西南交通大学 2005—2006 学年

第（一）学期考试试卷

一、（15 分）某种运动汽车有自动排挡和手动排挡以及红黑蓝白四种颜色可供选择，颜色和排挡搭配的概率如下表：

<table>
<tr><td></td><td></td><td colspan="4">颜色</td></tr>
<tr><td></td><td></td><td>白</td><td>蓝</td><td>黑</td><td>红</td></tr>
<tr><td rowspan="2">排挡类型</td><td>自动</td><td>0.15</td><td>0.10</td><td>0.10</td><td>0.10</td></tr>
<tr><td>手动</td><td>0.15</td><td>0.05</td><td>0.15</td><td>0.20</td></tr>
</table>

从这种运动汽车中随机选取一辆，记

$A=$“自动排挡”，　$B=$“黑色”，　$C=$“白色”。1.用字母表示以下事件：

（1）选中的车为深色手动排挡；（2）选中的车为白色自动排挡汽车；（3）选中的车为黑色手动排挡汽车；（4）选中的车为蓝色或红色自动排挡汽车。

2.计算：$P(A),P(B),P(A\cup B)$。

3. 计算：$P(A|C),P(A|\bar{C})$并说明其意义。

【解题过程】1.考查对随机事件的表示，（1）深色手动为深色和手动的积事件，深色又是白色的对立事件，手动为自动的对立事件；（2）白色自动为白色和自动的积事件；（3）黑色手动为黑色和手动的积事件；（4）蓝色或红色自动为蓝色或红色与自动的积事件，蓝色或红色为白色与黑色的积事件的对立事件。2.考查概率的计算和概率的加法公式，即对于任意两事件A,B有

$P\left(A\cup B\right)=P(A)+P(B)-P(AB)$。3.考查条件概率，设$A,B$是两个事件，且$P(A)>0$，称$(B|A)=\dfrac{P(AB)}{P(A)}$为在事件$A$发生的条件下事件$B$发生的条件概率。

【解题过程】1.（1）深色手动排挡为$\overline{A}\overline{C}$;(2)白色自动排挡为AC；（3）黑色手动排挡为$\overline{A}B$；（4）蓝色或红色自动排挡为$A\overline{BC}$；

2. $P(A)=0.15+0.10+0.10+0.10=0.45;$

$P(B)=0.1+0.15=0.25$;

$P(A\cup B)=P(A)+P(B)-P(AB)$

$=0.45+0.25-0.10=0.60$

3. $P(A|C)$表示一辆白色运动车是自动排挡的概率，即

$$P(A|C)=\frac{P(AC)}{P(C)}=\frac{0.15}{0.15+0.15}=\frac{1}{2};$$

$P(A|\overline{C})$表示一辆深色运动车是自动排挡的概率，即

$$P(A|\overline{C})=\frac{P(A\overline{C})}{P(\overline{C})}=\frac{0.10+0.10+0.10}{1-(0.15+0.15)}=\frac{3}{7}$$

二、（10 分）设A,B两人连续下象棋，直到两人中有一人赢 10 局。记$S=$“A赢一局”，$F=$“B赢一局”。假定在下棋的过程中各局的结果相互独立，$P(S)=p$，$P(F)=q$，$X=$“结束下棋时他们所下的局数”。

1.在下棋的过程中没有和棋$\left(p+q=1\right)$，求X的概率分布；

2.在下棋的过程中有和棋$\left(p+q\neq 1\right)$，求X的概率分布。

【解题过程】考查离散型随机变量的概率分布，设离散型随机变量X所有可能取的值为$x_k\left(k=1,2,\cdots\right)$，$X$取各个可能值得概率，即事件$\left\{X=x_k\right\}$的概率为$P\left\{X=x_k\right\}=p_k,k=1,2,\cdots$。此式为离散型随机变量$X$的分布律。此题考查二项分布，若没有和棋则$X$的取值$10,11,\cdots,19$，若有和棋则$X$可能的取值为$10,11,12,\cdots$。

【解题过程】1. X可能的取值为$10,11,\cdots,19$，则

$$P(X=k)=P(A\text{赢})+P(B\text{赢})$$
$$=pC_{k-1}^{9}p^{9}q^{k-10}+qC_{k-1}^{9}q^{9}p^{k-10}$$

$$=C_{k-1}^{9}\left(p^{10}q^{k-10}+q^{10}p^{k-10}\right)$$

2. X 可能的取值为 $10,11,12,\cdots$，则

$$P(X=k)=P(A\text{赢})+P(B\text{赢})$$
$$=pC_{k-1}^{9}p^{9}\left[\sum_{i=0}^{\min\{k-10,9\}}C_{k-10}^{i}q^{i}(1-p-q)^{k-10-i}\right]$$
$$+qC_{k-1}^{9}q^{9}\left[\sum_{i=0}^{\min\{k-10,9\}}C_{k-10}^{i}p^{i}(1-p-q)^{k-10-i}\right]$$

三、（20 分）设随机变量 X 具有密度函数

$$f(x)=\begin{cases}\dfrac{k}{x^{\alpha}},x\geq 5\\ 0,x<5\end{cases}$$ 。

1.求 k，对 α 的必要限制是什么？

2.求 X 的分布函数；

3.求 EX，若要求 EX 有限，对 α 有什么限制？

4.求 $\ln\left(\dfrac{X}{5}\right)$ 的分布。

【解题过程】1.2.考查连续型随机变量的分布函数及其性质，概率密度函数。设 X 为一个随机变量，对任意实数 x，称 $F(x)=P\left\{X\le x\right\}$ 为 X 的分布函数。分布函数具有单调不减性，有界性，右连续性等性质。设随机变量 X 的分布函数为 $F(x)$，如果存在非负可积函数 $f(x)$，使得对任意实数 x，均有 $F(x)=\int_{-\infty}^{x}f(t)dt$，则称 X 为连续型随机变量，其中函数 $f(x)$ 称为 X 的概率密度函数。3.考查了连续型随机变量的期望，设连续型随机变量 X 的概率密度为 $f(x)$，若积分 $\int_{-\infty}^{\infty}xf(x)dx$ 绝对收敛，则称积分 $\int_{-\infty}^{\infty}xf(x)dx$ 的值为随机变量 X 的数学期望，记为 $E(X)$。即 $E(X)=\int_{-\infty}^{\infty}xf(x)dx$。

4.考查连续型随机变量的函数的分布。即设随机变量 X 具有概率密度 $f_X(x)$，$-\infty < x < \infty$，又设函数 $g(x)$ 处处可导且恒有 $g(x)' > 0$（或恒有 $g(x)' < 0$），则 $Y = g(X)$ 是连续型随机变量，其概率密度为 $f_Y(y) = \begin{cases} f_X[h(y)]\left|h'(y)\right|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$，

其中 $\alpha = \min\{g(-\infty), g(\infty)\}$，$\beta = \max\{g(-\infty), g(\infty)\}$，$h(y)$ 是 $g(x)$ 的反函数。

【**解题过程**】1.由概率密度函数的性质可得：

$$\int_{-\infty}^{\infty} f(x)dx = \int_{5}^{\infty} \frac{k}{x^{\alpha}}dx = \begin{cases} +\infty, \alpha \le 1 \\ \dfrac{k}{\alpha-1}5^{1-\alpha}, \alpha > 1 \end{cases},$$

故 $\alpha > 1$，且 $k = \dfrac{\alpha-1}{5^{1-\alpha}}$。

2.当 $x < 5$ 时，$F(x) = 0$；

当 $x \ge 5$ 时，

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{5}^{x} \frac{k}{t^{\alpha}}dt = 1 - \left(\frac{x}{5}\right)^{1-\alpha};$$

所以 X 的分布函数为

$$F(x) = \begin{cases} 0, x < 5 \\ 1 - \left(\dfrac{x}{5}\right)^{1-\alpha}, x \ge 5 \end{cases}。$$

3.

$$EX = \int_{-\infty}^{\infty} xf(x)dx = \int_{5}^{\infty} x \cdot \frac{k}{x^{\alpha}}dx = \frac{k}{2-\alpha}x^{2-\alpha}\Big|_{5}^{\infty},$$

若要求 EX 有限，则 $\alpha > 2$。

4. 由于 $y'=\frac{1}{x}(x\geq 5)$，故 $y'>0$。且

$x=5e^y$，所以

$$f_Y(y)=\begin{cases} f_X[5e^y]\left|5e^y\right| \\ 0 \end{cases}=\begin{cases} \frac{(\alpha-1)5^{\alpha-1}}{\left(5e^y\right)^{\alpha}}\cdot 5e^y \\ 0 \end{cases}$$

$$=\begin{cases} (\alpha-1)e^{(1-\alpha)y}, y\geq 0 \\ 0, y<0 \end{cases}。$$

四、（10 分）对一枚均匀的硬币，至少要掷多少次才能保证正面出现的频率在 $0.4\sim 0.6$ 之间的概率不小于 0.9？（标准正态分布的分布函数取值为已知）

【解题过程】考查独立同分布的中心极限定理，设随机变量 $X_1,X_2,\cdots,X_n\cdots$ 相互独立，且服从同一分布，且具有数学期望和方差：

$E(X_k)=\mu$，

$D(X_k)=\sigma^2>0$ $(k=1,2,\cdots)$，则随机变量

之和 $\sum_{k=1}^{n} X_k$ 的标准化变量

$$Y_n=\frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}}=\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$ 的

分布函数 $F_n(x)$ 对于任意 x 满足

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

$$=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)$$ 。

也就是说，均值为 μ，方差为 $\sigma^2>0$ 的独

立同分布的随机变量 $X_1,X_2,\cdots,X_n\cdots$ 之和

$\sum_{k=1}^{n} X_k$ 的标准化变量，当 n 充分大时，有

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \sim N(0,1)。$$

【解题过程】设至少要掷 n 次，记随机变量

$$X_i = \begin{cases} 1,\text{第}i\text{次出现的是正面} \\ 0,\text{第}i\text{次出现的是反面} \end{cases},$$

$i = 1,2,\cdots,n$，则 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 表示掷 n 次正面出现的频率，

由于 $P(X_i = 0) = P(X_i = 1) = 0.5$,

$E(X_i) = 0.5$,

$D(X_i) = 0.5 - 0.5 \times 0.5 = 0.25$。

根据独立同分布中心极限定理，有:

$P\left(0.4 < \bar{X} < 0.6\right)$

$$= P\left(\frac{0.4-0.5}{0.5/\sqrt{n}} < \frac{\bar{X}-0.5}{0.5/\sqrt{n}} < \frac{0.6-0.5}{0.5/\sqrt{n}}\right)$$

$$= \Phi\left(\frac{0.6-0.5}{0.5/\sqrt{n}}\right) - \Phi\left(\frac{0.4-0.5}{0.5/\sqrt{n}}\right)$$

$= 2\Phi\left(\frac{\sqrt{n}}{5}\right) - 1 \geq 0.9$，即 $\Phi\left(\frac{\sqrt{n}}{5}\right) \geq 0.95$，

可解得 n。

五、（15 分）设总体 X 的具有密度函数

$$f(x,\lambda,\theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)}, x \geq \theta \\ 0, x < \theta \end{cases},$$

$X_1, X_2, \cdots, X_n$ 为来自总体 X 的样本。

1.求参数 λ, θ 的矩估计量。

2.求参数 λ, θ 的极大似然估计量。

3.如果$n=10$，样本观测值如下：3.11 0.64 2.55 2.20 5.44 3.42 10.39 8.93 17.82 1.30 计算参数λ,θ的极大似然估计值。

【解题过程】固定样本观察值$x_1,x_2,\cdots,x_n$，在θ取值的可能范围Θ内挑选使似然函数$L\left(x_1,x_2,\cdots,x_n;\widehat{\theta}\right)$达到最大的参数值，作为参数$\theta$的估计值。即取$\widehat{\theta}$使

$$L\left(x_1,x_2,\cdots,x_n;\widehat{\theta}\right)=\max_{\theta\in\Theta}L\left(x_1,x_2,\cdots,x_n;\theta\right)$$

。这样得到的$\widehat{\theta}$与样本值$x_1,x_2,\cdots,x_n$有关，常记为$\widehat{\theta}\left(x_1,x_2,\cdots,x_n\right)$，称为参数$\theta$的最大似然估计值，而相应的统计量$\widehat{\theta}\left(X_1,X_2,\cdots X_n\right)$称为参数$\theta$的最大似然估计量。还考查了矩估计值。

【解题过程】 1.由题知:

$$EX=\int_{\theta}^{+\infty}x\cdot\lambda e^{-\lambda(x-\theta)}dx$$

$$=x\cdot(-e^{-\lambda(x-\theta)})\Big|_{\theta}^{+\infty}+\int_{\theta}^{+\infty}e^{-\lambda(x-\theta)}dx$$

$$=\frac{1}{\lambda}+\theta$$

$$EX^2=\int_{\theta}^{+\infty}x^2\cdot\lambda e^{-\lambda(x-\theta)}dx$$

$$=x^2\cdot(-e^{-\lambda(x-\theta)})\Big|_{\theta}^{+\infty}+2\int_{\theta}^{+\infty}xe^{-\lambda(x-\theta)}dx$$

$$=\theta^2+\frac{2\theta}{\lambda}+\frac{2}{\lambda^2}$$

$$DX=EX^2-\left(EX\right)^2=\lambda^{-2}$$

令 $\begin{cases}\dfrac{1}{\widehat{\lambda}}+\widehat{\theta}=\overline{X}\\ \widehat{\lambda}^{-2}=S_n^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2\end{cases}$

则 $\begin{cases}\widehat{\theta}=\overline{X}-S_n\\ \widehat{\lambda}=\dfrac{1}{S_n}\end{cases}$

2.似然函数为

$$L\left(x,\lambda,\theta\right)=\prod_{i=1}^{n}f(x_i,\lambda,\theta)$$

$$=\begin{cases}\lambda^n e^{-\lambda\left(\sum x_i-n\theta\right)},\theta<x_i,\lambda>0\\0,\text{其他}\end{cases}$$

$\ln L=-\lambda\left(n\overline{x}-n\theta\right)+n\ln\lambda$ ，

令 $\dfrac{\partial\ln L}{\partial\theta}=n\lambda=0$ ，无解

$\dfrac{\partial\ln L}{\partial\lambda}=-(n\overline{x}-n\theta)+\dfrac{n}{\lambda}=0$ 则 $\dfrac{1}{\lambda}+\theta=\overline{x}$

又 $x_i>\theta$ 则 $\theta<\min\left(x_1,x_2,\cdots,x_n\right)$ ，得似然估计：

$\widehat{\theta}=\min\left(X_1,X_2,\cdots,X_n\right)=X_{(i)}$，

$\widehat{\lambda}=\dfrac{1}{\overline{X}-X_{(i)}}$。

3.由于 $x_{(i)}=0.64,\overline{x}=5.58$，极大似然估计值为

$\hat{\theta} = x_{(i)} = 0.64, \hat{\lambda} = \dfrac{1}{5.58 - 0.64} = 0.2024$。

六、（20 分）设 $X_1, X_2, \cdots, X_n$ 为来自服从 $[0, \theta]$ 上均匀分布总体的样本，记 $Y = \max\left\{X_1, X_2, \cdots, X_n\right\}$。

1.验证 $U = \dfrac{Y}{\theta}$ 的密度函数为

$$f_U(u) = \begin{cases} nu^{n-1}, 0 \le u \le 1 \\ 0, u \notin [0,1] \end{cases}$$ 。

2. 验证 $P\left\{\left(\dfrac{\alpha}{2}\right)^{\frac{1}{n}} < \dfrac{Y}{\theta} \le \left(1 - \dfrac{\alpha}{2}\right)^{\frac{1}{n}}\right\} = 1 - \alpha$，并利用这个结论给出 θ 的 $1-\alpha$ 水平置信区间估计。

3.验证 $P\left\{\alpha^{\frac{1}{n}} < \dfrac{Y}{\theta} \le 1\right\} = 1 - \alpha$ 并利用这个结论给出 θ 的 $1-\alpha$ 水平置信区间估计。

4.上述两个区间估计哪个更好？若有一组5个观测值中的最大值为4.2，计算上述两个95%置信区间。

【解题过程】1. 考查连续型随机变量的函数的分布。即设随机变量 X 具有概率密度 $f_X(x)$，$-\infty < x < \infty$，又设函数 $g(x)$ 处处可导且恒有 $g(x)' > 0$（或恒有 $g(x)' < 0$），则 $Y = g(X)$ 是连续型随机变量，其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)]\left|h'(y)\right|, & \alpha < y < \beta \\ 0, 其他 \end{cases},$$

其中 $\alpha = \min\{g(-\infty), g(\infty)\}$，

$\beta = \max\{g(-\infty), g(\infty)\}$，$h(y)$ 是 $g(x)$ 的反函数。2.3.4.考查中心极限定理及置信区间估计。

【解题过程】1.证明：由于

$$Y=\max\left\{X_1,X_2,\cdots,X_n\right\},$$

$X_1,X_2,\cdots,X_n\sim U\left[0,\theta\right]$,故

$$F_Y(y)=\left[F_X\left(y\right)\right]^n=\begin{cases}0,y<0\\ \left(\dfrac{y}{\theta}\right)^n,0\le y\le\theta\\ 1,y>\theta\end{cases},$$

所以 $f_Y(y)=\begin{cases}n\left(\dfrac{y}{\theta}\right)^{n-1},0\le y\le\theta\\ 0,其他\end{cases}$,

由 $u=\dfrac{y}{\theta}$ 得 $u'=\dfrac{1}{\theta}>0$，且 $y=\theta u$,

$\dfrac{dy}{du}=\theta$ ，所以 $f_U(u)=\begin{cases}nu^{n-1},0\le u\le 1\\ 0,u\notin\left[0,1\right]\end{cases}$。

2.证明：由于 $f_U(u)=\begin{cases} nu^{n-1}, 0\le u\le 1 \\ 0, u\notin[0,1] \end{cases}$ 所以

$$P\left\{\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}<\frac{Y}{\theta}\le\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}}\right\}$$

$$=\int_{\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}}^{\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}}} nx^{n-1}dx=1-\alpha$$

又 $\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}<\frac{Y}{\theta}\le\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}}$

即 $Y\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}}\le\theta<Y\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}$

故 θ 的 $1-\alpha$ 水平置信区间为

$$\left[Y\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}},Y\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}\right]。$$

3. 由于 $f_U(u)=\begin{cases} nu^{n-1}, 0 \le u \le 1 \\ 0, u \notin [0,1] \end{cases}$ 所以

$P\left\{\alpha^{\frac{1}{n}} < \frac{Y}{\theta} \le 1\right\} = \int_{\alpha^{\frac{1}{n}}}^{1} nx^{n-1} dx = 1-\alpha$，而

$\alpha^{\frac{1}{n}} < \frac{Y}{\theta} \le 1$ 即 $Y \le \theta < Y\alpha^{\frac{1}{n}}$ 故 θ 的 $1-\alpha$ 水平置信区间为 $\left[Y, Y\alpha^{\frac{1}{n}}\right]$。

4.前一个区间估计相对更好一些，因为前者取得的是双侧置信限，而后者取得的则是单侧置信限。若一组 5 个观测值中的最大者为 4.2，$\alpha = 0.05$，则 θ 的水平置信区间分别为

$$\left[Y\left(1-\frac{\alpha}{2}\right)^{\frac{1}{n}}, Y\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}\right]$$
$$=\left[4.2\times 0.975^{-0.2}, 4.2\times 0.025^{-0.2}\right]$$
$$=[4.22, 8.78]$$

$$\left[Y, Y\alpha^{-\frac{1}{n}}\right]=\left[4.2, 4.2\times 0.05^{-0.2}\right]=\left[4.2, 7.64\right]$$

七、（10 分）简述假设检验中的两类错误及犯两类错误的概率之间的关系。

【解题过程】考查假设检验中的两类错误及其概率之间的关系的理解。

【解题过程】假设检验中的第一类错误是指当原假设为真时，作出拒绝原假设 H_0 这一错误判断，也称之为“拒真”；第二类错误则是指当原假设实际上为假时，作出接受原假设 H_0 这一错误判断，也称之为“采伪”。当样本容量固定时，若减少犯第一类错误的概率则犯第二类错误的概率往往增大，若要使犯两类错误的概率都减小，除非增加样本容量。