《 概率论与数理统计》期末考试参考答案（A）

一、填空题〖每小题 5 分，共计 25 分〗

1、设 A,B,C 是三个随机事件，则恰有一个发生的事件可表示为 $A\overline{B}\overline{C}+\overline{A}B\overline{C}+\overline{A}\overline{B}C$ 。

2、设随机变量 X 与 Y 相互独立，且 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$。对不全为零的常数 a 和 b，则 $aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2)$。

3、某工厂在含碳量与合金强度关系时，选取 12 个生产小时作样本，测得数据如下：

$$\sum_{i=1}^{12}x_i=1.9,\sum_{i=1}^{12}y_i=589.5,\sum_{i=1}^{12}{x_i}^2=0.3194,$$

$$\sum_{i=1}^{12}y_i^2=29304.25,\sum_{i=1}^{12}x_iy_i=95.85\text{。}$$

假定合金强度 y 与含碳量 x 间具有近似线性关系，则 y 对 x 的线性回归方程 $\widehat{y}=27.6992+135.32x$ 。

4、设总体 X 的概率密度为 $p(x;\theta)=\begin{cases}\sqrt{\theta}x^{\sqrt{\theta}-1}, & 0<x<1\\ 0, & \text{其它}\end{cases}$，其中 $\theta>0$ 是未知参数，$X_1,\cdots,X_n$ 是来自总体 X 的简单随机样本，则 θ 的最大似然估计量

$\hat{\theta}_{\mathrm{MLE}}=\left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right)^{-2}$ **或** $\hat{\theta}_{\mathrm{MLE}}=\left(\overline{\ln x}\right)^{-2}$ 。

5、概率论与数理统计的基本研究对象是 随机事件 **；研究手段是** 随机试验 。

二、〖计 10 分〗设罐中有 b 只黑球、r 只红球，每次随机取出一只球，取出后将原球放回，再加入同色球 c 只。若连续从罐中取球三次，试求所取出的 3 只球是“2 只红球、1 只黑球”事件的概率。

解：记 A 为取得的是“2 只红球、1 只黑球”事件，B_k 和 R_k 表示第 k 次取到红球和黑球的事件。于是，$A=B_1R_2R_3+R_1B_2R_3+R_1R_2B_3$。进一步有加法公式知：$P(A)=P(B_1R_2R_3)+P(R_1B_2R_3)+P(R_1R_2B_3)$。有乘法公式可得：

$$P(B_1R_2R_3)=P(B_1)\cdot P(R_2\mid B_1)\cdot P(R_3\mid B_1R_2)$$

$$=\frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}\text{。}$$

同理可求：$P(R_1B_2R_3)=P(R_1R_2B_3)=P(B_1R_2R_3)$。即

$$P(A)=\frac{3br(r+c)}{(b+r)(b+r+c)(b+r+2c)}。$$

三、〖计 15 分〗设二维随机变量(X,Y)的联合概率密度函数是

$$p(x,y)=\begin{cases}kx^2y, & |x|\le y\le 1\\ 0, & 其它\end{cases}$$

其中k为常数，且$k>0$，试求：

（1）常数k；〖5 分〗

（2）判断随机变量X 和Y是否独立;〖5 分〗

（3）设$Z=X+Y$，求Z的概率密度函数$p_Z(z)$。〖5 分〗

解：(1) 由$1=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}p(x,y)dy=k\int_{-1}^{1}x^2dx\int_{|x|}^{1}ydy=\frac{2k}{15}$得：$k=\frac{15}{2}$。

（2）$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy$

$$=\begin{cases}\int_{|x|}^{1}\frac{15}{2}\cdot x^2ydy=\frac{15}{4}x^2(1-x^2), & -1\le x\le 1\\ 0, & 其它\end{cases},$$

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx=\begin{cases}\int_{-y}^{y}\frac{15}{2}\cdot x^2ydx=5y^4, & 0\le y\le 1\\ 0, & 其它\end{cases}。$$

因$p_X(x)\cdot p_Y(y)\neq p(x,y)$，则随机变量$X$ 与Y不是相互独立的。

（3）由$p_Z(z)=\int_{-\infty}^{+\infty}p(x,z-x)dx$

$$=\begin{cases}\int_{z-1}^{\frac{z}{2}}\frac{15}{2}\cdot x^2\cdot(z-x)dx=\frac{-55z^4+480z^2-640z+240}{128}, & 0\le z\le 2\\ 0, & 其它\end{cases}。$$

四、〖10 分〗设$x_1,\cdots,x_n,x_{n+1}$来自总体$X\sim N(\mu,\sigma^2)$的样本，且

$$\overline{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i，\quad s_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x}_n)^2$$

试求常数C使得$t_c=C\frac{x_{n+1}-\overline{x}_n}{s_n}$服从$t$分布，并指出分布的自由度。

解：因 $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$，$X_{n+1} \sim N(\mu, \sigma^2)$，有 $X_{n+1}-\overline{X}_n \sim N(0, \sigma^2+\frac{\sigma^2}{n})$，

即 $\dfrac{X_{n+1}-\overline{X}_n}{\sigma\sqrt{\frac{n+1}{n}}} \sim N(0,1)$，又因 $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$，且与 $X_{n+1}-\overline{X}_n$ 相互独立，

则由 t 分布定义知
$$\frac{\dfrac{X_{n+1}-\overline{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}}{\sqrt{\dfrac{(n-1)S_n^2}{\sigma^2}\Big/(n-1)}} = \sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\overline{X}_n}{S_n} \sim t(n-1),$$

故当 $c=\sqrt{\frac{n}{n+1}}$ 时，$c\dfrac{X_{n+1}-\overline{X}_n}{S_n}$ 服从自由度为 $n-1$ 的 t 分布.

五、〖计 14 分〗为比较两个小麦品种的产量，选择 18 块条件相似的试验田，采用相同的耕作方法做试验，结果播种甲品种的 8 块试验田的平均产量 $\overline{x}_{甲}=569.38$（单位：Kg）和样本方差 $s_{甲}^2=2140.55$，结果播种乙品种的 10 块试验田的平均产量 $\overline{x}_{乙}=487.00$（单位：Kg）和样本方差 $s_{乙}^2=3256.22$，假定每个品种的单位面积产量均服从正态分布，且两个品种单位面积产量的标准差相等。试求

（1）这两个品种平均单位面积产量差的置信区间。〖7 分〗

（2）并判断这两个小麦品种的产量有无显著差异？(取 $\alpha=0.05$) 〖7 分〗

解：（1）已知两个品种单位面积产量的标准差相等，则采用二样本 t 区间。

$$s_w=\sqrt{\frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-2}}=\sqrt{\frac{7\times2140.55+9\times3256.22}{16}}=52.6129$$

$$t_{1-\alpha/2}(m+n-2)=t_{0.975}(16)=2.1199$$

$$t_{1-\alpha/2}(m+n-2)s_w\sqrt{\frac{1}{m}+\frac{1}{n}}=2.1199\times52.6129\times\sqrt{\frac{1}{8}+\frac{1}{10}}=52.91$$

故 μ_1-μ_2 的 0.95 置信区间为：

$$\overline{x}-\overline{y}\pm t_{1-\alpha/2}(m+n-2)s_w\sqrt{\frac{1}{m}+\frac{1}{n}}=569.38-487\pm52.91=[29.47,135.29]$$

（2）双侧假设检验问题

原假设 $H_0: \mu_1 = \mu_2$;　　$H_1: \mu_1 \neq \mu_2$;

在$\sigma_1^2 = \sigma_2^2 = \sigma^2$但未知时，选择$t$检验统计量：$t = \dfrac{\bar{x} - \bar{y}}{s_w\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}$

取$\alpha = 0.05$，检验的拒绝域$W = \{|t| \geq t_{1-\frac{\alpha}{2}}(m+n-2)\} = \{|t| \geq 2.1199\}$

经过计算$t = \dfrac{\bar{x} - \bar{y}}{s_w\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}} = \dfrac{569.38 - 487}{52.6129\sqrt{\dfrac{1}{8} + \dfrac{1}{10}}} = 3.301$

由于$|t| \geq 2.1199$，落入拒绝域，故拒绝原假设，两个小麦品种的产量有显著差异。

六、〖计 10 分〗设从均值为μ，方差为σ^2（>0）的总体中分别抽取容量为n_1和 n_2的两独立样本，$\bar{x}_1$和 $\bar{x}_2$分别是这两个样本的均值。试证，对于任意常数 a,b（$a+b=1$），$Y = a\bar{x}_1 + b\bar{x}_2$ 都是μ 的无偏估计。并确定常 数a,b使$\mathrm{Var}(Y)$ 达到最小。

解：因$E(Y) = aE(\bar{X}_1) + bE(\bar{X}_2) = a\mu + b\mu = (a+b)\mu = \mu$，

故Y是μ的无偏估计；

因$\mathrm{Var}(Y) = a^2\,\mathrm{Var}(\bar{X}_1) + b^2\,\mathrm{Var}(\bar{X}_2) = a^2 \cdot \dfrac{\sigma^2}{n_1} + (1-a)^2 \cdot \dfrac{\sigma^2}{n_2} = \left(\dfrac{n_1+n_2}{n_1 n_2}a^2 - \dfrac{2}{n_2}a + \dfrac{1}{n_2}\right)\sigma^2$，

令$\dfrac{d}{da}\mathrm{Var}(Y) = \left(\dfrac{n_1+n_2}{n_1 n_2} \cdot 2a - \dfrac{2}{n_2}\right)\sigma^2 = 0$，得$a = \dfrac{n_1}{n_1+n_2}$，且$\dfrac{d^2}{d^2a}\mathrm{Var}(Y) = \dfrac{n_1+n_2}{n_1 n_2} \cdot 2\sigma^2 > 0$，

故当$a = \dfrac{n_1}{n_1+n_2}$，$b = 1 - a = \dfrac{n_2}{n_1+n_2}$时，$\mathrm{Var}(Y)$达到最小$\dfrac{1}{n_1+n_2}\sigma^2$.

七、〖计 10 分〗设随机变量X的期望μ和方差σ^2（>0）均存在。$x_1, x_2, \cdots, x_n$是来自总体X的一个样本，试利用特征函数证明：

$$Y_n = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \xrightarrow{L} N(0,1)\text{。}$$

其中“$\xrightarrow{L}$”表示“依分布收敛”。

证明：记$X - \mu$的特征函数为$\varphi(t)$，则由特征函数性质可知：$\varphi'(0) = 0$，$\varphi''(0) = -\sigma^2$。

由$Y_n=\dfrac{\sqrt{n}(\bar{x}-\mu)}{\sigma}=\sum\limits_{i=1}^{n}\dfrac{x_i-\mu}{\sigma\sqrt{n}}$，可得$Y_n$的特征函数为

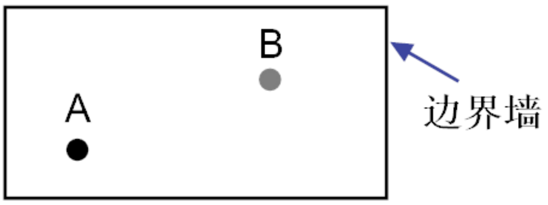
$$\varphi_{Y_n}(t)=\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right)=\left[\left(1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)^{-\frac{2n}{t^2}}\right]^{-\frac{t^2}{2}}$$

易得：$\lim\limits_{n\to\infty}\varphi_{Y_n}(t)=e^{-\frac{t^2}{2}}$。即当$n\to\infty$时，随机变量$Y_n$依分布收敛于标准正态分布。

八、〖计 6 分〗（A）和（B）两题中请任选一题作答。

（A）请比较概率论中的“概率的公理化定义”和假设检验中的“假设（hypothesis）”所体现的思想性异同，并谈谈它们在学科体系架构和科学研究中的地位。

（B）在一个长为 2a、宽为 a 的平面区域内，有两个醉鬼。假设两个醉鬼前进的方向和速度随时都会发生改变，且他们在遇到边界墙后会随机选一个方向继续前进。请设计程序（画出程序流程图即可），模拟醉鬼的运动轨迹，直至两个醉鬼相撞。



答：(A)要点：1) 公理是不能被证伪、亦不能被更基础的公理或定理等所证明的命题; 2) hypothesis 是基于有限经验的推测; 3) 公理化体系是源自欧几里得几何，是自然科学知识体系的主要架构方式; 4) hypothesis 是现代科学研究的主要抓手，是可以反复升级直至无限逼近真理的。

(B) 需引入方向改变和速度改变的随机变量，且需引入发生改变的时刻的随机变量。需定义醉鬼的所在区域。程序主体为迭代循环法。