# 中国人民大学硕士研究生入学考试
# 概率论与数理统计专业试题集*

## I. 2007年试题

一、（20分）两个不能分辨的盒子里都有 9 个球，其中一个是 5 红 4 白，另一个是 4 红 5 白。从两个盒子中随机抽一个，希望通过无放回抽样来猜测抽到底是哪个盒子。其规则是：无放回抽取三次，如果抽到的红球多，则认为盒子是 5 红 4 白；反之认为是 4 红 5 白。问这样猜错的概率有多大？如果用有放回抽样，猜错的概率又有多少？

二、（20分）相互独立的随机变量 $X$ 和 $Y$ 分别服从参数为 $\lambda_1$ 和 $\lambda_2$ 的泊松分布，证明随机变量 $X+Y$ 服从参数为 $\lambda_1+\lambda_2$ 的泊松分布。要求用两种方法证明，其中一种是特征函数。

三、（10分）二维随机变量 $(X,Y)$ 的联合概率密度为

$$f(x,y)=\begin{cases} cxy, & 0<x<2, 1<y<2 \\ 0, & \text{其他} \end{cases}$$

求 $Z=\min(X,Y)$ 的概率密度函数。

四、（15分）设随机变量序列 $\{\xi_n\}$ 及 $\{\eta_n\}$ 分别以概率收敛于随机变量 $\xi$ 和 $\eta$，证明：$\{\xi_n+\eta_n\}$ 以概率收敛于 $\xi+\eta$。

五、（15分）二维随机变量 $(X,Y)$ 的联合分布律为

| $X \backslash Y$ | 0 | 1 | 2 |
|---|---|---|---|
| $-1$ | 0.1 | 0.2 | 0 |
| 0 | 0 | 0.1 | 0.2 |
| 1 | 0.1 | 0.2 | 0.1 |

*没有答案，自己做出来才能验证自己是不是真的掌握了

求 $\mathrm{E}[Y|X]$ 和 $\mathrm{Var}[X|Y]$ 的分布律。

六、（20分）设 $X_1$，$X_2$，...，$X_n$是来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本，证明：

1. $\dfrac{1}{\sigma^2}\sum_{t=1}^{n}(X_t-\overline{X})\sim\chi^2(n-1)$
2. $\overline{X}$ 与 $S_n^2$ 相互独立。

七、（15分）设总体 $X$ 的分布函数为 $F(x)$，概率密度函数为 $f(x)$，$X_1$，$X_2$，...，$X_n$是总体 $X$ 的简单随机样本，证明第 $k$ 个次序统计量 $X_{(k)}$ 的概率密度函数为

$$f_k(x)=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)，k=1,2,\ldots,n$$

八、（20分）设总体 $X$ 服从正态总体 $N(\theta,\sigma^2)$ ，其中 $\sigma^2$ 已知。参数 $\theta$ 的先验分布为正态总体 $N(\mu,\tau^2)$ ，其中 $\mu$ 和 $\tau^2$ 已知。$X_1$，$X_2$，...，$X_n$是总体 $X$ 的简单随机样本，求

1. 参数 $\theta$ 的后验分布；
2. 在平方损失函数下求 $\theta$ 的贝叶斯估计；
3. 求 $\theta$ 的置信水平为 $1-\alpha$ 的区间估计。

九、（15分）某市作电视收视率调查，随机调查了 100 人，在晚七点二十分收看中央台新闻联播节目的人数是 45 人，根据以往经验，这一时刻节目的收视率是 50%，则

1. 能否认为收视率有了显著减少（显著性水平 $\alpha=0.05$ ）；
2. 假设检验问题与区间估计具有对偶性，给出与以上假设检验问题相对偶的单侧区间估计，并说明如何由此区间估计做上面的假设检验。
（已知 $\Phi(1.960)=0.975$，$\Phi(1.645)=0.95$ ）

## II. 2006年试题

一、（20分）设有编号 $1,\cdots,n$ 的 $n$ 个球和从左到右排列的编号为 $1,\cdots,N$ 的$N$个格子 $(N\geq n)$,每个球都以同样的概率 $1/N$ 落到到 $N$ 个格子中的一个格子中，试求：

1. 某指定的 $n$ 个格子中各有一个球的概率；
2. 任何 $n$ 个格子中各有一个球的概率；
3. 任何 $n$ 个格子中各有一个球并且球号从左到右严格上升的次序排列的概率；
4. 任何 $n$ 个格子中各有一个球并且球号从左到右严格上升的次序排列，同时编号为 $m$ 的球落在第 $M$ 个盒子中的概率 $(m\leq M, n-m\leq N-M)$ 。

二、（20分）在可靠性与生存分析中，所研究的寿命现象是非负随机变量，记作 $\xi$.其分布函数为 $F(x)$，密度函数为 $f(x)$，称 $S(x)=1-F(x)$ 为生存函数，这时常引入失效率函数 $\lambda(x)$，其定义为：$\lambda(x)=\frac{f(x)}{S(x)}$.

1. 给出失效率函数 $\lambda(x)$ 的直观解释，并推导用 $\lambda(x)$ 表示 $S(x)$ 的公式；
2. 并计算以下问题：某放射性物质在初始时刻的质量为 $m_0$ ，在单位时间内每个原子产生分裂核的概率为常数 $\lambda$ ，求经过时间 $x$ 后该放射性物质质量的期望。

三、（20分）设 $\xi_1$ 和 $\xi_2$ 不相关，分别对以下两种情况证明 $\xi_1$ 与 $\xi_2$ 独立:

1. $\xi_1$ 与 $\xi_2$ 都是只取0和1两个值的随机变量；
2. $\xi_1$ 是只取 $a_1$ 和 $b_1$ ( $b_1>a_1$ ) 这两个值的随机变量， $\xi_2$ 是只取 $a_2$ 和 $b_2$ ( $b_2>a_2$ ) 这两个值的随机变量.

四、（15分）一个同学采用如下的方法获得标准正态分布的随机数：每次产生12个(0,1)区间均匀分布的随机数，求和后再减6，作为标准正态分布的一个随机数。你认为该同学的是做法是否合理，说明原因.

五、（20分）假设 $X_1,X_2,\cdots,X_n$ 取自正态分布 $N(\mu,\sigma^2)$ 的简单随机样本， $\mu$ 和 $\sigma^2$ 未知，求

1. $\mu$ 和 $\sigma^2$ 的极大似然估计；
2. 在上述分布下，求用 $S_{n+1}^2=\dfrac{1}{n+1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 估计 $\sigma^2$ 的均方误差，比较 $S_{n+1}^2$ 和上一问中的 $\sigma^2$ 的极大似然估计的均方误差，谁大谁小.

六、（20分）以下是13名大学生刚入校时某项体能测试成绩和入校进行了三个月体育训练后的测试成绩，如下数据所示：

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 入校时： | 42 | 57 | 38 | 49 | 63 | 36 | 48 | 58 | 47 | 51 | 83 | 27 | 31 |
| 训练后： | 40 | 65 | 48 | 37 | 68 | 40 | 50 | 60 | 49 | 58 | 62 | 33 | 44 |

在水平为0.10之下，回答训练前后测试数据的均值是否存在显著性差异？

1. 观察两组数据，讨论所选检验方法的假定条件
2. 用1.中所选用的假定条件，在0.10的显著性水平下，判断训练前后测试数据的均值是否存在显著性差异？

七、（20分）设从均值为 $\mu$，方差为 $\sigma^2 > 0$ 的总体中，分别抽取样本量为 $n_1$ 和 $n_2$ 的两个独立样本，$\bar{X}_1$ 和 $\bar{X}_2$ 分别是两个样本的样本均值。确定常数 $a$ 和 $b$ 使得 $Y = a\bar{X}_1 + b\bar{X}_2$ 为 $\mu$ 的无偏估计且 $\mathrm{Var}(Y)$ 达到最小.

八、（15分）设总体 $X$ 的分布密度为:

$$p(x,\theta) = \begin{cases} \frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}} & x > 0, 0 < \theta < \infty \\ 0 & x \leq 0 \end{cases}$$

$(X_1, \cdots, X_n)$ 为总体 $X$ 的样本，求参数 $\theta$ 的置信度为 $1-\alpha$ 的置信区间.

## III. 2005年试题

一、（20分）证明几何分布是离散随机变量中唯一具有无记忆的分布。

二、（20分）记 $U_1, U_2$ 是相互独立的[0,1]区间均匀分布的随机数，令

$$\xi_1 = (-2\ln U_1)^{\frac{1}{2}}\cos(2\pi U_2)$$

$$\xi_2 = (-2\ln U_1)^{\frac{1}{2}}\sin(2\pi U_2)$$

证明：$\xi_1$ 与 $\xi_2$ 是相互独立的 $N(0,1)$ 随机变量.

三、（20分）设 $\xi_1, \xi_2, \cdots, \xi_n$ 是相互独立随机变量，$\xi_i$ 的方差 $\mathrm{Var}\xi_i = \sigma_i^2$，试找非负实数 $a_1, a_2, \cdots, a_n$（其中 $a_1 + a_2 + \cdots + a_n = 1$），使 $\xi = a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n$ 的方差最小.

四、（15分）设 $\{\xi_k, k\geq 1\}$ 是一列相互独立的随机变量序列，$\xi_k$ 服从区间 $[-k, k]$ 上的均匀分布，用林德贝格条件证明 $\{\xi_k\}$ 服从中心极限定理.

五、（20分）令 $X_1, X_2, \cdots, X_n$ 是从分布族 $p(x,\theta)$ 中抽取的独立同分布随机样本，$p(x,\theta)$ 如下所示：

$$p(x,\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x \leq \theta, 0 < \theta < \infty \\ 0, & \text{其他} \end{cases}$$

求参数 $\theta$ 的极大似然估计 $\hat{\theta}_{\mathrm{MLE}}$，判断 $\hat{\theta}_{\mathrm{MLE}}$ 是否 $\theta$ 的无偏估计，如果是，求出它的方差；如果不是，请构造一个无偏估计，并求出无偏估计的方差，在均方差的标准下说明谁更优。（提示：参数 $\theta$ 的估计量 $T(X_1, X_2, \cdots, X_n)$ 的均方差定义为 $\mathrm{E}(T(X_1, X_2, \cdots, X_n) - \theta)^2$）。

六、（15分）某超市为方便附近居民对某种商品的需求，调查了 100 家住户，得出每户平均需要量 $\overline{X}$ 为 15kg，$S^2$ 为 6.25kg，假如居民对该类商品的月需要量服从正态分布，如果该超市附近有 10000 户居民。

1. （8分）试求一户居民对该种商品的平均月需求量置信水平为0.99的区间估计。
2. （7分）本着节约库存的考虑，至少需要准备多少该类商品才能以0.99的概率满足附近居民的需要。

七、（20分）设 $X_1, \cdots, X_n$ 是取自正态总体 $N(\mu, 1)$ 独立同分布的样本，对假设检验问题$H_0: \mu \leq 0,\ H_1: \mu > 0$.

1. （10分）试给出一个水平为 $\alpha$ 的检验和检验拒绝域（要求：给出求解的全部过程）。

2. （10分）假设 $\mu$ 有先验分布 $N(1,2)$，求 $\mu$ 在平方损失函数下的 Bayes 估计。（提示：$\mu$ 的 Bayes 估计定义为：$\hat{\mu}_B$=E$(\mu|X)$）.

八、（20分）一个社会工作者选取10对夫妻，考察他们对婚姻状况的满意程度，婚姻满意度描述的是每个人在婚姻中的快乐。结果由下表给出：

| | 女性 | | 男性 | |
|---|---|---|---|---|
| 统计量 | 均值 | 标准差 | 均值 | 标准差 |
| 婚姻满意度 | 25.6 | 8.6 | 32.0 | 9.8 |

1. （8分）如果需要分析在婚姻满意状况中，男性的差异和女性的差异是否存在不同，请讨论可以选择怎样的假定和方法进行分析，你对如上的数据汇总方式是否满意？

2. （6分）根据你的假定和选择的方法回答是否可以认为在婚姻满意状况中男性的差异与女性的差异不同？

3. （6分）构造男性和女性对婚姻状况满意度之差的90%置信区间？

# IV. 2004年试题

一、 （15分）袋中有 $m+n$ 枚同型号硬币，$m$ 枚是正品，$n$ 枚是次品，次品的两面都是国徽。从袋中任取一枚，将它抛掷 $r$ 次，每次都出现国徽，求这枚硬币是正品的概率。

二、 (20分）设一个家庭有 $n$ 个小孩的概率为

$$P_n = \begin{cases} \alpha p^n, & n \geq 1 \\ 1 - \dfrac{\alpha p}{1-p}, & n = 0 \end{cases}$$

这里 $0 < p < 1$, $0<\alpha<\dfrac{1-p}{p}$，若认为生一个小孩为男孩或女孩是等可能的，求证一个家庭有 $k\,(k \geq 1)$ 个男孩的概率为

$$\frac{2\alpha p^k}{(2-p)^{k+1}}$$

三、 （20分）设二维随机变量 $(\xi, \eta)$ 的联合概率密度函数为

$$f(x,y) = \begin{cases} 8xy, & 0 \leq x \leq y \leq 1 \\ 0, & \text{其它} \end{cases}$$

求条件数学期望 $\mathrm{E}(\eta|\xi = x)$ 和条件方差 $\mathrm{Var}(\eta|\xi = x)$ .

四、 （20分）（伯恩斯坦定理）已知随机变量序列 $\{\xi_n, n\geq 1\}$ 的方差有界：$\mathrm{Var}(\xi_i) \leq C, (i = 1, 2, \cdots)$，并且当 $|i-j| \to \infty$ 时，$\xi_i$ 和 $\xi_j$ 的相关系数 $r_{ij} \to 0$ ，证明对 $\{\xi_n, n\geq 1\}$ 成立大数定律.

五、 （20分）假定钢铁制造厂A生产的钢材的强度服从 $N(\mu_1, \sigma_1^2)$ ，从中获得容量为16的样本，测定其强度，得到 $\overline{X}$=1190, $S_x^2$=$90^2$ （样本无偏方差）。钢铁制造厂B生产的同种钢材的强度服从 $N(\mu_2, \sigma_2^2)$ , 从中抽取容量为13的样本，测定其强度，得 $\bar{Y} = 1190, S_y^2 = 100^2$ .

1. （6分）求 $\sigma_1/\sigma_2$ 的置信水平为0.95的置信区间；
2. （7分）由上述置信区间是否可以假定 $\sigma_1 = \sigma_2$ ，请指出这样做的理由？
3. （7分）在 $\sigma_1 = \sigma_2$ 条件下求 $\mu_1 - \mu_2$ 的置信水平为0.95的置信区间.

六、 （20分）设$\theta \in (a,b)$,$T(X)$是$\theta$的无偏估计，令

$$S(x) = \begin{cases} T(x), & a \leq T(x) \leq b \\ a, & T(x) < a \\ b, & T(x) > b \end{cases}$$

证明：$\mathrm{E}(S(X)-\theta)^2 \le \mathrm{E}(T(X)-\theta)^2$

七、（20分）有一种专门用于动物治疗的新安眠药，据说在一定剂量下，能比某旧安眠药平均增加睡眠时间3小时。根据以往资料，用旧安眠药时平均睡眠时间为20.8小时。为了检验新安眠药是否到达疗效，收集到一组（8个）用新安眠药的睡眠时间分别为：26.7，22.0，24.1，21.0，27.2，25.0，24.3，24.5 。

1. （10分）假定睡眠时间为正态分布，试在显著性水平 $\alpha = 0.05$ 下，判断新安眠药是否已达疗效。

2. （10分）如果没有正态的假定，用符号检验给出检验，并和1.中的结果进行比较。

八、（15分）设总体密度函数为

$$p(x;\theta) = \frac{2}{\theta^2}(\theta - x) \qquad 0 < x < \theta$$

从中获得样本$X_1, X_2, \cdots, X_n$，试给出下列检验问题：

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

的广义似然比检验法则。

# V. 2003年试题

一、（20分）甲、乙两人下棋，每局获胜概率各为0.5，约定谁先胜5局则赢得全部8000元奖金。现已下4局，甲3胜1负，这时因故终止比赛。若按最终获胜概率的比例分配奖金，甲、乙两人各应分得多少奖金？

二、（20分）若 $\xi,\eta$ 独立，且均服从 $N(0,1)$ , 试证 $U=\xi^2+\eta^2$ 与 $V=\dfrac{\xi}{\eta}$ 相互独立。

三、（20分）设 $(\xi,\eta)$ 服从二元正态分布，$\mathrm{E}(\xi)=\mathrm{E}(\eta)=0, \mathrm{Var}(\xi)=\mathrm{Var}(\eta)=1$, 相关系数系数为 $r$ ，求 $\mathrm{E}\max(\xi,\eta)$ .

四、（15分）将编号为 $1,2,\cdots,n$ 的球随机放入编号为 $1,2,\cdots,n$ 的盒中，每盒只放一球。以 $S_n$ 表示球与盒的编号正好相同的个数，求证：

$$\frac{1}{n}(S_n-\mathrm{E}S_n)\xrightarrow{P}0\ ,\ (n\to\infty)$$

五、（20分）设 $X_1,X_2,\cdots,X_n$ 是相互独立的连续型随机变量，且 $X_i$ 的分布函数为 $F_i(x_i)$, $i$=$1,2,\cdots,n$ .试证明随机变量 $Y=-2\sum\limits_{i=1}^{n}\ln(F_i(x_i))$服从 $\chi^2_{(2n)}$ 分布。

六、（20分）设 $X_1,X_2,\cdots,X_n$ 为一组简单随机样本。总体分布密度为

$$p(x,\lambda)=\begin{cases}\lambda x^{\lambda-1}, & 0<x<1\\ 0, & \text{其它}\end{cases}$$

其他 $\lambda>0$ 为未知参数，求 $\lambda$ 的极大似然估计。

七、（20分）设 $X_1,X_2,\cdots,X_n$ 是取自总体 $X$ 的样本， $X$ 具有密度函数为

$$p(x)=\begin{cases}\mathrm{e}^{-(x-\theta)}, & x\geq\theta\\ 0, & \text{其它}\end{cases}$$

证明可取 $X_{(1)}-\theta$ 作为求 $\theta$ 区间估计的枢轴量（其分布与 $\theta$ 无关），并求 $\theta$ 的置信水平为 $1-\alpha$ 的下置信限。

八、（15分）有甲、乙两台机床，加工同样产品，其产品口径服从正态分布，从两台机床的产品中分别抽取6个和9个，测量口径尺寸分别为 $x_1,x_2,\cdots,x_6$, 和$y_1,y_2,\cdots,y_9$ ，并计算得

$$\sum_{i=1}^{6}x_i=204.60\qquad\sum_{i=1}^{6}x_i^2=6978.93$$

$$\sum_{i=1}^{9} y_i = 370.8 \qquad \sum_{i=1}^{9} y_i^2 = 15280.17$$

取显著性水平 $\alpha = 0.05$ ，试问这两台机床加工的零件其口径尺寸的方差是否有显著性差异？（注：连续型随机变量 $X$ 的分位数 $m_\alpha$ 如下定义：$P(X < m_\alpha) = \alpha$，题目可能用到的分位数 $F_{(6.9,0.025)} = 0.18, F_{(9.6,0.025)} = 0.23, F_{(5.8,0.025)} = 0.15, F_{(8.5,0.025)} = 0.21$ ）。

# VI. 2002年试题

一、（10分）甲、乙两袋中分别装有一黑一白两球，从两袋中各随机取一球，交换后分别放入两个袋中，重复做此过程，以 $p_n, q_n$ 和 $r_n$ 分别记 $n$ 次后，甲袋中有两个白球，两个黑球和黑白各一球的概率。试导出 $p_{n+1}, q_{n+1}$ 和 $r_{n+1}$ 用 $p_n, q_n$ 和 $r_n$ 表示的关系式，并求出当 $n \to \infty$ 时的结果。

二、（15分）已知某机器在 $\Delta t$ 时间内停止工作的概率为 $\alpha\Delta t + o(\Delta t), (\alpha > 0)$ , 并假定在不重叠的时间段内机器停止工作的事件是相互独立的，并假定在 $t_0$ 时刻机器在工作。求该机器直到 $t_0 + \Delta t$ 这段时间一直不停工作的概率。

三、（15分）令 $X_1, X_2, \cdots$，为一列独立同分布随机变量，且 $P\{X_j = 1\} = p, P\{X_j = 0\} = q = 1 - p, 0 < p < 1, j = 1, 2, \cdots$ , 记

$$S_n = \sum_{j=1}^{n} X_j, n = 1, 2, \cdots ,$$

试证对任意的 $x \in \mathbb{R}$ ，有

$$\lim_{n\to\infty} P\left\{\frac{S_n - np}{\sqrt{npq}} \le x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{u^2}{2})\,\mathrm{d}u$$

四、（8分）证明二元联合正态分布可完全由其边缘分布决定。

五、（7分）设已有一组由均匀分布 $U(0,1)$ 产生的随机数，说明如何由此得到一组服从正态分布 $N(0,1)$ 的随机数，并说明其原理。

六、（20分）设 $X = (X_1, X_2, \cdots, X_n)$ 是取自正态分布 $N(\mu, \sigma^2)$ 的子样，其中未知参数 $\theta = (\mu, \sigma^2)$ .

1. 证明 $T(X) = (\overline{X}, \sum_{i=1}^{n} X_i^2)$ 是 $\theta$ 的充分统计量，其中 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. 求出 $\mu$ 和 $\sigma^2$ 的无偏估计的方差下界。

七、（20分）设子样 $(X_1, \cdots, X_n)$ 取自均匀分布 $U(\theta, 1), \theta$ 为参数，记 $M_n = \max(X_1, \cdots, X_n)$ ，对检验问题 $H_0: \theta \le \theta_0$ 对 $H_1: \theta > \theta_0$ 取检验函数

$$\Phi(x) = \begin{cases} 1, & 若\ M_n \ge c, \\ 0, & 若\ M_n < c. \end{cases} \quad 这里c是常数.$$

1. 求检验函数的势函数，并证明它是关于 $\theta$ 单调；

2. 在检验 $H_0: \theta \leq \frac{1}{2}$ 对 $H_1: \theta > \frac{1}{2}$ 中选择什么样的 $c$ 能使检验函数的精度 $\alpha_0$ 恰好是0.05.

八、（20分）设子样 $(X_1, \cdots, X_n)$ 取自正态母体 $N(\mu, \sigma^2)$，试证

1. 当 $\mu = \mu_0$ 时，正态分布关于 $\sum_{i=1}^{n}(X_i - \mu_0)^2$ 具有单调似然比；
2. 当 $\sigma^2 = \sigma_0^2$ 时，正态分布关于 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 具有单调似然比.