概率统计B

Probability and Statistics

张思容
zhangsirong@buaa.edu.cn

数学与系统科学学院，北京航空航天大学
School of Mathematics and System Sciences, BUAA

January 6, 2015

复习和讲解

1. 内容复习
   - 概率论
   - 数理统计
2. 小测验讲解
3. 随机游戏 Taking Chances

考试事宜

- 考试时间: 1月21日13：30-15：30,J3-105,J3-106
- 答疑时间: 1月19日下午，20日上午，21日10:30-12:00
  班级答疑: 地点待定
- 联系方式：134-3920-1025. zhangsirong@buaa.edu.cn
- 期末考试不考内容：
  1. 第四章: 第三节中 $Z = max(X, Y)$与 $Z = min(X, Y)$，其中(X, Y)为连续型r.v. 求$F_Z(z)$， ***** 当X, Y不独立时不要求。
  2. 第五章: 第五节
  3. 第七章: χ^2, t, F分布的概率密度函数表达式不要求记
  4. 第八章: 第五节
  5. 第九章: 第三、四节

概率论(1): 概率模型

- 概念：样本空间+概率律；事件；独立与条件概率；
- 结论: 事件复合；加法公式，乘法公式；全概率和贝叶斯公式；排列组合公式；
- 例子: 抽签原理, 生日问题，配对问题，假阳性,重复射击；...

概率论(2): 随机变量与随机向量

- 概念：随机变量X,(累积)分布函数$F(x)$,概率密度函数$f(x)$, p_i. 事件$(-\infty, x)$
  随机向量,联合分布$F(x, y)$或表格, 事件$(-\infty, x) \times (-\infty, y)$;
- 结论: 边缘分布概率$F_X(x)$公式；条件分布概率公式$F(X|Y)$； 边缘密度公式, 条件密度公式及独立；
- 随机变量例子: 贝努利分布,二项分布,泊松分布; 均匀分布， 指数分布， 正态分布；
  随机向量例子: 二维离散分布； 二维正态分布; n维独立分布乘积；

概率论(3): 复合随机变量和计算

- 概念：随机变量的复合$g(X_1, X_2, \ldots)$
- 结论:
  独立随机变量的复合：X, Y独立则$g(X), h(Y)$独立。
  一维情形: 离散$P(g(X) = y) = \sum_{g^{-1}(y)} P(X = g^{-1}(y))$
  连续: $f_Y(y) = \sum_i f(h_i(y))|h_i'(y)|, h_i(y) = x$
  二维情形: $F_{Z=g(X,Y)}(z) = \int\int_{g(x,y)\le z} f(x, y)dxdy$
- 例子: 离散随机变量的简单复合；比如$X^2, X + Y, max(X, Y)$
  连续情形: 分布的线性组合$Z = aX + bY + c$； $Z = X^2 + Y^2$;
  特别独立正态分布的线性组合$N(\sum \mu_i, \sum \sigma_i^2)$,
  两个独立分布的加法(密度的卷积),极大值分布(分布函数乘积),极小值分布;
  常见分布复合：二项分布是贝努利分布的和；独立泊松分布的和是泊松分布；指数分布的极大极小值分布；

概率论(4): 数字特征

- 概念：期望算子EX, 矩DX, $Cov(X,Y)$, $\rho(X,Y)$,不相关；
- 结论: $E(g(X)) = \int g(x)f(x)dx$;
  线性$E(X+Y) = EX + EY$,$EXY = EXEY$(独立);
  $D(X+Y) = DX + DY - 2Cov(X,Y)$;
- 例子: 所有常见分布的期望与方差；常见分布的复合的期望；
  正态分布的高阶矩(密度函数为偶函数的期望(一阶矩)为零,一般有奇数矩为零);
  二维正态分布的相关系数ρ;
  （不要求) 理论证明用：特殊随机变量特征 $Ea = a, Da = 0$,
  $DX = 0 \Longleftrightarrow P(X = EX) = 1$,
  $Cov(X,Y) = 1 \Longleftrightarrow P(aX + b = Y) = 1$

概率论(5): 不等式与极限定理

- 概念：依概率收敛，独立同分布；
- 结论: 切比雪夫不等式的不同形式，(马尔可夫不等式**)
  切比雪夫大数定理:$\lim P(\bar{X}_n=\mu)=1, \bar{X}_n \xrightarrow{p} \mu$,
  中心极限定理：$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$
- 例子: 概率的频率解释；二项分布的正态分布逼近计算；

数理统计(1): 样本与三大分布

- 概念：总体与样本；简单抽样；样本矩$\bar{X}_n, s_n^2$, 统计量，分布的下分位数；
- 结论: 独立分布的复合：Z_i是标准正态分布；
  $\chi^2(n) = Z_1^2 + \cdots + Z_n^2; EX = n, DX = 2n$
  $t(n) = Z/(\chi^2(n)/n)$, 对称性；
  $F(m, n) = \frac{\chi^2(m)/m}{\chi^2(n)/n}$. 非负性*,互逆性*；
- 例子：正态总体的统计量分布:
  样本均值与样本方差独立；
  $\bar{X}_n \sim N(\mu, \sigma^2/n), (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$;
  $\to (\bar{X}_n - \mu)/(S/\sqrt{n}) \sim t(n-1)$.
  两个正态总体的均值差的分布,样本方差比的分布***;

数理统计(2): 点估计与评估

- 概念：参数统计量$\theta = g(X_1, \ldots, X_n)$;
  矩估计，极大似然估计;
  无偏性，有效性(方差小)，一致性(相合性);
- 结论: 矩估计都是一致的**;
  样本均值是期望的最小方差无偏估计MVU。
  样本方差是方差的无偏估计。
- 例子：常见分布的矩估计和极大似然估计(均匀分布*);
  正态总体的极大似然估计是有偏估计;
  给定函数的极大似然估计;

数理统计(2): 区间估计与假设检验

- 概念：置信区间与置信水平$1-\alpha$,
  零假设,备择假设,显著性水平α;
  置信区间的意义；假设检验的两类错误;
- 结论: 一个正态总体的统计推断分布:
  样本均值：已知方差，正态分布,正态检验;
  样本均值：未知方差，$t(n-1)$分布,t检验;
  样本方差：$\chi^2(n-1)$分布，卡方检验;
- 例子：以上三类的区间估计。
  置信水平与单侧检验*：$\alpha=0.05, z_{0.975}=1.96$

QUIZ 小测验 一

1. 设A, B 为任意两事件,则下列关系成立的有( )
   (A) $(A+B)-B=A$;(B) $(A+B)-B=A-B$;
   (C) $(A-B)+B=A$;(D) $(A-B)+B=AB$.
2. 从 $0 \to 9$ 这十个数码中任意取出4个排成一串数码,则数码恰成四位偶数的概率为: （A） $\frac{41}{90}$;（B） $\frac{1}{2}$; （C） $\frac{40}{90}$; （D） $\frac{32}{90}$
3. 一盒子内装有5个红球,15个白球,从中不放回取10次,每次取一个球,则第5次取到的是红球的概率为多少?
4. 袋中装有编号$1-8$的八个球,从中任取3个,则最小号码为偶数的概率为多少?

答案: 1. B, 2.A , 3 $5/(5+15)=1/4$, 4 $11/28$

QUIZ 小测验 二

1. 设随机变量X在$(-\pi/2, \pi/2)$上服从均匀分布,则$Y = \tan X$的概率密度 $f_Y(y) =$
2. 将红、白、黑三只球随机地逐个放入编号为1，2，3的三个盒内（每盒容纳球的个数不限），以X 表示有球盒子的最小号码，求：（1）随机变量X的分布律；（2）X的分布函数。
3. 某仪器上装有4只独立工作的同类元件。已知每只元件的寿命(以小时计) $X \sim N(5000, \sigma^2)$,当工作的元件不少于2只时,该仪器能正常工作。 则该仪器能正常工作5000小时以上的概率为_。
4. P63. 23题:设随机变量X的概率密度为$f(x) = \frac{a}{e^x + e^{-x}}$, (1)确定常数$a$; (2)求X的分布函数; (3)求$P(0 \leq X \leq \ln\sqrt{3})$.

QUIZ 小测验 二: 答案

1. $f_Y(y) = \frac{1}{\pi}\frac{1}{1+y^2}, -\infty < y < +\infty.$
2. （1）随机变量X的分布律;
   $P(X = 3) = 1/27; P(X = 2) = 7/27; P(X = 1) = 19/27;$
   （2）X的分布函数。
   $F(x) = 0, x < 1; 19/27, 1 \le x < 2; 26/27, 2 \le x < 3; 1, x \ge 3$
3. 一个元件正常工作概率$p = P(X > 5000) = 1/2;$
   元件工作个数Y服从$B(4, 0.5)$,
   仪器正常工作概率:
   $P(Y \ge 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - 1/16 - 4/16 = 11/16.$
4. $a = 2/\pi, F(x) = 2/\pi \arctan e^x, P = 1/6.$

QUIZ 小测验 三

1. 设随机变量$X \sim N(\mu, \sigma^2)$,则$E(X-\mu)^4 =$. (D)
   (A) σ^4; (B) $2\sigma^4$; (C) $6\sigma^4$; (D) $3\sigma^4$ 。
2. 设随机变量X存在数学期望EX 和方差DX ,则对任意正数ϵ有((c)),
   （A） $P(|X-EX| \geq \epsilon) > \frac{DX}{\epsilon^2}$, （B） $P(|X-EX| < \epsilon) > 1-\frac{DX}{\epsilon^2}$;
   （C） $P(|X-EX| \geq \epsilon\sqrt{DX}) \leq \frac{1}{\epsilon^2}$;
   （D） $P(|X| \geq \epsilon) \leq \frac{E(X-EX)^k}{\epsilon^k}(k \geq 1)$,
3. 设随机变量X, Y的二阶矩EX^2, EY^2存在， 证明： 成立不等式$|EXY| \leq \sqrt{EX^2 EY^2}$. (构造$E(X+tY)^2 \geq 0$)
4. 设X_n是相互独立的随机变量序列,且其分布律为
   $P(X_n = -\sqrt{n}) = \frac{1}{2^{n+1}}, P(X_n = \sqrt{n}) = \frac{1}{2^{n+1}}, P(X_n = 0) = 1-\frac{1}{2^n}$. 记
   $Y_n = \frac{1}{n}\sum_i X_i$。
   试求： （1） EX_n, DX_n;$(0, n/2^n)$ （2） EY_n, DY_n;$(0, \frac{1}{n^2}\sum_i i/2^i)$
   （3） 证明：对任给$\epsilon > 0$,成立$\lim_{n\to\infty} P(|Y_n| \geq \epsilon) = 0$。

Quiz 小测验 四

1. 设$X_1, X_2, \ldots, X_n$是正态总体$X \sim N(\mu, \sigma^2)$的样本，当$C = ()$时，$\hat{\mu}^2 = \bar{X}^2 + cQ^2$是$\mu^2$的无偏估计， 其中$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $Q^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$.
   （A） $-\frac{1}{n(n-1)}$，（B） $-\frac{1}{n-1}$ （C） $-\frac{1}{n^2}$ （D） $-\frac{1}{(n-1)^2}$ 。
2. 令$X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_m$分别为$N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$的简单样本，记$\bar{X} = \frac{1}{n}\sum_i X_i$, $\bar{Y} = \frac{1}{m}\sum_j Y_j$, $S_1^2 = \frac{1}{n-1}\sum_i(X_i - \bar{X})^2$. 试求：
   （1） $\bar{X}$服从的分布， $\bar{Y}$服从的分布； （2） $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{1/n+1/m}}$ 服从的分布； （3） $T = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_1\sqrt{1/n+1/m}}$ 服从的分布。
3. 根据大量调查得知，我国健康成年男子的脉搏平均为72次/分，标准差为6.75 次/分，现从某体院男生中，随机抽出25人，测得平均脉搏为69.3次/分。根据经验脉搏 服从正态分布.如果标准差不变，试问该体院男生的脉搏与一般健康成年男子的脉搏有无差异? 检验水平$\alpha = 0.05$, (已知$Z_{0.95} = 1.645$, $Z_{0.975} = 1.96$; $t_{0.95}(24) = 1.7109$, $t_{0.95}(25) = 1.7081$, $t_{0.975}(24) = 2.0639$; $t_{0.975}(25) = 2.0595$)

Quiz 小测验 四: 答案

1. (A). 利用$\bar{X} \sim N(\mu, \sigma^2/n)$, $E\bar{X}^2 = D\bar{X} + (E\bar{X})^2 = \sigma^2/n + \mu^2$
   利用$ES^2 = \sigma^2$, $EQ^2 = \sigma^2 * (n-1)^2$,计算可得。
2. $\bar{X} \sim N(\mu_1, \sigma^2/n)$, $\bar{Y} \sim N(\mu_2, \sigma^2/m)$.
   则$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2(1/n + 1/m))$,标准化得到正态分布;
   又$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, 正态分布除以卡方分布得到$t(n-1)$分布。
3. $H_0 : \mu = 72.vs.H_1 : \mu \neq 72$; 检验统计量$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$;
   检验水平对应分位数$z_{0.975} = 1.96$,计算
   有$|Z| = 2.7/6.75 * 5 = 2 > 1.96$, 检验显著, 否定原假设, 有差异。

随机游戏：Taking chances

1. 第一关：随机序列的奥秘；
2. 第二关：统计数字的分布；
3. 第三关：概率的大小；

beat the randomness! 打败随机性

- 第一关：随机序列的奥秘；
  潘尼游戏：（1969年发明)
  抛掷硬币多次直到出现一个特殊的(三位连续)组合为止。第一个玩家先从8个组合中任选一个，第二个玩家(庄家)再选一个不同组合。游戏中哪个组合先出现，那个人赢。
- 第二关：统计数字的分布；给定一组数据，请猜测其第一个有效数字(非零和小数点)最有可能是什么的两个选择。
  2010年联合国人口署给出237个国家人口的估计数。
  自然数$1/n, n = 1, 2, \ldots, 200$
  以6出发，每次乘以2，得到100个数字

第三关：概率的大小

请猜测有关的概率或平均值大小：

两副牌放在桌子上，54个人从每副牌中各选一张，平均多少人拿到一对完全相同的牌？

$EX = 1$

已知一个家庭有两个小孩，已知其中一个是男孩，问他有一个兄弟的概率是多少？假定生男或生女的概率一样。

$P = 1/3$

设四个人打牌(52张)，每人得到13张扑克牌。如果已知你对家有一个A,则他还有一个A的概率为p_1; 如果已知你对家有一个黑桃A，他还有一个A的概率为p_2，问p_1, p_2谁大？不考虑你的和其他人的牌。

$p_1 \sim 0.37, p2 \sim 0.56$

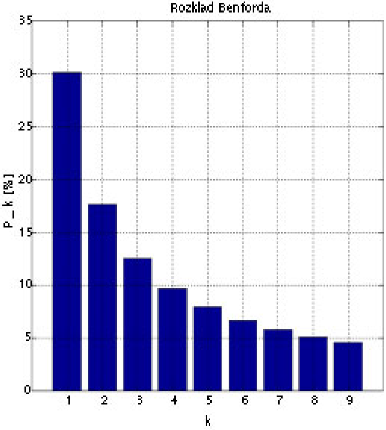
潘尼游戏： Penney's game

潘尼游戏：（1969年发明）抛掷硬币多次直到出现一个特殊的(三位连续)组合为止。第一个玩家先从8个组合中任选一个，第二个玩家(庄家)再选一个不同组合。游戏中哪个组合先出现，那个人赢。

第一个选择	第二个选择	赔率(庄家)
HHH	THH	7 to 1
HHT	THH	3 to 1
HTH	HHT	2 to 1
HTT	HHT	2 to 1
THH	TTH	2 to 1
THT	TTH	2 to 1
TTH	HTT	3 to 1
TTT	HTT	7 to 1

诀窍：$ABC \to \bar{B}AB$.

本福德规律 Benford's law



本福德规律(1938).

大多数真实数据的第一个有效数字不是平均分布的! (服从对数法则)。

世界60最高建筑的高度，人口，河流长度，股票价格，。。。
应用： 审计财务报表!税务局查税!

Taking chances

1. "Two roads diverged in a wood, and I... I took the one less traveled by, and that has made all the difference." Robert Frost
2. "Twenty years from now you will be more disappointed by the things you didn't do than by the ones you did. So throw off the bowlines, sail away from the safe harbor, catch the trade winds in your sails. Explore. Dream. Discover." Mark Twain
3. 电影《死亡诗社》Dead poet society:
   "Carpe diem, seize the day boys, make your lives extraordinary."