《概率论与数理统计》期末考试题解析

序言

在阅读本资料前请确保您已经在身上准备好三样东西：
1. 概率统计教材
2. 2014～2018 年真题及答案解析
3. 一颗准备好好提前复习和迎接挑战的心
我们可以开始了。

目录

卷壹·2014 年卷

一、 全概率公式

知识点摘要：

一个随机试验由先后两个阶段构成，第二阶段结果受到第一阶段的影响。但是，第一阶段的结果已知，想要求的是第二阶段的概率。

做题摘要：

1. 第一题考查全概率公式的应用。本题巧妙之处在于全概率公式涉及到的“两个阶段”中，第一阶段的划分的选取方式。第一个阶段的划分直接将 i 比 j 先发生（L）这一事件分为三个部分：第一次 i 发生条件下 L 发生；第一次 j 发生条件下 L 发生；第一次都不发生条件下 L 发生。这的确是一个划分，因为它包含了第一次试验的所有可能。
2. 第二个巧妙的地方在于条件概率的化简。第一次 i 发生概率 pi，在这个条件下，L（i 比 j 先发生）的概率是 1；第一次 j 发生概率 pj，在这个条件下，L 发生的概率是 0；如果第一次都不发生，那么在这个条件下 L 发生的概率仍为 P（L）。
3. 按上述化简等式，即可解出 P（L）；

技巧摘要：

本类型题做题分为以下步骤：

1. 将所求复杂事件划分为一组完备事件（互斥且并在一起是复杂事件本身）

2. 利用全概率公式和划分好的完备事件求复杂事件的概率

$$P(B)=\sum_{i=1}^{n} P(A_i)P(B|A_i)$$

二、 几何分布的无记忆性

知识点摘要：

$$P(X>n+m \mid X>n)=P(X>m)$$

几何分布的表示方法为：X～G(p)；如果 X 服从参数为 p 的几何分布，那么就可以将 X 看作是首次命中的射击次数，其中 p 是每次设计命中的概率。

做题摘要：

本题考查几何分布无记忆性的证明。详细证明过程及详细说明在教材 39 页十分清楚，故此处不再赘述。这一结论在后续习题中会至少出现 2 次。

三、 正态分布

知识点摘要：

$$U=X+Y\sim N(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2)$$

$$V=X-Y\sim N(\mu_X-\mu_Y,\sigma_X^2+\sigma_Y^2).$$

做题摘要：

本题考查正态分布的概率计算相关问题。

（1） 利用变量代换将正态分布标准化，然后查表即可带入计算。

（2） 根据正态分布的可加性结论（详见教材 91 页）：

a） Y 也是正态分布；

b） Y 的期望是 X 的-2 倍加 1；

c） Y 的方差是 X 的四倍；

d） 将 Y 的期望和方差带入正态分布概率密度函数中：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

得到答案。

四、

第四题考查方差的计算。根据经验，方差的计算在期末考试中出现，难度不会很大。如果计算思路简单的话（使用定义的计算公式 E（xˆ 2）-E（x）ˆ 2），计算复杂度可能相对大一些；如果计算思路简单的话（使用间接求法，如协方差），计算复杂度会小一些。

本题中我们看到，离散型随机变量 x 的取值只有-1、0、1 三种可能。所以我们可以直接根据概率固有的归一性（所有随机变量的概率之和为 1），将-1、0、1 分别带入概率密度函数做和求出未知参数 a 的值。

a 的值求出来后可以顺手求出 P（x=-1），P(x = 0)，P(x=-1)，发现这三个值都是 1/3。所以说伊布西隆的期望是 0. 这使我们更加确信使用定义法求方差。

五、

第五题考查离散型随机变量的联合概率分布求法，独立性判断。

（1） 已知两个随机变量分别的分布列求联合分布列无非是一道填空题。

$\xi \backslash \eta$	0	1	$p_{i\cdot}$
-1	$\frac{1}{4}$	0	$\frac{1}{4}$
0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$	

在列出这个 3*2 的联合分布列后，会有 12 个空需要填（把它当作一个 4*3 的矩阵）。天空顺序如下：

1. 我们已知的显性条件是两个变量乘积为零的概率是 1.即乘积不为零的概率是零。这样（1，2）、（3、2）处的概率填 0；
2. 由于我们已知两个变量的边际分布，所以可以直接填出 pi・和 p・j 的五个空。
3. 利用作差，可以填出（1，1），（3，1）；
4. 再次利用两个变量乘积为零的概率是 1，做差求出（2，2）；

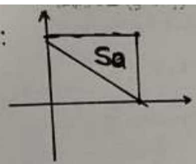
填表结束。

（2） 独立的条件是非常苛刻的，要求每一对边缘分布的乘积都与联合分布的矩阵中的值对应相等（具体细节详见教材 69 页）。很明显 P(X = 1, Y = 1) != P(X = 1) * P(Y = 1)，所以是不相互独立的。

（3） 第三问无非是在第一问的基础上定义了一个新的离散型随机变量。它的取值只能是 0，1. 将联合分布列的矩阵中心矩形的 6 个值分别对应到 0 和 1，分别求其概率和即可。注意，不要漏掉任何一种情况，新随机变量取 0 时对应 2 种情况，取 1 时读经 4 种情况。0 和 1 的概率之和是 1，这一点用来做验算。

六、

解:



$$\cancel{E}(\xi,\eta)=E(\xi)+E(\eta)=\frac{1}{2}+\frac{1}{2}=\frac{1}{4} \times$$

$$(\xi,\eta)\sim p(x,y)=\begin{cases}2 & (x,y)\in G\\ 0 & (x,y)\notin G\end{cases}$$

$$1-1-y^2+2y+2y^2$$

$$E(\xi,\eta)=\int_0^1 dy\int_{1-y}^1 (x+y)\cdot 2\,dx=\int_0^1 dy\cdot\left(x^2\Big|_{1-y}^1+2y(y)\right)=\int_0^1 dy\cdot\left[1-(1-y)^2+2y^2\right]$$

$$=\int_0^1 y^2+2y\,dy=\frac{1}{3}+1=\frac{4}{3}$$

本题使用了定理 4.1.2（教材 98 页（2））。想要求期望和方差共分为以下 个步骤：

1. 求出联合概率密度函数：
   a) 判断边缘分布类型——均匀分布
   b) 求出均匀分布面积
   c) 画出可积域函数图像
   d) 写出均匀分布概率密度函数（例子参见教材 84 页-例 3.4.3）
2. 带入二重积分公式求出期望。二重积分的积分域是整个均匀分布可行域，不要与求联合密度函数的可行域混淆。
3. 利用定义计算式，根据期望求出方差。

七、矩估计法

知识点摘要：

矩估计法的基本思想是替换原理，用样本矩替换同阶总体矩。如果总体 X 的未知参数多于一个，如 k 个，那么需要假设 X 的前 k 阶总体矩都存在，令各阶样本矩与同阶总体矩相等。这样就可以用 k 个方程解出 k 个参数，求出对应的矩估计量。

做题摘要：

本题的特点在于一阶总体矩等于样本矩的话无法引入参数。所以果断灵活使用二阶样本矩等于二阶总体矩。计算过程不再赘述。

第二问要求矩估计量的平方的期望。矩估计量平方后刚好消掉根号，把 3/n 提出来然后求里面的“样本观测值平方再求和”。根据这些样本观测值的独立同分布的性质，我们可以把和的期望转化为期望的和。样本观测值的平方我们在第一问中已经用方差的方法间接求出了，所以就得到了答案。

八、

（1）

已知：

$$\mu = 80.$$
$$X^{bar} = 78.25$$
$$S^2 = 2.5^2.$$

单正态分布公式

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

带入化简（注意 t 分布的对称性，使得置信区间具有对称性）为关于 sigma 方的不等式，注意在化简的过程中不等号变号。

（2）

假设：总体均值$\mu = 80.$

样本观测值：Xbar = 78.25. $S^2 = 2.5^2.$

原假设：$H_0: \mu = 80.$

备择假设：$H_1: \mu \neq 80$.

显著性水平：0.05.

选取公式：单总体正态分布（3）

代入 n=25、$S = 2.5$ 、$\mu_0 = 80$, $X^{bar} = 78.25$

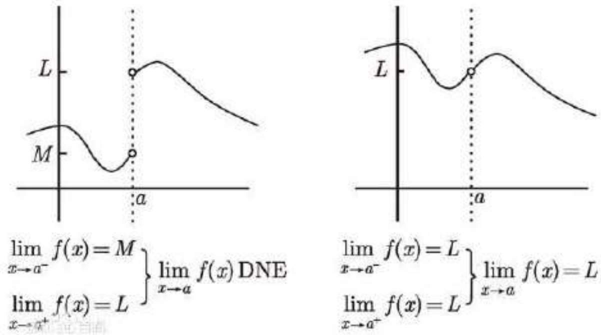
落入拒绝域，小概率事件，H_0被拒绝，备择假设成立。

卷贰·2015 年卷

一、

1. 从结果出发，化简结果，找到需要的概率。再从已知出发，化简已知，得到需要的概率带入结果化简式计算。
2. 第二问考查混合型随机变量已知分布列求某点概率的问题。乍一看还是很奇怪的。因为这是一个离散和连续混合的随机变量，它的概率密度函数是既有连续部分，又有离散部分的。

本题的解法使用是，如果该点事区间的非连续点（通常只会间断点），那么用该点的分布函数值减去该点的分布函数左极限。我们在工数（上）求断点处的跳跃值大小时曾经使用过这个计算方法。当时需要我们判断该函数是左连续还是右连续。分布函数一定是右连续的，所以我们只需要在断点处关心左极限即可。



这里我们做一个复习，左极限是指从间断点横坐标的左侧趋紧得到的极限值，用（x-）来表示（负号表示从负方向趋近，即左侧）；右极限是指从间断点横坐标的右侧趋近得到的极限值，用（x+）来表示。由于概率密度函数是右连续函数，即：

$$\lim_{x\to x_0^+} f(x) = f(x_0)$$

所以间断点在右侧能取到值，且这个值根据概率的非负性一定比左侧的极限大。这个差值，即跳跃值即是间断点处概率大小。

3. 第三问的核心在于 p 这个未知量的求解，根据题干中 5/9 的概率和二项分布的分布列可以求解。由于二项分布的伯努利实验只有两次（n = 2）所以 x 的取值只能是 0、1、2；所以 P（X = 0）=4/9 即 q 方等于 4/9，q = 2/3，p = 1/3.所以 Y 这个几何分布的方差是 q/p^ 2 = 6；
   注意：二项分布和几何分布的方差容易记混。二项分布的方差是 npq，几何分布的方差是 q/p^ 2；

二、

对于一维连续型随机变量函数的分布的概率密度计算分为以下几步：

1. 根据 X 的分布类型计算 X 的概率密度函数；
2. 根据 X 和 Y 的函数关系，使用 Y 表示出 X（反函数），带入到 X 的概率密度函数表示出 Y 的分布函数；
3. Y 的取值区间是由关于 X 的函数决定的，注意 Y 在什么范围内存在分布函数，讨论要全面。
4. 将分布函数的分段函数求出后对于每段求倒计算出分段的关于 Y 的密度函数。

三、

第三题考查二维离散型随机变量已知分布列求边际分布列的计算。根据教材上 66 页公式 3.2.2 和 3.2.3 可以求出边际分布列。

1. 在求 P（X = n）时可以把 n 当作一个像常数一样的常量，将二维降至一维。

$$常用公式1: \frac{n!}{(n-i)!\,i!} = C_n^i$$

$$常用公式2: \sum_{i=0}^{n} C_n^i a^i (b-a)^{n-i} = b^n$$

2. 利用条件概率计算公式计算，带入第一问的结果并化简。注意 5^n = 5^m*5^(n-m);利用二维离散型随机变量的条件分布列还可以求出根据乘法公式求出二维离散型随机变量的联合分布列。
3. 考查二维离散型随机变量函数的分布，技巧可以参考教材 81 页例 3.4.1 的证明。将其中一个变量作为累加变量 i，另一个变量就可以根据函数关系用 i 表示。最终将随机变量函数的分布列变为二维随机变量的分布列的求和。

$$formula_3: \sum_{k=0}^{\infty} \frac{a^k}{k!} = e^k$$

四、

1. 考查二维连续型随机变量已知联合密度函数求边际密度函数。通过联合密度求边际密度的思路无论是离散的还是连续的随机变量都是一样的，将要求的变量当作已知，对另一变量作为积分变量求积分。只不过连续时，最好把可行域的面积画出来，便于上下限的选取。注意在表示边际密度是也要分段表示。
2. 考查二维连续型随机变量函数的密度函数计算。思路和离散时也是一模一样的，只不过这次是将随机变量函数作为一个条件限制来缩小可行积分域，本质上还是转化为对二维连续型随机变量分布函数的计算，将随机变量函数的自变量作为积分上下限进行计算。注意：可能会出现根据自变量进行分段的可能。

五、

第五题考查利用独立性求方差和期望（性质）。当分段时，我们发现 ZW 其实永远等于 XY。在求方差时使用技巧

$$trick_1: E(x^2) = D(x) + E(x)^2$$

六、双侧检验

假设：总体方差$\sigma^2 = 4000.$
样本观测值： $S^2 = 5200.$
原假设：$H_0: \sigma^2 = 4000.$
备择假设：$H_1: \sigma^2 \neq 4000$.
显著性水平：0.05.
选取公式：单总体正态分布（2）
代入 n=21、$S^2 =\ 5200$
未落入拒绝域， H_0被接受，原假设成立，没有显著差异。

七、

已知：
样本均值：8.65
总体方差:0.6ˆ 2
求总体均值置信区间
使用单正态分布公式：

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

带入化简时不等号变号，注意标准正态分布分位点的对称性。

八、

知识点摘要：
“极大似然估计值”的想法是根据“概率最大的事件最可能出现”的原理才成立的。我们首先根据密度函数构造出极大似然函数$L(\theta)$,然后利用对数似然函数（如果$L(\theta)$比较复杂，直接求导难度较大的话，可以对$L(\theta)$取对数，得到**对数似然函数$\boldsymbol{ln}L(\theta)$**，由于它和$L(\theta)$有相同的最大值点，因此求对数似然函数的最大值点即可得到(θ)的最大似然估计）。

做题摘要：
本题在计算上并没有设置难点。需要使用对数似然函数，求出来的对数似然函数关于μ的一阶导为 n。那也就是说μ越大越好喽，μ最大不能超过样本观测值，这是题目中给出的条件，所以很自然的得出让μ的极大似然估计值为所有样本观测值中的极小值即可。

答案中海使用密度函数的归一性求出了θ，实际上并没有十分的必要，因为题干中已经明确θ只是已知常数而已，但是利用归一性求未知参数却是一种常用的技巧。（当然也很好用）

卷叁·2016 年卷

一、

1. 第一题考查对二项分布的理解。将整个过程拆分是一种巧妙的技巧：

- 第六次射击命中概率：1/3
- 且前五次射击命中 2 次

这两个事件的积事件（且的关系）即是我们要求的过程的概率。

2. 和卷贰中的没有本质区别，从结论出发化简结论寻找需要的条件。然后从已知出发去构造这些条件。本题中的关键在于从结论中找到 P（A+B）=P（A）这一必须要从已知中获取的条件。也就是 A 包含 B 的关系。
3. 和之前预言的一样，这是几何分布无记忆性的再次上演。X>6|X>3 = X>3 这一关键的关系是解决问题的关键。带入到几何分布的分布列中求出概率，实际上就是求以 pq^3 为首项，（1-p）为公差的等比数列前 n 项求和。

二、

本题考查的知识点是之前卷贰提到的利用条件概率的乘法公式求边际分布列的问题。已知的两个分布：X 的泊淞分布是 Y 的二项分布的条件。如果要计算 Y 的边际分布列就要利用条件概率和 X 边际分布列的乘积计算联合分布列，再根据联合分布列求出 Y 的边际分布列。（请注意：在 X 已知的条件下，Y 的个数才能以二项分布计算）化简联合分布列计算式时用到公式：

$$常用公式1: \frac{n!}{(n-i)!\,i!} = C_n^i$$

化简 Y 边际分布列计算式时用到公式：

$$formula_3: \sum_{k=0}^{\infty} \frac{a^k}{k!} = e^k$$

注意，在 Y 边际分布计算式化简时只有 n-m 是未知参量，m 可以作为已知参量。

三、

1. 考查由联合密度函数求边际密度函数。老方法，求谁把谁当作已知，然后对另一个变量积分。
2. 第二问计算的技巧在于发现联合密度函数均匀分布的本质。然后只需要求出 y=zx 与 y=1-x 的围城的面积（下部分）与整个积分域（面积 1/2）的比值即是 z 的分布函数（注意分段）。

四、

本题的特殊之处在于 X，Y，Z 是一个三维均匀分布。但是定义就是定义，根据定义法求方差是屡试不爽的。这就需要求出 x 的期望和 x^2 的期望。按照方差的定义求出来的 X 的期望是 0（因为本质上是对奇函数 X 求二重积分，必然为零）；X 平方的期望根据《工科数学分析 2》里用到的 X、Y、Z 三个变量完全等价的方法，求出三个变量平方和的积分（即球半径的平方）。注意分母上永远都是表面积，因为题干中说在单位球面上等可能的任取一点，这就把二维的面积扩展为三位的表面积。

五、

求三次幂的期望，我们貌似并没有遇到过这样的题目，担当我们看到对数函数（反函数是指数函数）和正态分布结合在一起时，我们就仿佛明白了什么。把 lnx 换成一个新的随机变量 Y 是必要的，因为 Y 是符合正态分布的，而且 X 与 Y 是指数关系。

按照定义求期望是本题的方法。注意 X 和 e 的 Y 次幂是两个完全等价的随机变量，所以求 X 的三次幂就等价于求 e 的 3Y 次幂的期望。那么事情就这样解决了，将 e 的 3Y 次幂乘以 Y 的概率密度函数，一个 N（0，4）的正态分布再对其从负无穷到正无穷积分即可。算出来的即是 e 的 3Y 次幂的期望同时也是 X 的三次幂的期望。

在计算积分的时候用到了一个技巧就是构造正态分布的积分是其在负无穷到正无穷上界分为一。恰巧我们构造出的是 e 的 18 次幂与这样一个 N（12，4）的正态分布的乘积的积分的形式。所以积分结果就是 e 的 18 次幂。

六、

知识点摘要：无偏估计量

参数的估计量也是作为一个统计量存在的，因为它是基于不同的样本观测值求得的。如果想要确定估计量的优劣就应该参考它的均值，所以我们关注估计量

的期望（当样本足够多时，均值是以一定概率收敛于方差的）是否恰好就是被估计的那个参数的期望。我们把“是”的情况中的估计量称为**无偏估计量**。

做题摘要：

1. 本题第一问计算没有难度，注意密度函数中前面的 1/sigma 在似然函数中要做 n 次方，然后使用对数似然函数进行化简，然后求对数似然函数的一阶导，令一阶导为零，得到极大似然估计量。
2. 本题的第二问考查估计量的无偏性。就是求出极大似然估计量的期望。注意，在利用期望定义求积分时关注奇偶性进行化简是事半功倍的效果。

七、双侧检验

（1）

已知：

$$S^2 = 2.12^2$$

样本均值：50.4

求总体均值置信区间。

单正态分布公式：

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

带入化简，化简时不等式变号，注意 t 分布的对称性。

（2）

假设：总体方差$\sigma^2 = 4.$

样本观测值：$S^2 = 2.12^2.$

原假设：$H_0 : \sigma^2 = 4.$

备择假设：$H_1 : \sigma^2 \neq 4$.

显著性水平：0.05.

选取公式：单总体正态分布（2）

代入 n=16、$S^2 = 2.12^2$

未落入拒绝域，H_0被接受，原假设成立，没有显著差异。

2017 年卷·卷肆

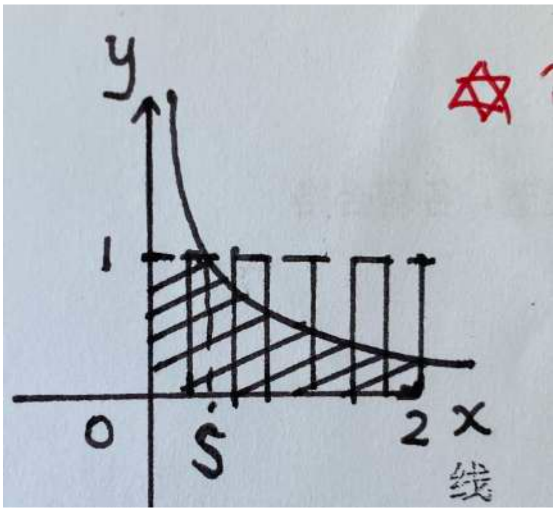
一、

本题和卷三第五题如出一辙。首先是用 Y 来限定 X 的取值范围，然后结合 X 的积分域和概率密度函数求得 Y 的分布函数，再求导求出 Y 的概率密度函数。要注意时刻关注 Y 的取值范围。

二、

本题是一道积分域有特色的题目，考查二维均匀分布的随机变量函数的密度函数求法。解题步骤如下：

1. 可以求出（X，Y）的联合密度函数；
2. S = XY；
3. 求出 S 的分布函数；
   a） 画出积分域



   b） 根据积分域（下方面积）带入联合密度函数求积分：[0:S]和[s:2]分开求；
   c） 将两部分积分相加作为 S 的密度函数；
4. 求导求出 S 的概率密度函数；

并不喜欢使用答案提供的那种方法，因为这和我们之前反复使用的方法是不同的。我们为了练习之前反复使用的“根据联合密度函数求二维随机变量函数的密度函数”的相关方法解决这道题目。

三、

本题正向和逆向使用条件概率公式。Y 是已知的分布，X 是 Y 条件下的均匀分布。本题的结论是求在 X 的条件下，Y 的概率密度函数。从结论出发，利用条件概率公式，将结论转化为求 X 和 Y 的联合密度和 X 独立的概率密度函数的问题。前者可以利用已知的两种分布使用条件概率的乘法公式求得，后者可以使用对求得的联合密度函数作关于 Y 的积分的方式求得。具体流程如下：

1. 写出已知的两个概率密度函数；
2. 利用乘法公式将两个概率密度函数乘积求得对应区间的联合密度函数；
3. 对联合密度函数关于积分域求 Y 的积分，求出 X 的边际密度函数；
4. 将 2、3 带入条件密度函数计算公式求出结果；

四、双侧检验

假设：总体均值$\mu = 6$.
样本观测值：Xbar = 6.36. $S^2 = 0.72^2$.
原假设：$H_0 : \mu = 6$.
备择假设：$H_1 : \mu \neq 6$.
显著性水平：0.05.
选取公式：单总体正态分布（3）
代入 n=9、$S = 0.72$ 、$\mu_0 = 6$, $X^{bar} = 6.36$
未落入拒绝域，H_0被接受，原假设成立。

五、

1. 从结论出发化简，非常简单
2. 这道题妙就妙在其实除了 2 号球、4 号球，之外的球都全部扔掉即可，因为它们对于问题毫无用处。问题可以重新叙述成这个样子“袋中有 5 个球，标有 2 号（3 个）和 4 号（2 个），每次从中取出一个然后放回，则 2 号球比 4 号球线被摸到的概率是多少？”这就是非常简单的一道题了，因为第一个球就直接决定了 2 号球和 4 号球谁先被摸到，第一个球必须只能是 2 号球，它的概率是 3/5.
3. 根据几何分布的无记忆性（第三次出现了），P（X>3）的概率是 1-P（X=1）-P（X=2）-P（X=3）=1-0.9-0.09-0.009=1-0.999=0.001；请记住：几何分布是离散分布。几何分布的随机变量是从 1 开始取值，而不是 0；射击时，第一次命中对应的射击次数是几何分布的应用。
4. 本质上，Fy（3）和 Fx（3）是完全相等的。这是为什么呢？Y 的取值是 X 和 2 的较大者，那么当 X<2 时，Y=2，它的概率是和 X<2 的密度函数的积分相等的，即和 Fx（2）相等，就好像离散型随机变量要把所有能取到相同随机变量值的概率加起来一样。而在（2，3）的区间内，y 的概率密度和 X 是一模一样的，所以本质上 Fy（3）就是把（0，3）的所有 X 的概率密度函数做个积分，即 Fx（3）。
5.
6. 双正态分布公式：

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

已知样本方差（S），求总体方差比值，不等式在化简中不变号。

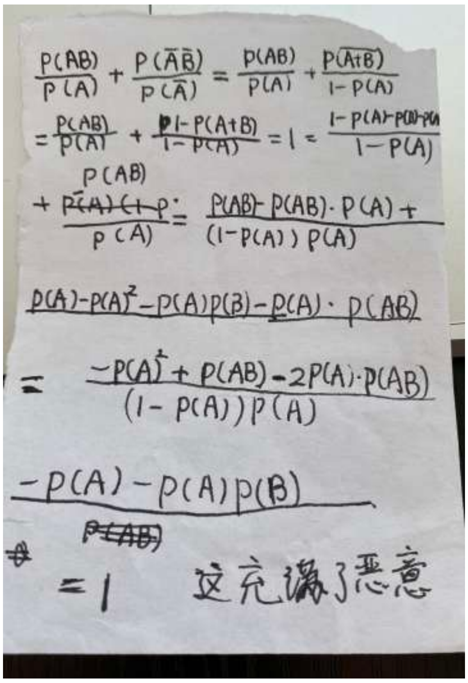
7. 最简单的一阶矩等于均值的估计方式，求出样本观测值的平均值就是估计值。
8. 本题是作业题，按照连续函数极大似然估计将所有样本观测值代入密度函数后相乘，取对数化简后得到：

$$L(\theta) = n\theta - \sum_{i=1}^{n} x_i$$

显然这是一个关于θ的递增函数，所以当θ取极大值时，$L(\theta)$极大。所以让θ等于所有样本观测值的最小值即可（因为 X<θ）。

六、

1. 这题非常技巧。把 A、B 独立作为条件代进去化简结论，发现正好是 B 的概率和 B 的补事件的概率和，为 1。如果要严格推导的话费时费力：



对立时事件是指概率和为 1 的事件。相等事件是 A 和 B 的概率相等。交集为空是二者互不相容。

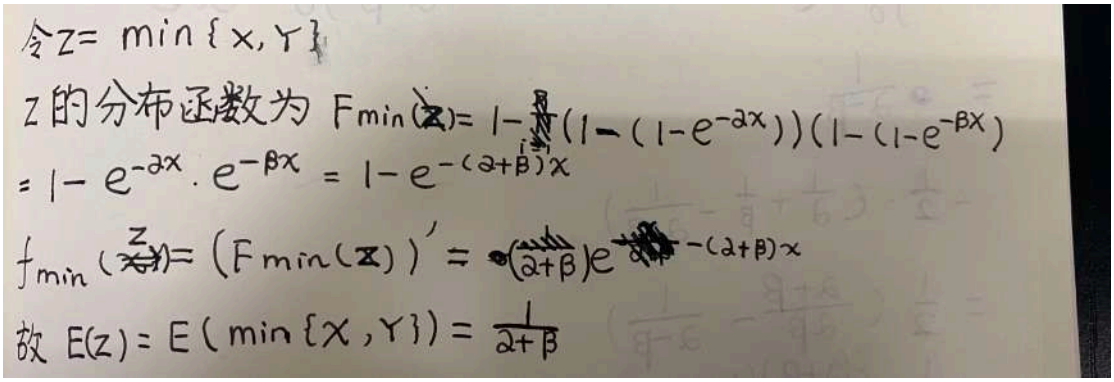
2. 化简有 P（AB）=P（A）这能说明 A 是 B 的**子集**，可不一定是真子集。
3. 敬请期待
4. 将 Y 的期望分段计算，X 在 0，1 处的概率和，是 Y 取 2 的概率；X 在大于 1 时与 Y 同概率。所以 Y 的期望可以使用 X 的期望计算。将 X 的期望计算公式中的 X 换成 2 和 2^X 分段计算求和即可。
5. 本题考查协方差的分配律。将 U、V 放到 Cov（U，V）里然后利用协方差的分配律化简 Cov（U，V），这样就能使用 X 和 Y 的方差和相关系数表示协方差。而协方差等于零时，X 和 Y 独立时协方差。所以令协方差为零。解出相关系数 7/2.
6. 本题考查知识点正态总体的抽样分布。t(n-1)需要使用一个标准正态分布除以一个自由度为 n-1 的卡方分布除以 n-1 再开根号。那么标准正态分布可以使用正态分布标准化来构造。首先想到 X 的均值由于 X 时正态分布的缘故也是一个$N\sim\left(\mu,\frac{\sigma^2}{n}\right)$，而从结论入手发现需要构造 Xn-1-Xbar，所以我们根据

正态分布的可加性判断出这也是一个正态分布$N\sim\left(0,\frac{(n+1)\sigma^2}{n}\right)$。然后进行标准化就使得结论里不含$\mu$。
而对于卡方分布的构造直接使用教材 135 页的单总体正态分布定理的第二结论即可。上下两个统计量相除即可计算出 t（n-1）

2018 年卷·卷伍

一、

1. A 与 B 不同时发生可以求出 AB 同时发生的概率是 1/9. 又由于 AB 独立且概率相等，所以求出 A 和 B 的概率是 1/3. 那么 A 与 B 同时不发生的概率就是 A 非乘 B 非的概率即（A+B）非的概率，等于 1-P（A）-P（B）+P（A）P（B）
2. 本题利用条件概率公式，从结论出发需要得到 P（AB）的概率。这根据 P（A|B）能够求出。
3. 本题考查利用协方差计算方差的问题。方差和 X、Y 相同的协方差是等价的，所以将方差化为协方差再利用协方差的分配律进行计算是最方便的。
$$Cov(X-3Y,X-3Y)=Cov(X,X)+9Cov(Y,Y)-6Cov(X,Y)$$
$$=D(X)+9D(Y)-6\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$
4. 敬请期待
5. 我们已知的是样本方差S^2，也只有样本方差。想要求的是总体方差σ^2。使用公式
$$P\left(\chi_{1-\frac{\alpha}{2}}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi_{\frac{\alpha}{2}}(n-1)\right)=1-\alpha$$
带入$S^2=45,n=8$查表得到$\chi_{\frac{\alpha}{2}}(n-1)$和$\chi_{1-\frac{\alpha}{2}}(n-1)$，$\alpha=0.05$即可计算出置信区间。
6. 本题答案的方法并不可取。因为指数分布不具有可加性，无法保证两个独立的指数分布随机变量的差仍然满足指数分布，因为 X-Y 的取值范围不再是（0，+∞）而是 R。所以可以使用极大极小值分布原理，现求出 min（X，Y）的分布函数然后对分布函数求导求出概率密度函数，最后根据概率密度函数尝试判断方差。



令 $Z=\min\{X,Y\}$

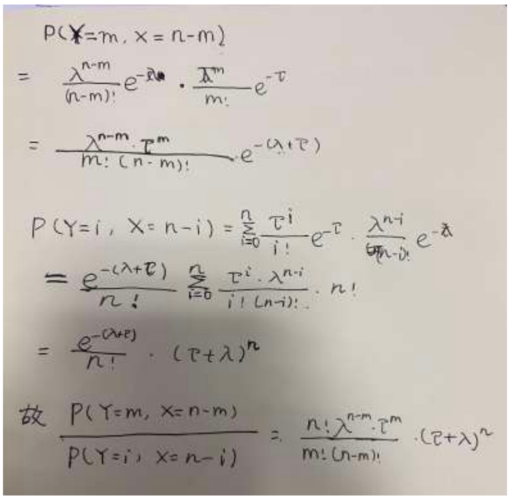
Z 的分布函数为 $F_{min}(z)=1-(1-(1-e^{-\alpha x}))(1-(1-e^{-\beta x}))$

$=1-e^{-\alpha x}\cdot e^{-\beta x}=1-e^{-(\alpha+\beta)x}$

$f_{min}(z)=(F_{min}(z))'=(\alpha+\beta)e^{-(\alpha+\beta)x}$

故 $E(Z)=E(\min\{X,Y\})=\frac{1}{\alpha+\beta}$

7. 本题的结论可以转化为 P（X=m，Y=n-m）/P（X+Y=n）



$P(Y=m, X=n-m)$

$=\frac{\lambda^{n-m}}{(n-m)!}e^{-\lambda}\cdot\frac{\tau^m}{m!}e^{-\tau}$

$=\frac{\lambda^{n-m}\cdot\tau^m}{m!\,(n-m)!}e^{-(\lambda+\tau)}$

$P(Y=i, X=n-i)=\sum_{i=0}^{n}\frac{\tau^i}{i!}e^{-\tau}\cdot\frac{\lambda^{n-i}}{(n-i)!}e^{-\lambda}$

$=\frac{e^{-(\lambda+\tau)}}{n!}\sum_{i=0}^{n}\frac{\tau^i\cdot\lambda^{n-i}}{i!\,(n-i)!}\cdot n!$

$=\frac{e^{-(\lambda+\tau)}}{n!}\cdot(\tau+\lambda)^n$

故 $\frac{P(Y=m, X=n-m)}{P(Y=i, X=n-i)}=\frac{n!\,\lambda^{n-m}\cdot\tau^m}{m!\,(n-m)!}\cdot(\tau+\lambda)^n$

8.

9. **做题摘要：**

本题仍然是考查极大似然估计，和前几套的极大似然估计的类型完全一样，详细解题过程可参见卷壹到卷肆。

10. **做题摘要：**

$$P(\text{次品出现两次正面}\mid\text{出现两次正面})=\frac{P(\text{次品出现两次正面})}{P(\text{出现两次正面})}$$

$$=\frac{P(\text{次品出现两次正面})}{P(\text{次品出现两次正面})+P(\text{正品出现两次正面})}=\frac{\frac{1}{16}}{\frac{1}{16}+\frac{1}{4}}$$

$$=\frac{1}{5}$$

二、已知概率密度求一元随机变量函数的分布密度

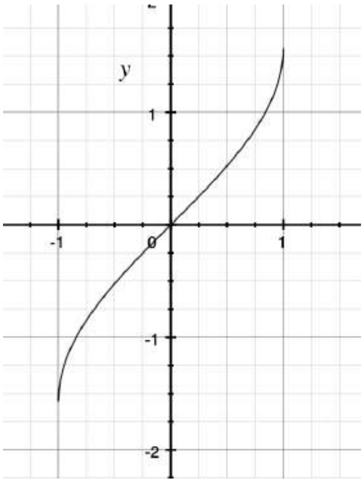
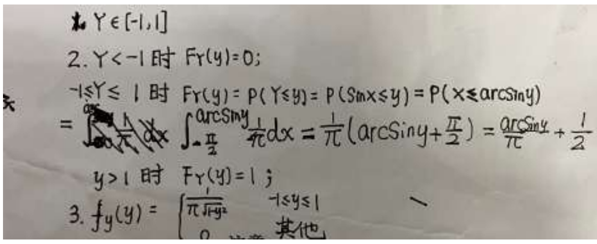


Figure 1 反正弦函数图像

1. Y 的取值范围：[-1, 1]；
2. 将 Y 分段表示出 Y 的分布函数；
3. 对 Y 求导表示出 Y 的密度函数；



$Y \in [-1,1]$

2. $Y<-1$ 时 $F_Y(y)=0$;

$-1\le Y\le 1$ 时 $F_Y(y)=P(Y\le y)=P(\sin x\le y)=P(x\le \arcsin y)$

$=\int_{-\frac{\pi}{2}}^{\arcsin y}\frac{1}{\pi}dx=\frac{1}{\pi}(\arcsin y+\frac{\pi}{2})=\frac{\arcsin y}{\pi}+\frac{1}{2}$

$y>1$ 时 $F_Y(y)=1$;

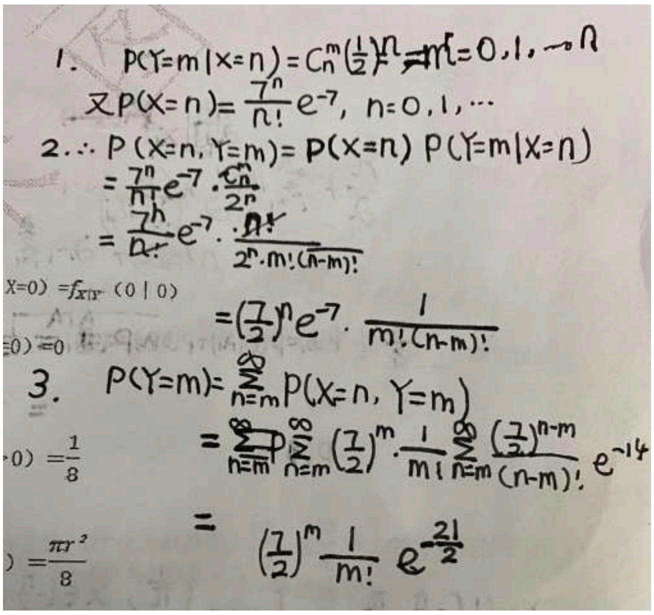
3. $f_Y(y)=\begin{cases}\frac{1}{\pi\sqrt{1-y^2}} & -1\le y\le 1\\ 0 & 其他\end{cases}$

Figure 2 二题图解

三、二维离散型随机变量的条件分布列

做题摘要：

1. 计算条件密度函数；
2. 根据条件密度函数乘法公式计算联合密度函数
3. 对联合密度函数求和计算边际密度函数



1. $P(Y=m|X=n)=C_n^m(\frac{1}{2})^n, m=0,1,\cdots n$

又 $P(X=n)=\frac{7^n}{n!}e^{-7}, n=0,1,\cdots$

2. $\because P(X=n,Y=m)=P(X=n)P(Y=m|X=n)$

$=\frac{7^n}{n!}e^{-7}\cdot\frac{C_n^m}{2^n}$

$=\frac{7^n}{n!}e^{-7}\cdot\frac{n!}{2^n\cdot m!(n-m)!}$

$=(\frac{7}{2})^n e^{-7}\cdot\frac{1}{m!(n-m)!}$

3. $P(Y=m)=\sum_{n=m}^{\infty}P(X=n,Y=m)$

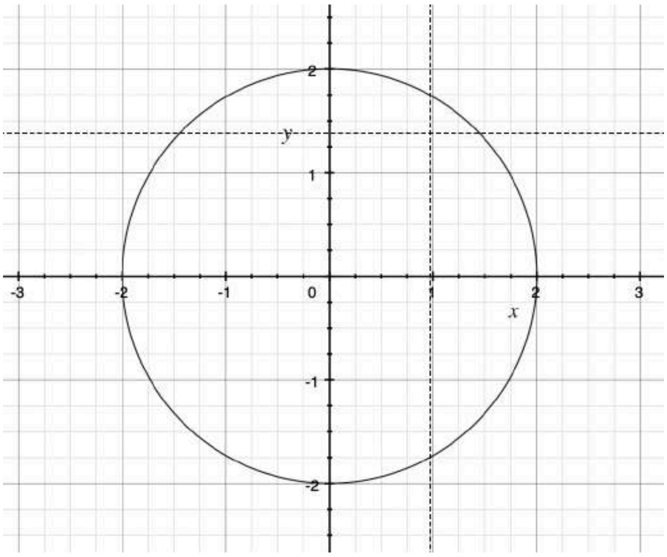
$=\sum_{n=m}^{\infty}\sum_{n=m}^{\infty}(\frac{7}{2})^m\cdot\frac{1}{m!}\sum_{n=m}^{\infty}\frac{(\frac{7}{2})^{n-m}}{(n-m)!}e^{-14}$

$=(\frac{7}{2})^m\frac{1}{m!}e^{-\frac{21}{2}}$

Figure 3 三题图解

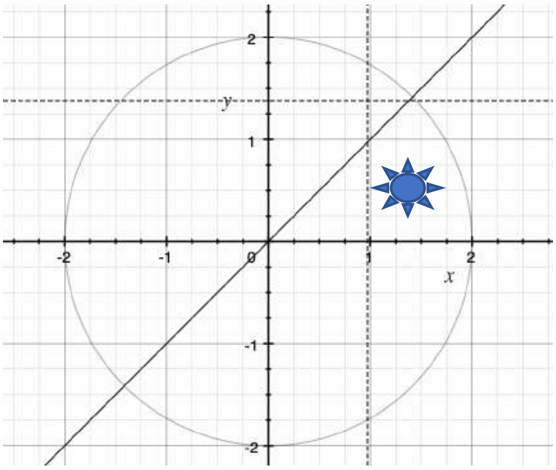
四、

本题棘手的地方在于 X 与 Y 的关系不是函数关系。



1. 第一问是考查对于独立性和相关性的理解。X 与 Y 不是独立的因为给定的平方和关系就决定了它们必须是有关联的。但是有关联的意思是相互影响，而不是线性相关。相关的意思其实是线性相关，这在教材 110 页可以体现出来。所以 X 和 Y 既不独立，也不（线性）相关。
2. 第二问中 Y 在 X 一直的条件下是均匀分布，这是合理的,因为一旦 X 的值已知，Y 的区间就是定长，由于（X，Y）是均匀分布，那么当 X 一直时 Y 的那条线段上各点的取值概率是相等的，因为也是连续函数。Y 的边际分布是将 X 取各值时 Y 的概率，由于 X 取各个值时，Y 的大小并不相等，所以 Y 的边际分布并不满足在各点概率相等，因而不是均匀分布。

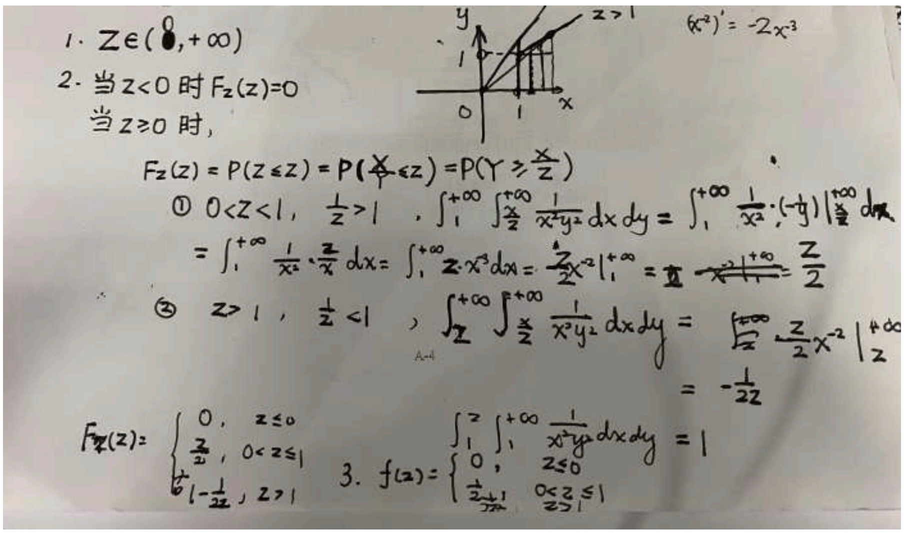
3. 承接 2，当 X=0 时，Y 无非是有大于零和小于零两种选择，而且是均匀分布，所以每种选择的概率是 1/2.
4. 第四问其实是将整个圆分为八个卦象，落在太阳所在的那个卦象的概率为 1/8。



五、二维连续型随机变量函数的分布

做题摘要：

1. Z 的取值范围；
2. 根据 Z 的取值范围分段，求密度函数非零区域的 Z 的分布函数；
3. 对分布函数求导得到密度函数



1. $Z \in (0, +\infty)$

2. 当 $z<0$ 时 $F_Z(z)=0$

当 $z \geq 0$ 时，

$F_Z(z) = P(Z \leq z) = P(\frac{X}{Y} \leq z) = P(Y \geq \frac{X}{z})$

① $0<z<1$，$\frac{1}{z}>1$，$\int_1^{+\infty}\int_{\frac{x}{z}}^{+\infty} \frac{1}{x^2y^2}dxdy = \int_1^{+\infty} \frac{1}{x^2}\cdot(-\frac{1}{y})\Big|_{\frac{x}{z}}^{+\infty} dx$

$= \int_1^{+\infty} \frac{1}{x^2}\cdot\frac{z}{x}dx = \int_1^{+\infty} z\cdot x^{-3}dx = -\frac{z}{2}x^{-2}\Big|_1^{+\infty} = \frac{z}{2}$

② $z>1$，$\frac{1}{z}<1$，$\int_z^{+\infty}\int_{\frac{x}{z}}^{+\infty} \frac{1}{x^2y^2}dxdy = -\frac{z}{2}x^{-2}\Big|_z^{+\infty}$

$= -\frac{1}{2z}$

$(x^{-2})' = -2x^{-3}$

$\int_1^{z}\int_1^{+\infty} \frac{1}{x^2y^2}dxdy = 1$

$F_Z(z) = \begin{cases} 0, & z \leq 0 \\ \frac{z}{2}, & 0<z\leq 1 \\ 1-\frac{1}{2z}, & z>1 \end{cases}$

3. $f(z) = \begin{cases} 0, & z\leq 0 \\ \frac{1}{2}, & 0<z\leq 1 \\ \frac{1}{2z^2}, & z>1 \end{cases}$

Figure 4 五题图解

六、双侧检验

做题摘要：

假设：总体均值$\mu = 6.$
样本观测值：Xbar = 5.89. $S^2 = 1.1^2.$
原假设：$H_0\!:\mu \geq 6.$
备择假设：$H_1\!:\mu < 6$.
显著性水平：0.05.
选取公式：单总体正态分布（3）

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

代入 n=16、$S = 1.1$、$\mu_0 = 6$，$X^{bar} = 5.89$
未落入拒绝域，H_0被接受，原假设成立,产品寿命均值合乎标准。

论假设检验中的做题技巧

假设检验这个类型题是按照一个主线来做题的。这条主线是：

- 判断哪一侧的检验
- 根据题目中关键字建立原假设和备择假设
- 套用单正态或双正态的六个公式
- 利用“小概率事件在一次实验中不可能发生”的原理构造出分位点
- 求出拒绝域
- 根据拒绝域判断原假设是否成立。

判断哪一侧

双侧检验最简单，判断方差，标准差，均值的与他题目中的假设是否一样即可；单侧则又分为左侧和右侧，所谓左右侧是指的是备择假设中是大于号就是右侧，小于号就是左侧；

备择假设和原假设的建立

在判断清楚哪一侧后就可以建立原假设和备择假设。备择假设首先建立是比较方便的。因为备择假设是由***题干中样本估计值与经验值的大小关系***决定的。如果大于就是一个右侧检验，小于就是一个左侧检验。原假设是我们用拒绝域判断真伪的假设，它和备择假设互斥。

套用公式

公式的运用规则在于能否***充分***使用题干中的所有已知条件。

分位点的构造

如果是双侧检验就用左小于 1-a/2，右大于 a/2 的方式构造出一个为 a 的概率，进而求出拒绝域。如果是左侧检验就用小于 1-a；右侧就用大于 a；

求出拒绝域

拒绝域是用来拒绝或者接受***原假设***的。如果带入数据后，刚好落在拒绝域里，说明***原假设***被拒绝，备择假设正确，反之亦然。