}\le x\}=P\{\min\{X_1,X_2,\cdots,X_n\}\le x\}\\&=1-P\{\min\{X_1,X_2,\cdots,X_n\}>x\}\\&=1-P\{X_1>x,X_2>x,\cdots X_n>x\}\\&=1-P\{X_1>x\}P\{X_2>x\}\cdots P\{X_n>x\}\\&=1-[1-F(x)]^n=\begin{cases}1-e^{-2n(x-\theta)}, x>\theta \\ 0, \quad x\le\theta\end{cases}\end{aligned}$$

（3）$\hat{\theta}$的概率密度为 $f_{\hat{\theta}}(x)=F'_{\hat{\theta}}(x)=\begin{cases}2ne^{-2n(x-\theta)}, x>\theta \\ 0, \quad x\le\theta\end{cases}$，因为

$E(\hat{\theta})=\int_{-\infty}^{+\infty} xf_{\hat{\theta}}(x)dx=\int_{0}^{+\infty} 2nxe^{-2n(x-\theta)}dx=\theta+\frac{1}{2n}\ne\theta$，所以$\hat{\theta}$不具有无偏性。