清华大学本科生考试试题专用纸 （ A 卷）

考试课程： 概率论与数理统计　　考试时间： 2007 年 1 月 10 日 14:30-16:30

姓名 ________ 学号 200__0________ 班级 __________

标准正态分布 $N(0,1)$ 的分布函数值：	x	1.46	1.645	1.96	2	2.46
	$\Phi(x)$	0.928	0.950	0.975	0.977	0.993

一、填空。（每空 2 分，共 38 分。直接将答案填在划线处。如果最终结果用小数表示，请精确到小数点后第 3 位。）

1. 设 $X \sim U(0,1)$ 。以下判断正确的是 ____________ 。
   A： “事件 $\{X = 0.5\}$ 是不可能事件”； B： “事件 $\{X = 0.5\}$ 是零概率事件”。
2. 设事件 A_1, A_2, A_3 相互独立，且 $P(A_i) = 1/3$（ $i = 1,2,3$ ）。则这三个事件中至少有一个发生的概率为 _______ ；这三个事件中恰好有一个发生的概率为 _______ 。
3. 工厂 A 和工厂 B 的产品次品率分别是 1% 和 2% 。现从 A 、B 两厂产品各占 60% 和 40% 的一批产品中随机选取一件，则该产品是次品的概率为 _____% 。如果发现这个产品是次品，那么它是工厂 A 生产的概率为 ____________ 。
4. 设随机变量 X 和 Y 均服从 $N(0,1)$, 则以下判断正确的是 ___________ 。
   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 都服从 χ^2 分布；　　D： X^2/Y^2 服从 F 分布。
5. 某人有 n 把钥匙，其中只有一把能打开自己的家门，他随意地试开。如果他每次把试过的钥匙又混杂进去，则打开门所需试开次数 X 的数学期望为 ______ 。如果他每次把不能打开的钥匙剔除，则打开门所需试开次数 Y 的数学期望为 __________ 。
6. 现有一个容量为 9 、来自正态总体 $N(\mu, 0.81)$ 的简单随机样本，其样本均值为 5 ，则未知参数 μ 的置信水平为 0.95 的置信区间为 _____________________ 。
7. 设 $x_1, \ldots, x_4$ 为来自正态总体 $N(\mu, 4^2)$ 的简单随机样本，$\overline{x}$ 为样本均值. 在显著性水平 $\alpha = 0.05$ 下考虑检验问题： $H_0: \mu = 5$ vs $H_1: \mu \neq 5$. 其拒绝域为 $\{|\overline{x} - 5| \geq k\}$, 则 $k =$_______ 。该检验问题在真值 $\mu = 6$ 时犯第二类错误的概率为 _______ 。
8. 设 X, Y 相互独立， $P(X=1) = P(X=2) = 1/2$ ， Y 服从指数分布 $\mathrm{Exp}(1)$ 。则 XY 的分布函数 $F(z) =$______________ ，密度函数 $f(z) =$______________ 。
9. 设 $X_1, \ldots, X_{100}$ 是来自正态总体 $N(\mu, \mu^2)$（ $\mu > 0$ ）的简单随机样本，样本均值为 $\bar{X}$ 。则 $10(\mu^{-1}\bar{X} - 1)$ 服从 _________________（分布名称及参数）。如果正数 C 满足 $P(\bar{X}/C \geq \mu) = 5\%$ ，则 $C =$________ 。
10. 设 $g: [0,1] \to \mathbb{R}$ 连续，记 $A = \int_0^1 g(x)dx$ ， $B = \int_0^1 [g(x)]^2 dx$ ， $B \neq A^2$ 。设 $X_1, X_2, \ldots$ 独立，都服从均匀分布 $U[0,1]$ 。则当 $n \to \infty$ 时， $Q_n := \frac{1}{n}\sum_{k=1}^{n} g(X_k)$ 依概率收敛于 ______ 。当 n 充分大时，$\sqrt{n}(Q_n - A)$ 的近似分布为 ________（分布名称及参数）。

11. 设 X 的概率密度为 $f(x) = C/x^2$ ， $|x| \geq 1$ 。则常数 $C =$____ 。以下判断中正确的是 ______ 。
    A： X 的数学期望存在且等于零； B： X 不存在数学期望。

二、（ 8 分）一生产线生产的产品成箱包装，每箱的重量是随机的。设每箱的平均重量为 50 公斤，标准差为 5 公斤。若用最大载重量为 5000 公斤的汽车承运，试用中心极限定理说明每辆车最多可以装多少箱，才能保证不超载的概率大于 0.977 ？

三、（ 16 分, 每问 4 分 ）独立重复掷一枚硬币, 每次抛掷得到正面的概率为 $0 < p < 1$ ，到第 X 次掷币时首次得到正面，到第 Y 次掷币时刚好得到累计两次正面。

1. 求 X, Y 的联合概率分布列以及边缘概率分布列；
2. 求在已知 $X = k$ 的条件下， $Y - X$ 的条件分布列；并判断 $X, Y - X$ 是否独立；
3. 求 EX 和 EY ；
4. 求 X, Y 的相关系数。

四、（ 18 分，每问 3 分 ）设二维随机向量 (X, Y) 的联合概率密度函数为

$$f(x,y) = \begin{cases} C(x+y), & \text{若} 0 < x < 1 \text{且} 0 < y < 1; \\ 0, & \text{否则}. \end{cases}$$

1. 求 C 的值；
2. 求 X 和 Y 的边缘概率密度函数，并判断 X, Y 是否独立；
3. 求 $E(X^m Y^n)$ ，其中 m, n 是非负整数；
4. 求使得 $E(Y - aX)^2$ 达到最小值时的 a 值；
5. 求条件期望 $E(Y|X)$ ；
6. 求 $U = X + Y, V = X - Y$ 的联合概率密度函数。

五、（ 20 分，每问 4 分 ）设 $X_1, \ldots, X_n$ 是来自总体 $U(-\sqrt{\theta}, \sqrt{\theta})$ 的简单随机样本，其中 $\theta > 0$ 是未知参数。

1. 求 θ 的矩估计量 $\hat{\theta}_1$ ，以及当样本容量 $n \to \infty$ 时 $\hat{\theta}_1$ 的渐近分布；
2. 求 θ 的极大似然估计量 $\hat{\theta}_2$ ；
3. 求常数 c_1, c_2 使得 $\eta_1 = c_1\hat{\theta}_1$ 和 $\eta_2 = c_2\hat{\theta}_2$ 均为 θ 的无偏估计；
4. 上述两个无偏估计 η_1 和 η_2 中，哪一个更有效性？
5. 问： $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是不是 θ 的相合估计？为什么？

2006-2007 学年秋季学期概率论与数理统计
试题答案和评分标准

AB 一、填空。

题号	答案	题号	答案
A1	B	B1	B
A2	19/27， 4/9	B2	19/27， 4/9
A3	1.4， 3/7	B3	1.4， 3/7
A4	C	B4	3.92， 0.921
A5	n， $(n+1)/2$	B5	$1-(e^{-z}+e^{-z/2})/2, z>0$， $(e^{-z}+e^{-z/2})/2, z>0$
A6	$[4.412, 5.588]$	B6	$N(0,1)$， 1.1645
A7	3.92， 0.921	B7	C
A8	$1-(e^{-z}+e^{-z/2})/2, z>0$， $(e^{-z}+e^{-z/2})/2, z>0$	B8	n， $(n+1)/2$
A9	$N(0,1)$， 1.1645	B9	$[4.412, 5.588]$
A10	A， $N(0, B-A^2)$	B10	A， $N(0, B-A^2)$
A11	$1/2$， B	B11	$1/2$， B

AB 二 设 $X_1, \ldots, X_n$ 是各箱重量，它们独立同分布，不超载的概率为

$$P(X_1+\cdots+X_n \le 5000) \ge 0.977 = \Phi(2).$$

由中心极限定理知对充分大的 n，

$$P\left(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \le \frac{5000-50n}{5\sqrt{n}}\right) \approx \Phi\left(\frac{5000-50n}{5\sqrt{n}}\right),$$

其中， $\mu = EX_1 = 50$ 公斤， $\sigma = \sqrt{DX_1} = 5$ 公斤。因此，当

$$\frac{5000-50n}{5\sqrt{n}} > 2,$$

即

$$25n^2 - 5001n + 25\times 10^4 > 0, \qquad 5000-500n > 0,$$

即

$$n < \frac{5001-\sqrt{5001^2-4\times 25^2\times 10^4}}{50} = \frac{5001-\sqrt{5001+5000}}{50} \approx \frac{5001-100}{50} = 98.02,$$

即 $n \le 98$ 时，不超载的概率大于 0.977 。

A 三 B 四 (1) X 服从几何分布 $G(p)$，分布列为

$$P(X=k) = pq^{k-1}, \qquad k=1,2,\ldots.$$

Y 的概率分布列为

$$P(Y=n) = C_{n-1}^1 p^2 q^{n-2} = (n-1)p^2q^{n-2}, \qquad n=2,3,\ldots.$$

联合概率分布为

$$P(X=k, Y=n) = p^2q^{n-2}, \qquad k=1,2,\ldots,n-1; n=2,3,\ldots.$$

(2)

$$P(Y-X=n|X=k) = \frac{P(X=k, Y=n+k)}{P(X=k)} = \frac{p^2q^{n+k-2}}{pq^{k-1}} = pq^{n-1}.$$

即在已知 $X=k$ 发生的条件下， $Y-X$ 服从几何分布 $G(p)$ 。

(3) 由 (2) 知， $X=k$ 时 $Y-X$ 的条件分布与 k 无关，所以 $Y-X$ 与 X 独立。

$$P(X=k, Y-X=n) = P(X=k, Y=n+k) = p^2q^{n+k-2} = pq^{k-1}\cdot pq^{n-1}, \quad \forall k,n\ge 1,$$

所以 X 与 $Y-X$ 独立。

(4)

$$EX = \sum_{k=1}^{\infty} kpq^{k-1} = \frac{1}{p},$$
$$EY = EX + E(Y-X) = \frac{2}{p}.$$

(5)

$$DX = EX(X-1) + EX - (EX)^2 = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} + \frac{1}{p} - \frac{1}{p^2}$$
$$= \frac{2pq}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2},$$
$$DY = D[X+(Y-X)] = DX + D(Y-X) = \frac{2q}{p^2},$$
$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,X) + \mathrm{Cov}(X, Y-X) = DX = \frac{q}{p^2},$$

于是

$$r_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{DX\cdot DY}} = \frac{1}{\sqrt{2}}.$$

A 四 B 三 (1)

$$1 = \int_0^1 \int_0^1 C(x+y)dxdy = 2C\int_0^1 xdx \int_0^1 dy = C.$$

(2)

$$f_X(x) = \int_0^1 C(x+y)dy = C\left(x+\frac{1}{2}\right) = x+\frac{1}{2}, \quad 0<x<1.$$

同理,

$$f_Y(y) = y+\frac{1}{2}, \quad 0<y<1,$$

X,Y 不独立。

(3)

$$E(X^mY^n) = \int_0^1\int_0^1 x^my^nf(x,y)dxdy = \int_0^1\int_0^1 x^{m+1}y^n + x^my^{n+1}dxdy$$
$$= \frac{1}{(m+2)(n+1)} + \frac{1}{(m+1)(n+2)} = \frac{2mn+3(m+n)+4}{(m+1)(m+2)(n+1)(n+2)},$$

(4) 由 (3) 知 $EX^2 = EY^2 = \frac{1}{4}+\frac{1}{6} = \frac{5}{12}$， $E(XY) = \frac{2}{6} = \frac{1}{3}$，

$$E(Y-aX)^2 = a^2EX^2 - 2aE(XY) + EY^2 = \frac{5}{12}a^2 - \frac{2}{3}a + \frac{5}{12} = \frac{5}{12}(a-4/5)^2 + \frac{3}{20}$$

在 $a=4/5$ 时达到最小值 $3/20$ 。

(5) 条件密度函数

$$f_{Y|X}(y|x) = \frac{x+y}{x+1/2}, \qquad 0<y<1, 0<x<1.$$

因此条件期望

$$E(Y|X=x) = \int_0^1 yf_{Y|X}(y|x)dy = \frac{x}{2x+1} + \frac{1}{3x+3/2} = \frac{3x+2}{6x+3}.$$

从而

$$E(Y|X) = \frac{3X+2}{6X+3}.$$

(6)

$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right)\frac{1}{2} = \frac{u}{2}I_{0<u+v<2, 0<u-v<2}.$$

AB 五 (1) 由

$$EX=0, \quad EX^2 = DX = \frac{(2\sqrt{\theta})^2}{12} = \frac{\theta}{3},$$

得到 θ 的矩估计为

$$\hat\theta_1 = \frac{3}{n}\sum_{i=1}^n X_i^2,$$

再由

$$EX^4 = \int_{-\sqrt\theta}^{\sqrt\theta} \frac{x^4}{2\sqrt\theta}dx = \frac{\theta^2}{5}, \qquad D(X^2) = \frac{\theta^2}{5} - \frac{\theta^2}{9} = \frac{4\theta^2}{45},$$

及中心极限定理，得到渐近分布 $N\left(\theta, \frac{4\theta^2}{5n}\right)$ 。

(2) 似然函数

$$L(\theta) = \prod_{i=1}^n \frac{1}{2\sqrt\theta}I_{|x_i|<\sqrt\theta} = \frac{1}{2^n\theta^{n/2}}I_{\max\limits_{1\le i\le n} x_i^2<\theta},$$

因其单调性，在 $\max\limits_{1\le i\le n} x_i^2$ 处取最大值，因此

$$\hat\theta_2 = \max_{1\le i\le n} X_i^2$$

是 θ 的极大似然估计。

(3)

$$F_{\hat\theta_2/\theta}(t) = P(\max_{1\le i\le n} X_i^2 \le t\theta) = \left[P(|X_1|\le\sqrt{t\theta})\right]^n = \left(\frac{2\sqrt{t\theta}}{2\sqrt\theta}\right)^n = t^{n/2}, \quad 0<t<1;$$
$$f_{\hat\theta_2/\theta}(t) = \frac{n}{2}t^{\frac{n}{2}-1}, \quad 0<t<1.$$

(4) 由

$$E\hat\theta_1 = 3EX^2 = \theta, \quad E\hat\theta_2 = \theta\int_0^1 tf_{\hat\theta_2/\theta}(t)dt = \frac{n\theta}{n+2}\int_0^1 \frac{n+2}{2}t^{\frac{n+2}{2}-1}dt = \frac{n\theta}{n+2},$$

得 θ 的两个无偏估计： $\hat\theta_1$ 和 $\frac{n+2}{n}\hat\theta_2$ ，它们的方差为

$$D\hat\theta_1 = E(\hat\theta_1-\theta)^2 = \frac{9}{n}D(X^2) = \frac{4\theta^2}{5n}.$$
$$D\left(\frac{n+2}{n}\hat\theta_2\right) = \theta^2\int_0^1\left(\frac{(n+2)}{n}t-1\right)^2\frac{n}{2}t^{\frac{n}{2}-1}dt$$
$$= \theta^2\left(\frac{(n+2)^2}{n(n+4)}\int_0^1\frac{n+4}{2}t^{\frac{n+4}{2}-1}dt - 2\int_0^1\frac{n+2}{2}t^{\frac{n+2}{2}-1}dt + \int_0^1\frac{n}{2}t^{\frac{n}{2}-1}dt\right)$$
$$= \theta^2\left(\frac{(n+2)^2}{n(n+4)} - 1\right) = \frac{4\theta^2}{n(n+4)},$$

后者比前者有效。

(5) 根据大数定律或 (1) 中的渐近分布，$\hat{\theta}_1$ 是 θ 的相合估计。

由

$$P(|\hat{\theta}_2-\theta|\le\varepsilon)=P(\theta-\varepsilon<\hat{\theta}_2<\theta)=\int_{\theta-\varepsilon}^{\theta}\frac{n}{2\theta}\left(\frac{t}{\theta}\right)^{\frac{n}{2}-1}dt=1-\left(\frac{\theta-\varepsilon}{\theta}\right)^{\frac{n}{2}}\xrightarrow[n\to\infty]{}1$$

知 $\hat{\theta}_2$ 是 θ 的相合估计。

也可以这样：当 $n>4\theta/\varepsilon-2$ 时

$$|E\hat{\theta}_2-\theta|=\frac{2\theta}{n+2}<\frac{\varepsilon}{2},$$

于是

$$P(|\hat{\theta}_2-\theta|>\varepsilon)\le P(|\hat{\theta}_2-E\hat{\theta}_2|>\frac{\varepsilon}{2})\le\frac{4}{\varepsilon^2}D\hat{\theta}_2,$$

而

$$D\hat{\theta}_2=E(\hat{\theta}_2-\theta)^2-(E\hat{\theta}_2-\theta)^2=\frac{4n\theta^2}{(n+4)(n+2)^2}\xrightarrow[n\to\infty]{}0,$$

所以 $\hat{\theta}_2$ 是 θ 的相合估计。