概率论与数理统计

期末真题题型全解析

整理人：思之心忧

题型分类参考：SAVIA 大佬的概统讲义

链接：https://savia7582.github.io/Exterior/Math/

涵盖内容：2020-2025 年期末真题

2026 年 1 月 5 日

目录

1 随机事件与概率

基本每次都会考一道填空题，考察概率的简单计算，可分成以下三种题型，基本都是送分题。

(1) 古典概型

(2) 抽象事件概率计算（集合关系）

(3) 条件概率计算（全概率公式与贝叶斯公式）

1.1 题型一：古典概型概率计算

古典概型，又称简单事件概率计算，指的是在一个样本空间中，每个事件的概率都是一个固定的常数。可能用到一些排列组合知识。

真题 1.1：23-24 春夏, 3

独立重复投掷一颗均匀的骰子 3 次，令 X 表示这三次点数之和，Y 表示三次中最小的点数，Z 表示出现 1 点的次数。则 $E(X) =$ ________，$P(Y=2) =$ ________，$P(X=6|Z=1) =$ ________。

解题思路：

(1) 单次掷骰子期望为 $\dfrac{7}{2}$，利用期望可加性求 $E(X)$；

(2) $P(Y=2)$ 表示三次点数均 2 且至少有一次 =2；

(3) $Z=1$ 表示恰有一次掷出 1 点，其余两次点数和为 5，求该情况下的概率。

详细解答：

1. 计算 $E(X)$：单次掷骰子期望 $E(\xi) = \dfrac{1+2+3+4+5+6}{6} = \dfrac{7}{2}$，故 $E(X) = 3 \times \dfrac{7}{2} = \dfrac{21}{2} =$ 10.5。

2. 计算 $P(Y=2)$：$P(Y \geq 2) = \left(\dfrac{5}{6}\right)^3$，$P(Y \geq 3) = \left(\dfrac{4}{6}\right)^3$，故 $P(Y=2) = P(Y \geq 2) - P(Y \geq 3) = \dfrac{125}{216} - \dfrac{64}{216} = \dfrac{61}{216}$。

3. 计算 $P(X=6|Z=1)$：$Z=1$ 的总情况数：$\mathrm{C}_3^1 \times 5 \times 5 = 75$（选 1 次掷 1 点，其余两次掷 2-6 点），满足 $X=6$ 的情况：其余两次点数和为 5，可能组合为 (2,3),(3,2),(4,1),(1,4),(5,0)（舍去无效组合），实际有效组合为 (2,3),(3,2)，共 $\mathrm{C}_3^1 \times 2 = 6$ 种，故 $P(X=6|Z=1) = \dfrac{6}{75} = \dfrac{2}{25}$。

故答案依次为 10.5（或 $\dfrac{21}{2}$），$\dfrac{61}{216}$，$\dfrac{2}{25}$

1.2 题型二：抽象事件概率计算

本类题主要考察集合关系的概率计算，可利用集合运算的概率性质简化计算。有些时候可以采用“正难则反”的方法，即求“不发生”的概率，再用总概率减去不发生的概率，即可得到“发生”的概率。

真题 1.2：24-25 春夏, 1

已知 $P(A)=P(B)=P(C)=\dfrac{1}{4}$，$P(AB)=0$，$P(BC)=P(AC)=\dfrac{1}{12}$。则 A,B,C 恰好有一个事件发生的概率为 __________。

解题思路：

题目要求“恰好有一个事件发生”，需利用集合运算的概率性质拆分事件。“恰好一个发生”可表示为 $A\overline{BC}\cup\overline{A}B\overline{C}\cup\overline{AB}C$，三者互斥，故概率为各事件概率之和。利用 $P(AB)=0$ 可推出 $P(ABC)=0$，进而简化单事件概率计算。

详细解答：

首先计算单个“恰好发生”事件的概率：$P(A\overline{BC})=P(A)-P(AC)-P(AB)+P(ABC)=\dfrac{1}{4}-\dfrac{1}{12}-0+0=\dfrac{1}{6}$，$P(\overline{A}B\overline{C})=P(B)-P(BC)-P(AB)+P(ABC)=\dfrac{1}{4}-\dfrac{1}{12}-0+0=\dfrac{1}{6}$，$P(\overline{AB}C)=P(C)-P(AC)-P(BC)+P(ABC)=\dfrac{1}{4}-\dfrac{1}{12}-\dfrac{1}{12}+0=\dfrac{1}{12}$。

三者互斥，故总概率为：$P(\text{恰好一个发生})=\dfrac{1}{6}+\dfrac{1}{6}+\dfrac{1}{12}=\dfrac{5}{12}$。

故答案为 $\dfrac{5}{12}$

真题 1.3：23-24 春夏, 1

已知 $P(A)=0.7$，$P(B)=0.4$，$P(A-B)=0.5$。则 $P(A\cup\overline{B})=$ ________，$P(B|\overline{A})=$ ________。

解题思路：

利用概率的基本公式推导：$P(A-B)=P(A)-P(AB)$ 可求 $P(AB)$；$P(A\cup\overline{B})$ 用并集公式，结合补集概率；$P(B|\overline{A})$ 用条件概率公式，需先求 $P(\overline{A})$ 和 $P(B\overline{A})$。

详细解答：

1. 求 $P(AB)$：由 $P(A-B)=P(A)-P(AB)$，得 $0.5=0.7-P(AB)$，故 $P(AB)=0.2$。

2. 计算 $P(A\cup\overline{B})$：$P(A\cup\overline{B})=P(A)+P(\overline{B})-P(A\overline{B})$，其中 $P(\overline{B})=1-P(B)=0.6$，$P(A\overline{B})=P(A-B)=0.5$，故 $P(A\cup\overline{B})=0.7+0.6-0.5=0.8$。

3. 计算 $P(B|\overline{A})$：$P(\overline{A}) = 1 - P(A) = 0.3$，$P(B\overline{A}) = P(B) - P(AB) = 0.4 - 0.2 = 0.2$，故 $P(B|\overline{A}) = \dfrac{P(B\overline{A})}{P(\overline{A})} = \dfrac{0.2}{0.3} = \dfrac{2}{3}$。

故答案依次为 0.8，$\dfrac{2}{3}$

真题 1.4：21-22 春夏, 1

设 A, B, C 为三个相互独立的随机事件，$P(A) = P(B) = P(C) = p$。则 A, B, C 至少有一个发生的概率为 ________。

解题思路：

“至少有一个发生”的对立事件是“全不发生”，利用独立性计算对立事件概率，再用 1 减去对立事件概率。

详细解答：

对立事件“A, B, C 全不发生”的概率为 $P(\overline{ABC}) = P(\overline{A})P(\overline{B})P(\overline{C}) = (1-p)^3$，故至少有一个发生的概率为 $1-(1-p)^3$。

故答案为 $1-(1-p)^3$

1.3 题型三：全概率公式与贝叶斯公式

本类题主要考察全概率公式与贝叶斯公式的应用，需根据题目给定的条件，利用全概率公式计算总概率，再利用贝叶斯公式计算条件概率，关键在于理解条件概率的定义。

真题 1.5：24-25 秋冬, 2

已知 $P(A) = 0.4$，$P(B|A) = 0.5$，$P(B|\overline{A}) = 0.3$，$P(C) = 0.6$，且 AB 与 C 独立。则 $P(AC|B) =$ ________。

解题思路：

利用条件概率公式 $P(AC|B) = \dfrac{P(ACB)}{P(B)}$，结合乘法公式、全概率公式计算分子分母，再利用独立性 $P(ACB) = P(AB)P(C)$。

详细解答：

1. 计算 $P(AB)$ 和 $P(\overline{A}B)$：$P(AB) = P(A)P(B|A) = 0.4 \times 0.5 = 0.2$，$P(\overline{A}B) = P(\overline{A})P(B|\overline{A}) = (1-0.4) \times 0.3 = 0.18$。

2. 由全概率公式求 $P(B)$：$P(B) = P(AB) + P(\overline{A}B) = 0.2 + 0.18 = 0.38$。

3. 利用独立性求 $P(ACB)$：因 AB 与 C 独立，故 $P(ACB)=P(AB\cap C)=P(AB)P(C)=0.2\times 0.6=0.12$。

4. 条件概率计算：$P(AC|B)=\dfrac{P(ACB)}{P(B)}=\dfrac{0.12}{0.38}=\dfrac{6}{19}$。

故答案为 $\dfrac{6}{19}$

❓ 真题 1.6：23-24 春夏, 5

某人出门去甲地。若天气好，就骑共享单车去，所花时间（单位：分钟）服从均匀分布 $U(20,40)$；若天气不好，就步行至地铁站坐地铁，所花时间服从 $U(30,50)$。已知天气好的概率为 0.8。

(1) 求此人出门半小时后还没到甲地的概率；

(2) 若已知此人出门半小时后还没到甲地，求他骑共享单车的概率。

解题思路：

(1) 用全概率公式，分天气好和天气不好两种情况计算“半小时未到”的概率；

(2) 用贝叶斯公式，结合 (1) 的结果计算条件概率。

详细解答：

(1) 设 $A=$“天气好”，$\overline{A}=$“天气不好”，$B=$“出门半小时后未到甲地”。

$P(A)=0.8$，$P(\overline{A})=0.2$；

天气好时，$P(B|A)=P(X>30)$（$X\sim U(20,40)$），均匀分布概率 $P(X>30)=\dfrac{40-30}{40-20}=0.5$；

天气不好时，$P(B|\overline{A})=P(Y>30)$（$Y\sim U(30,50)$），$P(Y>30)=\dfrac{50-30}{50-30}=1$；

全概率公式：$P(B)=P(A)P(B|A)+P(\overline{A})P(B|\overline{A})=0.8\times 0.5+0.2\times 1=0.6$。

(2) 贝叶斯公式：$P(A|B)=\dfrac{P(A)P(B|A)}{P(B)}=\dfrac{0.8\times 0.5}{0.6}=\dfrac{2}{3}$。

答案：(1) 0.6；(2) $\dfrac{2}{3}$

❓ 真题 1.7：22-23 秋冬, 1

甲盒中有 3 个红球、2 个白球，乙盒中有 2 个红球、2 个白球。从甲盒中不放回地取出 2 个球放入乙盒，再从乙盒中取出 2 个球。则：

(1) 从甲盒取出 2 个红球的概率为 ________；(2) 从乙盒取出 2 个红球的概率为 ________。

解题思路：

(1) 甲盒取 2 个红球的概率用组合数计算；

(2) 用全概率公式，分甲盒取出“2 红”“1 红 1 白”“2 白”三种情况，分别计算乙盒红球数，

再求取 2 个红球的概率。

详细解答：

(1) 甲盒取 2 个红球的概率：$P(A_1)=\dfrac{C_3^2}{C_5^2}=\dfrac{3}{10}$。

(2) 计算从乙盒取出 2 个红球的概率（记 $B=$"乙盒取 2 红"）：

需分三类情况，用全概率公式 $P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)$，

其中：$A_2=$"甲盒取 1 红 1 白"，$A_3=$"甲盒取 2 白"，三者互斥且穷尽。

第一步：计算三类先验概率：$P(A_2)=\dfrac{C_3^1C_2^1}{C_5^2}=\dfrac{3\times 2}{10}=\dfrac{6}{10}=\dfrac{3}{5}$；$P(A_3)=\dfrac{C_2^2}{C_5^2}=\dfrac{1}{10}$。

第二步：计算各类条件概率（乙盒原有 4 个球，放入 2 个后共 6 个球）：

若 A_1 发生：乙盒变为 4 红 2 白，取 2 红的组合数 $C_4^2=6$，故 $P(B|A_1)=\dfrac{C_4^2}{C_6^2}=\dfrac{6}{15}=\dfrac{2}{5}$；

若 A_2 发生：乙盒变为 3 红 3 白，取 2 红的组合数 $C_3^2=3$，故 $P(B|A_2)=\dfrac{C_3^2}{C_6^2}=\dfrac{3}{15}=\dfrac{1}{5}$；

若 A_3 发生：乙盒变为 2 红 4 白，取 2 红的组合数 $C_2^2=1$，故 $P(B|A_3)=\dfrac{C_2^2}{C_6^2}=\dfrac{1}{15}$。

第三步：代入全概率公式计算：

$$\begin{aligned}P(B)&=\frac{3}{10}\times\frac{2}{5}+\frac{3}{5}\times\frac{1}{5}+\frac{1}{10}\times\frac{1}{15}\\&=\frac{6}{50}+\frac{3}{25}+\frac{1}{150}\\&=\frac{18}{150}+\frac{18}{150}+\frac{1}{150}\\&=\frac{37}{150}\end{aligned}$$

故答案依次为 $\dfrac{3}{10}$，$\dfrac{37}{150}$

真题 1.8：21-22 秋冬, 1

一种病的得病率 4%，得病的人被查出的概率是 95%，没得病的人被查出得病的概率是 5%。已知这个人被查出得病，求他真的得病的概率为 __________。

解题思路：

用贝叶斯公式，设 $A=$"得病"，$B=$"被查出得病"，求 $P(A|B)$。

详细解答：

$P(A)=0.04$，$P(\overline{A})=0.96$，$P(B|A)=0.95$，$P(B|\overline{A})=0.05$，

由贝叶斯公式：

$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\overline{A})P(\overline{A})}=\frac{0.95\times 0.04}{0.95\times 0.04+0.05\times 0.96}=\frac{0.038}{0.038+0.048}=\frac{19}{43}$$

故答案为 $\frac{19}{43}$

真题 1.9：20-21 秋冬, 1

在甲乙的微信通讯录中，共同好友分别占甲的 30% 和乙的 20%。在甲乙中任选一人，打开其微信通讯录并随机点中一人，则点中的是他们共同朋友的概率为 ________；已知点中的是他们的共同朋友，则打开的是甲的微信通讯录的概率为 ________。

解题思路：

用全概率公式计算第一问，贝叶斯公式计算第二问，设 $A=$“选甲”，$B=$“选乙”，$C=$“点中共同好友”。

详细解答：

$P(A)=P(B)=0.5$，$P(C|A)=0.3$，$P(C|B)=0.2$，

(1) 全概率：$P(C)=0.5\times0.3+0.5\times0.2=0.25$；

(2) 贝叶斯：$P(A|C)=\dfrac{0.5\times0.3}{0.25}=0.6$。

故答案依次为 0.25，0.6

小结：

如你所见，本章内容相对简单，但知识点可能有些多，古典概型、集合关系、全概率公式、贝叶斯公式等，但都是些基本内容，希望这送分题能送到你手里。

2 一维随机变量的概率分布

本讲内容可能会考一到两题，分为离散和连续两种题型。

其中，包含对随机变量的分布律、期望、方差等的求解，也可能涉及期望方差的性质运算等内容，不熟悉者请自行复习好相关概念。

当然，还会存在既不连续也不离散的类型，可参考教材第 4 章课后题【B10】，这里不赘述。

2.1 题型四：一维离散型随机变量分布求解

真题 2.1：24-25 春夏, 2

有一生物繁育，设 X 为子代个数，已知 $P(X=0)=\frac{1}{4}$，$P(X=1)=\frac{1}{2}$，$P(X=2)=\frac{1}{4}$。X_1,X_2 分别代表第一代、第二代，每一代之间独立且满足相同分布。则 $P(X_2=0)=$ ________，$P(X_2=4)=$ ________。

解题思路：

第二代子代个数由第一代决定，利用全概率公式计算 $P(X_2=0)$（需考虑第一代所有可能个数的条件概率）；$X_2=4$ 仅当第一代有 2 个子代，且每个子代再产生 2 个子代时发生（独立事件），直接利用乘法公式。

详细解答：

1. 计算 $P(X_2=0)$：由全概率公式，考虑第一代所有可能情况：$P(X_2=0)=P(X_1=0)P(X_2=0|X_1=0)+P(X_1=1)P(X_2=0|X_1=1)+P(X_1=2)P(X_2=0|X_1=2)$ 代入已知概率：$=\frac{1}{4}\cdot 1+\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{4}\cdot\left(\frac{1}{4}\right)^2=\frac{1}{4}+\frac{1}{8}+\frac{1}{64}=\frac{25}{64}$。

2. 计算 $P(X_2=4)$：仅当 $X_1=2$ 且每个子代均产生 2 个子代时发生：$P(X_2=4)=P(X_1=2)P(X_2=4|X_1=2)=\frac{1}{4}\cdot\left(\frac{1}{4}\right)^2=\frac{1}{64}$。

故答案依次为 $\frac{25}{64}$，$\frac{1}{64}$

真题 2.2：24-25 春夏, 8

鸟等概率从三个窗口飞出去，X 为飞出窗户时尝试的次数。

(1) 如果这只鸟没记忆，求 X 的分布律；

(2) 若鸟有记忆，每个窗户最多飞一次，此时设 Y 为飞出窗户时尝试的次数，求 Y 的分布律。

解题思路：

(1) 无记忆时，每次尝试成功（飞出）概率为 $\frac{1}{3}$，失败概率为 $\frac{2}{3}$，符合几何分布；

(2) 有记忆时，尝试次数仅为 1,2,3，依次计算每次尝试成功的概率。

详细解答：

(1) 无记忆时（几何分布）：每次尝试成功概率 $p=\frac{1}{3}$，失败概率 $1-p=\frac{2}{3}$，故分布律为：$P(X=k)=\frac{1}{3}\left(\frac{2}{3}\right)^{k-1}$，$k=1,2,3,\ldots$。

(2) 有记忆时（有限次尝试）：

$P(Y=1)$：第一次就选中正确窗口，概率为 $\frac{1}{3}$；

$P(Y=2)$：第一次选错（概率 $\frac{2}{3}$），第二次选对（剩余 2 个窗口，概率 $\frac{1}{2}$），故 $P(Y=2)=\frac{2}{3}\times\frac{1}{2}=\frac{1}{3}$；

$P(Y=3)$：前两次均选错（概率 $\frac{2}{3}\times\frac{1}{2}$），第三次必选对，故 $P(Y=3)=\frac{2}{3}\times\frac{1}{2}\times 1=\frac{1}{3}$。

Y 的分布律整理如下：

Y	1	2	3
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

真题 2.3：21-22 秋冬, 2

某班级同学考试成绩均值 35，标准差 12。现对成绩做处理：先减 5，再乘以 2。则处理后成绩的均值为 ________，标准差为 ________。

解题思路：

利用期望和方差的性质：$E(aX+b)=aE(X)+b$，$Var(aX+b)=a^2Var(X)$，标准差为方差的平方根。

详细解答：

设原成绩为 X，处理后为 $Y=2(X-5)=2X-10$，

$E(Y)=2E(X)-10=2\times 35-10=60$，

$Var(Y)=4Var(X)=4\times 12^2=576$，标准差为 $\sqrt{576}=24$。

故答案依次为 60，24

注意：

这里简单考查了期望和方差的性质，“考试成绩”属于一维离散型随机变量，故归类于此。

真题 2.4：20-21 秋冬, 3

设随机变量 X 的分布函数 $F(x)=\begin{cases}0, & x<-1,\\ 0.4, & -1\le x<0,\\ 0.6, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$ 则：

(1) $P(X=0)=$ ________；

(2) 对 X 独立重复观测 3 次，至多有一次观测到“0”的概率为 ________。

解题思路：

(1) 离散型随机变量在点 x 的概率为 $F(x)-F(x^-)$；

(2) 服从二项分布 $B(3,p)$，计算 $P(Y\le 1)$。

详细解答：

(1) $P(X=0)=F(0)-F(0^-)=0.6-0.4=0.2$；

(2) $Y\sim B(3,0.2)$，$P(Y\le 1)=P(Y=0)+P(Y=1)=0.8^3+\mathrm{C}_3^1\times 0.2\times 0.8^2=0.896$。

故答案依次为 0.2，0.896

2.2 题型五：一维连续型随机变量分布求解

真题 2.5：24-25 秋冬, 1

设随机变量 X 的概率密度为

$$f(x)=\begin{cases}2xe^{-x^2}, & x>0,\\ 0, & .\end{cases}$$

则 $E\left(\dfrac{1}{X}\right)=$ __________。

解题思路：

利用期望的定义计算连续型随机变量函数的期望，积分区间为 $x>0$，直接代入公式 $E(g(X))=\int_{-\infty}^{\infty}g(x)f(x)dx$。

详细解答：

由期望定义：$E\left(\dfrac{1}{X}\right)=\int_0^{+\infty}\dfrac{1}{x}\cdot 2xe^{-x^2}dx=\int_0^{+\infty}2e^{-x^2}dx$。

已知反常积分 $\int_0^{+\infty} e^{-x^2}dx = \frac{\sqrt{\pi}}{2}$，(其实不难看出，正态分布的概率密度函数长这样)

故：$E\left(\frac{1}{X}\right) = 2 \times \frac{\sqrt{\pi}}{2} = \sqrt{\pi}$。

故答案为 $\sqrt{\pi}$

注意：

本题考察到了反常积分 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$，但熟悉正态分布概率密度公式的不难发现其中的技巧。

真题 2.6：24-25 秋冬，8(1-3)

设连续型随机变量 X 的分布函数为

$$F_X(x) = \begin{cases} ae^x, & x < 0, \\ b - ae^{-3x}, & x \geq 0. \end{cases}$$

(1) 求 a, b；

(2) 求 X 的概率密度函数 $f_X(x)$；

(3) 求 $E(X)$。

解题思路：

(1) 利用分布函数的性质：$\lim_{x\to-\infty} F_X(x) = 0$，$\lim_{x\to+\infty} F_X(x) = 1$，连续性 $F_X(0^-) = F_X(0^+)$；

(2) 密度函数为分布函数的导数；

(3) 利用期望定义积分计算。

详细解答：

(1) 求 a, b：

- $\lim_{x\to-\infty} ae^x = 0 \implies a$ 任意（后续确定）；
- $\lim_{x\to+\infty}(b - ae^{-3x}) = b = 1 \implies b = 1$；
- 连续性：$F_X(0^-) = ae^0 = a$，$F_X(0^+) = 1 - ae^0 = 1 - a$，故 $a = 1 - a \implies a = \frac{1}{2}$。

(2) 概率密度 $f_X(x)$：

- $x < 0$ 时，$f_X(x) = F_X'(x) = \frac{1}{2}e^x$；
- $x \geq 0$ 时，$f_X(x) = F_X'(x) = \frac{3}{2}e^{-3x}$；

即

$$f_X(x)=\begin{cases}\dfrac{1}{2}e^x, & x<0,\\ \dfrac{3}{2}e^{-3x}, & x\geq 0.\end{cases}$$

(3) 计算 $E(X)$：$E(X)=\int_{-\infty}^{0}x\cdot\dfrac{1}{2}e^x dx+\int_0^{+\infty}x\cdot\dfrac{3}{2}e^{-3x}dx$，

分部积分计算得：$\int_{-\infty}^{0}xe^x dx=-1$，$\int_0^{+\infty}xe^{-3x}dx=\dfrac{1}{9}$，

故 $E(X)=\dfrac{1}{2}\times(-1)+\dfrac{3}{2}\times\dfrac{1}{9}=-\dfrac{1}{2}+\dfrac{1}{6}=-\dfrac{1}{3}$。

答案： (1) $a=\dfrac{1}{2}$，$b=1$；(2) 见上述密度函数；(3) $-\dfrac{1}{3}$。

❓ 真题 2.7：22-23 秋冬, 3

设随机变量 X 的概率密度为

$$f(x)=\begin{cases}x, & 0<x<1,\\ x-3, & 3<x<4,\\ 0, & else.\end{cases}$$

则：(1) $E[\min(X,1)]=$ ________；(2) 当 $3<x<4$ 时，分布函数 $F(x)=$ ________。

解题思路：

(1) 分段计算期望，$\min(X,1)$ 在 $0<x<1$ 时为 x，$x\geq 1$ 时为 1；

(2) 分布函数 $F(x)=P(X\leq x)$，分段积分计算。

详细解答：

(1) 期望计算：

$$E[\min(X,1)]=\int_0^1 x\cdot xdx+\int_3^4 1\cdot(x-3)dx=\int_0^1 x^2dx+\int_3^4(x-3)dx=\frac{1}{3}+\frac{1}{2}=\frac{5}{6}$$

(2) 分布函数 $F(x)$（$3<x<4$）：

$$F(x)=P(X\leq x)=\int_0^1 xdx+\int_3^x(t-3)dt=\frac{1}{2}+\frac{(x-3)^2}{2}$$

故答案依次为 $\dfrac{5}{6}$，$\dfrac{1}{2}+\dfrac{(x-3)^2}{2}$

小结：

本章内容也相对基础，掌握一维随机变量的概率分布，是求解二维随机变量的基础。考察方式也相对固定，主要是根据定义计算随机变量的分布律、分布密度、期望、方差等性质。

3 二维随机变量的概率分布

本章通常也会考一到两道题，同样，也分为离散型和连续型两种题型 (不包含常见类型)。

可能会涉及到联合分布律、边缘分布律、条件分布律、期望、方差、协方差、相关性等内容。

相比一维的，二维随机变量的分布求解相对复杂，需要对随机变量的联合分布律有一定的理解，同时对二重积分有一定要求。

3.1 题型六：求解二元离散型非常见随机变量分布

? 真题 3.1：23-24 春夏, 6

设 X,Y 的联合分布律如下表所示：

$X \setminus Y$	0	1	2
0	$\frac{1}{9}$	$\frac{2}{9}$	a
1	$\frac{1}{3}$	$\frac{2}{9}$	0

(1) 求常数 a；

(2) 求 $Cov(X,Y)$，并判断 X 与 Y 的相关性；

(3) 求 $Y=1$ 的条件下，X 的条件分布函数 $F_{X|Y}(x|1)$。

解题思路：

(1) 利用联合分布律的性质（所有概率之和为 1）求解常数 a；

(2) 先计算 X,Y 的边缘分布律，再通过期望公式计算 $E(X)$、$E(Y)$、$E(XY)$，代入协方差公式判断相关性；

(3) 先求 $Y=1$ 时 X 的条件概率，再按分布函数定义分区间构造条件分布函数。

详细解答：

(1) 求常数 a：

联合分布律中所有概率之和为 1，将已知概率统一分母为 9：

$\frac{1}{9}+\frac{2}{9}+a+\frac{3}{9}+\frac{2}{9}+0=1$，化简得 $\frac{8}{9}+a=1$，解得 $a=\frac{1}{9}$。

(2) 协方差与相关性：

X 的边缘分布：$P(X=0)=\frac{1}{9}+\frac{2}{9}+\frac{1}{9}=\frac{4}{9}$，$P(X=1)=\frac{3}{9}+\frac{2}{9}+0=\frac{5}{9}$；

Y 的边缘分布：$P(Y=0)=\frac{1}{9}+\frac{3}{9}=\frac{4}{9}$，$P(Y=1)=\frac{2}{9}+\frac{2}{9}=\frac{4}{9}$，$P(Y=2)=\frac{1}{9}+0=\frac{1}{9}$；

计算期望：$E(X)=0\times\frac{4}{9}+1\times\frac{5}{9}=\frac{5}{9}$，$E(Y)=0\times\frac{4}{9}+1\times\frac{4}{9}+2\times\frac{1}{9}=\frac{6}{9}=\frac{2}{3}$，

$E(XY)=0\times 0\times\frac{1}{9}+0\times 1\times\frac{2}{9}+0\times 2\times\frac{1}{9}+1\times 0\times\frac{3}{9}+1\times 1\times\frac{2}{9}+1\times 2\times 0=\frac{2}{9}$；

协方差与相关性：$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{2}{9}-\frac{5}{9}\times\frac{2}{3}=\frac{2}{9}-\frac{10}{27}=-\frac{4}{27}$，

因 $Cov(X,Y)\neq 0$，故 X 与 Y 相关。

(3) 条件分布函数 $F_{X|Y}(x|1)$：

计算条件概率：$P(X=0|Y=1)=\frac{P(X=0,Y=1)}{P(Y=1)}=\frac{\frac{2}{9}}{\frac{4}{9}}=\frac{1}{2}$，

$P(X=1|Y=1)=\frac{P(X=1,Y=1)}{P(Y=1)}=\frac{\frac{2}{9}}{\frac{4}{9}}=\frac{1}{2}$；

构造条件分布函数：根据分布函数定义 $F_{X|Y}(x|1)=P(X\leq x|Y=1)$，分区间得：

$$F_{X|Y}(x|1)=\begin{cases}0, & x<0,\\ \frac{1}{2}, & 0\leq x<1,\\ 1, & x\geq 1.\end{cases}$$

答案： (1) $a=\frac{1}{9}$；(2) $Cov(X,Y)=-\frac{4}{27}$，X 与 Y 相关；(3) 见上述条件分布函数。

? 真题 3.2：21-22 春夏，6

假设投掷以概率 $p\left(\frac{1}{2}<p<1\right)$ 出现正面的非均匀硬币 2 次，出现正面的次数记为 X。在 2 次投掷中，每当出现正面时，立即投掷另一枚以概率 $\frac{1}{2p}$ 出现正面的硬币，记出现的正面次数为 Y。

(1) 求 (X,Y) 的联合分布律；

(2) 分别求 X,Y 的边际分布律；

(3) 判断 X,Y 是否独立？说明理由。

解题思路：

(1) X 的可能取值为 0,1,2，Y 的取值随 X 变化（$X=0$ 时 $Y=0$；$X=1$ 时 $Y=0,1$；$X=2$ 时 $Y=0,1,2$），利用乘法公式计算联合概率；

(2) 边际分布律由联合分布律累加得到；

(3) 验证是否满足 $P(X=i,Y=j)=P(X=i)P(Y=j)$。

详细解答：

(1) 联合分布律：

$P(X=0,Y=0)=P(X=0)=(1-p)^2$（无正面，不投第二枚）；

$P(X=1,Y=0)=P(X=1)P(Y=0|X=1)=\mathrm{C}_2^1p(1-p)\times\left(1-\frac{1}{2p}\right)=2p(1-p)\times\frac{2p-1}{2p}=(1-p)(2p-1)$；

$P(X=1,Y=1)=P(X=1)P(Y=1|X=1)=2p(1-p)\times\frac{1}{2p}=1-p$；

$P(X=2,Y=0)=P(X=2)P(Y=0|X=2)=p^2\times\left(1-\frac{1}{2p}\right)^2=p^2\times\frac{(2p-1)^2}{4p^2}=\frac{(2p-1)^2}{4}$；

$P(X=2,Y=1)=P(X=2)P(Y=1|X=2)=p^2\times\mathrm{C}_2^1\times\frac{1}{2p}\times\left(1-\frac{1}{2p}\right)=p^2\times2\times\frac{2p-1}{4p^2}=\frac{2p-1}{2}$；

$P(X=2,Y=2)=P(X=2)P(Y=2|X=2)=p^2\times\left(1-\frac{1}{2p}\right)^2=p^2\times\frac{1}{4p^2}=\frac{1}{4}$。

整理为表格：

$X\setminus Y$	0	1	2
0	$(1-p)^2$	0	0
1	$(1-p)(2p-1)$	$1-p$	0
2	$\frac{(2p-1)^2}{4}$	$\frac{2p-1}{2}$	$\frac{1}{4}$

(2) X 的边际分布：

$P(X=0)=(1-p)^2$，

$P(X=1)=(1-p)(2p-1)+(1-p)=2(1-p)p$，

$P(X=2)=\frac{(2p-1)^2}{4}+\frac{2p-1}{2}+\frac{1}{4}=p^2$；

Y 的边际分布：

$P(Y=0)=(1-p)^2+(1-p)(2p-1)+\frac{(2p-1)^2}{4}=\frac{(2-2p+2p-1)^2}{4}=\frac{1}{4}$，

$P(Y=1)=(1-p)+\frac{2p-1}{2}=\frac{2-2p+2p-1}{2}=\frac{1}{2}$，

$P(Y=2)=\frac{1}{4}$。

(3) 独立性判断：取 $X=0,Y=0$，$P(X=0,Y=0)=(1-p)^2$，$P(X=0)P(Y=0)=(1-p)^2\times\frac{1}{4}$，因 $\frac{1}{2}<p<1$，$(1-p)^2\neq\frac{1}{4}(1-p)^2$，故 X,Y 不独立。

答案：(1) 见上述联合分布律；(2) X：$P(X=0)=(1-p)^2$，$P(X=1)=2p(1-p)$，$P(X=2)=p^2$；Y：$P(Y=0)=\frac{1}{4}$，$P(Y=1)=\frac{1}{2}$，$P(Y=2)=\frac{1}{4}$；(3) 不独立，理由见上述分析。

真题 3.3：21-22 秋冬，4

设 (X, Y) 的联合分布律如下表所示：

$X \setminus Y$	-1	0	1
-1	a	0	0
0	0	$2a$	0
1	0	0	a

(1) 求常数 a；

(2) 设 $Z = \max\{X, Y\}$，求 Z 的分布函数；

(3) 求 X, Y 的相关系数，判断是否相关。

解题思路：

(1) 联合分布律概率和为 1;

(2) Z 的可能取值为-1,0,1，分区间求分布函数；

(3) 计算协方差，非零则相关。

详细解答：

(1) 由对称性和概率和为 1：$a + 2a + a = 1 \implies a = \dfrac{1}{4}$。

(2) Z 的可能取值为-1,0,1

$F_Z(z) = 0$，$z < -1$；

$F_Z(z) = P(Z = -1) = P(X = -1, Y = -1) = 0.25$，$-1 \le z < 0$；

$F_Z(z) = P(Z \le 0) = 0.25 + 0.5 = 0.75$，$0 \le z < 1$；

$F_Z(z) = 1$，$z \ge 1$。

(3)$E(X) = E(Y) = 0$，$E(XY) = 0.5$，$Var(X) = Var(Y) = 0.5$，$Cov(X, Y) = 0.5$

相关系数为 $\dfrac{0.5}{\sqrt{0.5^2 \times 0.5^2}} = 1$，故相关。

答案：(1) $a = 0.25$；(2) $F_Z(z) = \begin{cases} 0, & z < -1, \\ 0.125, & -1 \le z < 0, \\ 0.625, & 0 \le z < 1, \\ 1, & z \ge 1 \end{cases}$；(3) 相关系数为 1，相关。

3.2 题型七：求解二元连续型非常见随机变量分布

真题 3.4：24-25 春夏, 9

已知二维随机变量 (X, Y) 的联合概率密度为：

$$f(x,y) = \begin{cases} 8xy, & 0 < y < x < 1, \\ 0, & . \end{cases}$$

(1) 求 $P(X < 2Y)$；

(2) 求 $f_X(x)$ 和 $f_Y(y)$，并说明 X 和 Y 是否独立；

(3) 求 $Cov(X, Y)$，并说明 X 和 Y 是否相关。

解题思路：

(1) 利用二重积分计算区域概率，确定积分限为 $0 < x < 1$，$\frac{x}{2} < y < x$；

(2) 边缘密度通过联合密度积分得到，独立性判定：验证 $f(x, y) = f_X(x)f_Y(y)$ 是否恒成立；

(3) 相关系数的核心是计算协方差，利用 $Cov(X, Y) = E(XY) - E(X)E(Y)$，通过积分求期望。

详细解答：

(1) 计算 $P(X < 2Y)$：

积分区域为 $0 < x < 1$，$\frac{x}{2} < y < x$，故：

$P(X < 2Y) = \int_0^1 \int_{\frac{x}{2}}^{x} 8xydydx$

先对 y 积分：$\int_{\frac{x}{2}}^{x} 8xydy = 8x \cdot \left[\frac{y^2}{2}\right]_{\frac{x}{2}}^{x} = 4x\left(x^2 - \frac{x^2}{4}\right) = 3x^3$，再对 x 积分：$\int_0^1 3x^3dx = \frac{3}{4}$。

(2) 求边缘密度并判定独立性：

关于 X 的边缘密度（$0 < x < 1$）：$f_X(x) = \int_0^x 8xydy = 8x \cdot \frac{x^2}{2} = 4x^3$，其他区域 $f_X(x) = 0$；

关于 Y 的边缘密度（$0 < y < 1$）：$f_Y(y) = \int_y^1 8xydx = 8y \cdot \left[\frac{x^2}{2}\right]_y^1 = 4y(1 - y^2)$，其他区域 $f_Y(y) = 0$；

独立性判定：取 $x = 0.5$，$y = 0.2$，$f(0.5, 0.2) = 8 \times 0.5 \times 0.2 = 0.8$，

而 $f_X(0.5)f_Y(0.2) = 4 \times 0.125 \times 4 \times 0.2 \times (1 - 0.04) = 0.5 \times 0.768 = 0.384 \neq 0.8$，故 X 和 Y 不独立。

(3) 计算协方差并判定相关性：

首先计算期望：

$E(X)=\int_0^1 x\cdot 4x^3dx=4\int_0^1 x^4dx=\frac{4}{5}$，
$E(Y)=\int_0^1 y\cdot 4y(1-y^2)dy=4\left(\int_0^1 y^2dy-\int_0^1 y^4dy\right)=4\left(\frac{1}{3}-\frac{1}{5}\right)=\frac{8}{15}$，
$E(XY)=\int_0^1\int_0^x xy\cdot 8xydydx=8\int_0^1 x^2dx\int_0^x y^2dy=8\times\frac{1}{3}\int_0^1 x^5dx=\frac{4}{9}$。

协方差：

$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{4}{9}-\frac{4}{5}\times\frac{8}{15}=\frac{4}{225}\neq 0$，

故 X 和 Y 相关。

答案： (1) $\frac{3}{4}$；(2) $f_X(x)=4x^3(0<x<1)$，$f_Y(y)=4y(1-y^2)(0<y<1)$，不独立；(3) $Cov(X,Y)=\frac{4}{225}$，相关。

? 真题 3.5：24-25 秋冬, 9(1-2)

设二元随机变量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}12y^{-5}, & 1<x<y,\\ 0, & .\end{cases}$$

(1) 求 $f_X(x)$，$f_Y(y)$，并判断 X,Y 是否独立；

(2) 求 $P(1<X<1.5|Y=2)$；

(3) 设 $Z=X+Y$，求 Z 的概率密度函数 $f_Z(z)$。

解题思路：

(1) 边缘密度通过联合密度积分得到，独立性判定需验证 $f(x,y)=f_X(x)f_Y(y)$；

(2) 条件密度 $f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$，再积分求概率；

详细解答：

(1) 边缘密度与独立性：

$f_X(x)$（$x>1$）：$f_X(x)=\int_x^{+\infty}12y^{-5}dy=12\cdot\left[\frac{y^{-4}}{-4}\right]_x^{+\infty}=3x^{-4}$，其它为 0；

$f_Y(y)$（$y>1$）：$f_Y(y)=\int_1^y 12y^{-5}dx=12y^{-5}(y-1)$，其它为 0；

独立性:取 $x=2$,$y=3$,$f(2,3)=12\times3^{-5}=\frac{12}{243}$,$f_X(2)f_Y(3)=3\times2^{-4}\times12\times3^{-5}\times(3-1)\neq\frac{12}{243}$，故不独立。

(2) 条件概率 $P(1<X<1.5|Y=2)$：

$f_Y(2)=12\times2^{-5}\times(2-1)=\frac{12}{32}=\frac{3}{8}$；

条件密度 $f_{X|Y}(x|2)=\frac{f(x,2)}{f_Y(2)}=\frac{12\times2^{-5}}{3/8}=\frac{12/32}{3/8}=1$（$1<x<2$）；

概率：$P(1<X<1.5|Y=2)=\int_1^{1.5}1dx=0.5$。

答案： (1) $f_X(x)=3x^{-4}(x>1)$，$f_Y(y)=12y^{-5}(y-1)(y>1)$，不独立；(2) 0.5；

? 真题 3.6：23-24 春夏, 7

设 X,Y 的联合密度函数为

$$f(x,y)=\begin{cases}e^{-x}, & x>0\ x<y<x+1,\\ 0, & .\end{cases}$$

(1) 分别求 X 与 Y 的边际密度函数 $f_X(x)$ 与 $f_Y(y)$，并判断两者是否独立；

(2) 求 $P(X<0.5|Y=1)$；

(3) 求 $E(XY)$。

解题思路：

(1) 边际密度通过联合密度积分得到，独立性验证 $f(x,y)=f_X(x)f_Y(y)$；

(2) 先求条件密度 $f_{X|Y}(x|1)$，再积分求概率；

(3) 利用二重积分计算 $E(XY)=\iint xyf(x,y)dxdy$。

详细解答：

(1) 边际密度：

$f_X(x)$（$x>0$）：$f_X(x)=\int_x^{x+1}e^{-x}dy=e^{-x}(x+1-x)=e^{-x}$，其他为 0；

$f_Y(y)$：

当 $0<y<1$ 时，$f_Y(y)=\int_0^y e^{-x}dx=1-e^{-y}$；

当 $y\geq 1$ 时，$f_Y(y)=\int_{y-1}^{y}e^{-x}dx=e^{-(y-1)}-e^{-y}$；

其他为 0；

独立性：取 $x=0.5$，$y=0.8$，$f(0.5,0.8)=e^{-0.5}$，$f_X(0.5)f_Y(0.8)=e^{-0.5}(1-e^{-0.8})\neq e^{-0.5}$，故不独立。

(2) 条件概率 $P(X<0.5|Y=1)$：

$f_Y(1)=e^0-e^{-1}=1-e^{-1}$；

条件密度 $f_{X|Y}(x|1)=\dfrac{f(x,1)}{f_Y(1)}=\dfrac{e^{-x}}{1-e^{-1}}$（$0<x<1$）；

概率：$P(X<0.5|Y=1)=\int_0^{0.5}\dfrac{e^{-x}}{1-e^{-1}}dx=\dfrac{1-e^{-0.5}}{1-e^{-1}}=\dfrac{e-e^{0.5}}{e-1}$。

(3) 计算 $E(XY)$：

$$E(XY)=\int_{x=0}^{+\infty}\int_{y=x}^{x+1}xye^{-x}dydx$$

先对 y 积分：$\int_x^{x+1}ydy=\dfrac{(x+1)^2-x^2}{2}=x+\dfrac{1}{2}$，

再对 x 积分：$\int_0^{+\infty}x\left(x+\dfrac{1}{2}\right)e^{-x}dx=\int_0^{+\infty}x^2e^{-x}dx+\dfrac{1}{2}\int_0^{+\infty}xe^{-x}dx=2+\dfrac{1}{2}\times 1=\dfrac{5}{2}$。

答案：(1) $f_X(x)=\begin{cases}e^{-x}, & x>0,\\ 0, & \text{其他}\end{cases}$，$f_Y(y)=\begin{cases}1-e^{-y}, & 0<y<1,\\ e^{-(y-1)}-e^{-y}, & y\geq 1,\\ 0, & \text{其他}\end{cases}$，不独立；

(2) $\dfrac{e-e^{0.5}}{e-1}$；(3) $\dfrac{5}{2}$

真题 3.7：22-23 秋冬，7

设二元随机变量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{xy+1}{3}, & 0<x<1,0<y<2,\\ 0, & else.\end{cases}$$

(1) 求 $P(X>Y)$；

(2) 求 $f_X(x)$，$f_Y(y)$，并判断 X,Y 是否独立；

(3) 设 $M=\max\{X,Y\}$，求 $F_M(m)$ 与 $f_M(m)$。

解题思路：

(1) 积分区域 $0<y<x<1$，计算二重积分；

(2) 边缘密度积分得到，独立性验证 $f(x,y)=f_X(x)f_Y(y)$；

(3) 分布函数 $F_M(m)=P(\max\{X,Y\}\leq m)$，分区间计算。

详细解答：

(1) 计算 $P(X>Y)$：

$$P(X>Y)=\int_0^1\int_0^x\frac{xy+1}{3}dydx=\frac{1}{3}\left(\int_0^1 x\cdot\frac{x^2}{2}dx+\int_0^1 xdx\right)=\frac{1}{3}\left(\frac{1}{8}+\frac{1}{2}\right)=\frac{5}{24}$$

(2) $f_X(x)$（$0<x<1$）：$f_X(x)=\int_0^2\dfrac{xy+1}{3}dy=\dfrac{1}{3}(x\cdot 2+2)=\dfrac{2x+2}{3}$；

$f_Y(y)$（$0<y<2$）：$f_Y(y)=\int_0^1\dfrac{xy+1}{3}dx=\dfrac{1}{3}\left(y\cdot\dfrac{1}{2}+1\right)=\dfrac{y+2}{6}$；

独立性：$f(x,y)\neq f_X(x)f_Y(y)$，故不独立。

(3) 当 $m \le 0$ 时，$F_M(m) = 0$;

当 $0 < m < 1$ 时，$F_M(m) = P(X \le m, Y \le m) = f_X(x)f_Y(y)$ 积分得 $\dfrac{m^2(m^2+4)}{12}$；

当 $1 \le m < 2$ 时，$F_M(m) = P(X \le 1, Y \le m) = \dfrac{m^2+4m}{12}$；

当 $m \ge 2$ 时，$F_M(m) = 1$；

密度 $f_M(m) = F'_M(m)$：

$$f_M(m) = \begin{cases} \dfrac{m^3+2m}{3}, & 0 < m < 1, \\ \dfrac{m+2}{6}, & 1 \le m < 2, \\ 0, & else. \end{cases}$$

答案：(1) $\dfrac{5}{24}$；(2) $f_X(x) = \dfrac{2x+2}{3}(0 < x < 1)$，$f_Y(y) = \dfrac{y+2}{6}(0 < y < 2)$，不独立；(3) 见上述 $F_M(m)$ 和 $f_M(m)$。

❓ 真题 3.8：21-22 秋冬, 6

设二元随机变量 (X, Y) 的联合密度函数为

$$f(x,y) = \begin{cases} \dfrac{1}{2}(1-xy), & -1 < x < 1, 0 < y < 1, \\ 0, & else. \end{cases}$$

(1) 求 $P(Y \ge X)$；

(2) 求条件密度 $f_{Y|X}(y|x)$，并计算 $P(X \le Y | X = 0.5)$；

(3) 判断 $|X|$ 与 Y 是否独立。

解题思路：

(1) 分 $x < 0$ 和 $x \ge 0$ 积分计算概率；

(2) 先求边缘密度 $f_X(x)$，再得条件密度，积分求条件概率；

(3) 验证 $f_{|X|}(|x|)f_Y(y) = f_{|X|,Y}(x,y)$。

详细解答：

(1) 按 x 的范围拆分积分区域，联合密度非零区域为 $x \in (-1,1)$，$y \in (0,1)$：

当 $x \in (-1, 0)$ 时，$Y \ge X$ 恒成立（因 $y \ge 0 > x$），积分区域为 $y \in (0,1)$；

当 $x \in (0,1)$ 时，$Y \ge X$ 对应 $y \in (x, 1)$。

计算积分：

$$P(Y \geq X) = \int_{x=-1}^{0} \int_{y=0}^{1} \frac{1}{2}(1-xy)dydx + \int_{x=0}^{1} \int_{y=x}^{1} \frac{1}{2}(1-xy)dydx$$

$$= I_1 + I_2$$

计算 I_1（$x \in (-1,0)$）：

$$\int_{y=0}^{1} \frac{1}{2}(1-xy)dy = \frac{1}{2}\left[y - \frac{xy^2}{2}\right]_0^1 = \frac{1}{2}\left(1 - \frac{x}{2}\right)$$

$$I_1 = \int_{-1}^{0} \frac{1}{2}\left(1-\frac{x}{2}\right)dx = \frac{1}{2}\left[x - \frac{x^2}{4}\right]_{-1}^{0} = \frac{1}{2}\left(0 - \left(-1 - \frac{1}{4}\right)\right) = \frac{5}{8}$$

计算 I_2（$x \in (0,1)$）：

$$\int_{y=x}^{1} \frac{1}{2}(1-xy)dy = \frac{1}{2}\left[y - \frac{xy^2}{2}\right]_x^1 = \frac{1}{2}\left(1 - \frac{x}{2} - x + \frac{x^3}{2}\right) = \frac{1}{2}\left(1 - \frac{3x}{2} + \frac{x^3}{2}\right)$$

$$I_2 = \int_0^1 \frac{1}{2}\left(1 - \frac{3x}{2} + \frac{x^3}{2}\right)dx = \frac{1}{2}\left[x - \frac{3x^2}{4} + \frac{x^4}{8}\right]_0^1 = \frac{1}{2}\left(1 - \frac{3}{4} + \frac{1}{8}\right) = \frac{3}{16}$$

总概率：$P(Y \geq X) = \frac{5}{8} + \frac{3}{16} = \frac{13}{16}$。

(2) X 的边缘密度 $f_X(x)$：

$$f_X(x) = \int_{y=0}^{1} \frac{1}{2}(1-xy)dy = \frac{1}{2}\left(1 - \frac{x}{2}\right), \quad -1 < x < 1$$

其他区域 $f_X(x) = 0$。

条件密度 $f_{Y|X}(y|x)$：

当 $-1 < x < 1$ 且 $0 < y < 1$ 时：

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{1}{2}(1-xy)}{\frac{1}{2}\left(1-\frac{x}{2}\right)} = \frac{2-2xy}{2-x}$$

其他区域 $f_{Y|X}(y|x) = 0$。

当 $X = 0.5$ 时，条件密度为：

$$f_{Y|X}(y|0.5) = \frac{1-0.5y}{1-0.25} = \frac{4(1-0.5y)}{3} = \frac{4-2y}{3}, \quad 0 < y < 1$$

条件概率为 $Y \geq 0.5$ 的积分：

$$P(X \leq Y|X=0.5) = \int_{0.5}^{1} \frac{4-2y}{3}dy = \frac{1}{3}\left[4y - y^2\right]_{0.5}^{1} = \frac{1}{3}(3 - 1.75) = \frac{5}{12}$$

(3) Y 的边缘密度 $f_Y(y)$：

$$f_Y(y) = \int_{x=-1}^{1} \frac{1}{2}(1-xy)dx = \frac{1}{2}\left[x - \frac{yx^2}{2}\right]_{-1}^{1} = \frac{1}{2}(2-0) = 1, \quad 0 < y < 1$$

其他区域 $f_Y(y)=0$。

$|X|$ 的取值范围为 $0<u<1$，对应 $x=u$ 和 $x=-u$：

$$f_{|X|}(u)=f_X(u)+f_X(-u)=\frac{1-\frac{u}{2}}{2}+\frac{1-\frac{-u}{2}}{2}=1,\quad 0<u<1$$

其他区域 $f_{|X|}(u)=0$。

计算 $F_{|X|,Y}(u,y)$：

$$F_{|X|,Y}(u,y)=P(|X|\le u,Y\le y)=\int_{-u}^{u}\int_{0}^{y}\frac{1}{2}(1-xy)dxdy=uy,\quad 0<u<1,0<y<1$$

即 $F_{|X|,Y}(x,y)=xy$, 所以 $f_{|X|,Y}(x,y)=\dfrac{\partial^2 F}{\partial x\partial y}=1$

$f_{|X|}(|x|)f_Y(y)=f_{|X|,Y}(x,y)$ 成立，故 $|X|$ 与 Y 独立。

答案： (1) $P(Y\ge X)=\dfrac{13}{16}$；(2) 条件密度 $f_{Y|X}(y|x)=\dfrac{2-2xy}{2-x}$（$-1<x<1,0<y<1$），$P(X\le Y|X=0.5)=\dfrac{5}{12}$；(3) $|X|$ 与 Y 独立。

? 真题 3.9：20-21 秋冬, 8

设随机变量 X 的概率密度函数 $f_X(x)=\begin{cases}6x(1-x), & 0<x<1,\\ 0, & else.\end{cases}$ 当 $X=x$ 时，Y 在区间 $(0,x)$ 上服从均匀分布。

(1) 求 $P(X+Y<1)$；

(2) 求边际密度函数 $f_Y(y)$；

(3) 求条件密度函数 $f_{X|Y}(x|0.5)$。

解题思路：

(1) 联合密度 $f(x,y)=f_X(x)f_{Y|X}(y|x)$，积分计算概率；

(2) 对 x 积分得边际密度；

(3) 先求 $f_Y(0.5)$，再得条件密度。

详细解答：

联合密度 $f(x,y)=6x(1-x)\times\dfrac{1}{x}=6(1-x)$（$0<y<x<1$），

(1)$P(X+Y<1)=\int_0^{0.5}\int_y^{1-y}6(1-x)dxdy=0.75$。

(2) 边际密度 $f_Y(y)=\int_y^1 6(1-x)dx=6\times\dfrac{(1-y)^2}{2}=3(1-y)^2$（$0<y<1$）;$f_Y(y)=0,else$。

故 $f_Y(y)=\begin{cases}3(1-y)^2, & 0<y<1,\\ \\ 0, & else.\end{cases}$

(3) $f_Y(0.5)=3(1-0.5)^2=0.75$，

条件密度 $f_{X|Y}(x|0.5)=\dfrac{f(x,0.5)}{f_Y(0.5)}=\dfrac{6(1-x)}{0.75}=8(1-x)$（$0.5<x<1$）；$f_{X|Y}(x|0.5)=0, else$。

答案：(1) 0.75；(2) 见上；(3) 见上。

</> 小结：

上一章与本章内容可能相对重要，建议在复习时重点关注，其中难点在于理解二维随机变量的关系，灵活掌握运用方差、协方差、相关性等公式。

考试时可能会出两道计算题，分别考察一维、二维，离散、连续的一个组合题，同时可能涉及随机变量函数内容。

4 常见随机变量的概率分布

按照 SAVIA 大佬的概统讲义，本章将常见随机变量的概率分布单独列成一章，其本质上属于第二、三章节内容，主要概率分布包括：二项分布、泊松分布、均匀分布、指数分布、正态分布、二元正态分布等。要求掌握其概率密度函数、分布函数、期望、方差等公式，其中二元正态分布还涉及到相关系数、协方差、线性不变性等。

4.1 题型八：常见随机变量分布求解

真题 4.1：24-25 秋冬, 3

设 X, Y 服从参数为 2 的泊松分布，且 X 与 Y 相互独立。则 $P(X \le 1) =$ ________，$P(X = 0|X+Y=2) =$ ________。

解题思路：

泊松分布概率公式为 $P(X=k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$；条件概率需先求 $P(X+Y=2)$，利用独立泊松变量和的分布仍为泊松分布（参数相加）。

详细解答：

1. 计算 $P(X \le 1)$：$\lambda = 2$，故：$P(X \le 1) = P(X=0) + P(X=1) = e^{-2} + \dfrac{2e^{-2}}{1!} = 3e^{-2}$。

2. 计算 $P(X=0|X+Y=2)$：因 X, Y 独立，$X+Y \sim P(4)$，故：$P(X+Y=2) = \dfrac{4^2 e^{-4}}{2!} = 8e^{-4}$，$P(X=0, X+Y=2) = P(X=0, Y=2) = P(X=0)P(Y=2) = e^{-2} \times \dfrac{2^2 e^{-2}}{2!} = 2e^{-4}$。

条件概率：$P(X=0|X+Y=2) = \dfrac{2e^{-4}}{8e^{-4}} = \dfrac{1}{4}$。

故答案依次为 $3e^{-2}$，$\dfrac{1}{4}$

真题 4.2：24-25 秋冬, 5

设二元正态变量 $(X, Y) \sim N(1, 0, 16, 25, 0.25)$，则 $2X - Y - 1 \sim$ __________ 分布（写出参数）。

解题思路：

二元正态变量的线性组合仍为正态分布，需计算期望和方差。参数含义：$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，其中 ρ 为相关系数。

详细解答：

1. 计算期望 $E(2X-Y-1)$：$E(2X-Y-1) = 2E(X) - E(Y) - 1 = 2 \times 1 - 0 - 1 = 1$。

2. 计算方差 $Var(2X-Y-1)$：$Var(2X-Y)=4Var(X)+Var(Y)-4Cov(X,Y)$，$Cov(X,Y)=\rho\sigma_1\sigma_2=0.25\times4\times5=5$，故 $Var(2X-Y-1)=4\times16+25-4\times5=64+25-20=69$。

因此 $2X-Y-1\sim N(1,69)$。

故答案为 $N(1,69)$

? 真题 4.3：23-24 春夏，2

若公交车站候车人数 $X\sim P(\lambda)$（泊松分布），且已知恰有 1 人候车的概率与恰有 2 人候车的概率相等。则 $E(X^2)=$ ________，该公交车站至少有 4 人候车的概率为 ________，现已知至少 4 人候车，则恰有 4 人候车的概率为 ________。

解题思路：

泊松分布概率公式 $P(X=k)=\dfrac{\lambda^k e^{-\lambda}}{k!}$，由 $P(X=1)=P(X=2)$ 求 λ；$E(X^2)=Var(X)+[E(X)]^2$（泊松分布 $Var(X)=\lambda$）；后两问用条件概率和泊松概率累加计算。

详细解答：

1. 求 λ：由 $P(X=1)=P(X=2)$，得 $\dfrac{\lambda e^{-\lambda}}{1!}=\dfrac{\lambda^2 e^{-\lambda}}{2!}$，解得 $\lambda=2$。

2. 计算 $E(X^2)$：$E(X)=\lambda=2$，$Var(X)=\lambda=2$，故 $E(X^2)=2+2^2=6$。

3. 至少 4 人候车的概率：$P(X\geq4)=1-P(X=0)-P(X=1)-P(X=2)-P(X=3)$，代入 $\lambda=2$，得 $P(X\geq4)=1-e^{-2}(1+2+2+\frac{8}{6})=1-\frac{19}{3}e^{-2}$。

4. 已知至少 4 人候车时恰有 4 人的概率：$P(X=4|X\geq4)=\dfrac{P(X=4)}{P(X\geq4)}=\dfrac{\frac{2^4e^{-2}}{4!}}{1-\frac{19}{3}e^{-2}}=\dfrac{2}{3e^2-19}$。

故答案依次为 6，$1-\dfrac{19}{3}e^{-2}$，$\dfrac{2}{3e^2-19}$

? 真题 4.4：22-23 秋冬，2(1)

X 服从参数为 1.5 的泊松分布。则：(1) $Var(2X-4)=$ ________；

解题思路：

(1) 泊松分布 $Var(X)=\lambda$，方差性质 $Var(aX+b)=a^2Var(X)$；

详细解答：

(1) 方差计算：$Var(X)=1.5$，故 $Var(2X-4)=4\times1.5=6$。

故答案为 6

? 真题 4.5：22-23 秋冬, 4

设 $(X,Y)\sim N(1,2,1,4,\frac{1}{2})$（二元正态分布）。则：

(1) 边缘密度 $f_X(x)=$ ________；

(2) 当 $a=$ ________ 时，$aX-Y$ 与 $X+Y$ 相互独立；

(3) $3\left(\dfrac{X-1}{Y-X-1}\right)^2$ 服从 ________ 分布（写出参数）。

解题思路：

(1) 二元正态边缘分布为正态，参数取对应均值和方差；

(2) 独立的充要条件是协方差为 0；

(3) 构造独立正态变量，比值的平方服从 F 分布。

详细解答：

(1) 边缘密度 $f_X(x)$：$X\sim N(1,1)$，故 $f_X(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\dfrac{(x-1)^2}{2}}$。

(2) 求 a：$Cov(aX-Y,X+Y)=aVar(X)+(a-1)Cov(X,Y)-Var(Y)=0$，

代入 $Cov(X,Y)=\rho\sigma_X\sigma_Y=1$，得 $a\times 1+(a-1)\times 1-4=0$，解得 $a=2.5$（或 $\frac{5}{2}$）。

(3) 分布判断：令 $U=X-1\sim N(0,1)$，$V=Y-X-1\sim N(0,3)$（计算期望和方差得），且 U,V 独立，故 $\dfrac{U}{\sqrt{V/3}}\sim t(1)$，平方后乘以 3 得 $F(1,1)$ 分布。

故答案依次为 (1) $\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}$；(2) $\frac{5}{2}$；(3) $F(1,1)$

? 真题 4.6：22-23 秋冬, 6

设 X,Y 同服从 $B(2,0.5)$（二项分布），且 $P(XY\le 1)=1$。

(1) 求 X,Y 的联合分布律；(2) 求 $F(1.5,1)$；(3) 求 X,Y 的相关系数。

解题思路：

(1) 由 $P(XY\le 1)=1$ 知 $P(X=2,Y=2)=0$，结合边缘分布律求联合概率；

(2) 分布函数 $F(1.5,1)=P(X\le 1.5,Y\le 1)$；

(3) 计算协方差和标准差，求相关系数。

详细解答：

(1) 边缘分布：$P(X=0)=P(X=2)=0.25$，$P(X=1)=0.5$（同理 Y）；

结合已知条件 $P(XY\le 1)=1$ 推导得联合分布律：

$X \setminus Y$	0	1	2
0	0	0	0.25
1	0	0.5	0
2	0.25	0	0

(2) $F(1.5,1) = P(X \le 1.5, Y \le 1) = P(X=0,Y=0) + P(X=0,Y=1) + P(X=1,Y=0) + P(X=1,Y=1) = 0+0+0+0.5 = 0.5$。

(3) 期望：$E(X) = E(Y) = 1$，$E(XY) = 0 \times 0.5 + 1 \times 0.5 = 0.5$；

方差：$Var(X) = Var(Y) = 0.5$；

协方差：$Cov(X,Y) = 0.5 - 1 \times 1 = -0.5$；

相关系数：$\rho = \dfrac{-0.5}{\sqrt{0.5 \times 0.5}} = -1$。

答案： (1) 见上述联合分布律；(2) 0.5；(3) -1

真题 4.7：21-22 春夏, 2

设随机变量 X 在区间 $[-1,2]$ 上服从均匀分布。则当 $-1 < x < 2$ 时，分布函数 $F(x) =$ __________。

解题思路：

均匀分布的分布函数为 $F(x) = \dfrac{x-a}{b-a}$（$a<x<b$），其中 a,b 为区间端点。

详细解答：

$X \sim U(-1,2)$，区间长度为 $2-(-1)=3$，故当 $-1<x<2$ 时，$F(x) = \dfrac{x-(-1)}{3} = \dfrac{x+1}{3}$。

故答案为 $\dfrac{x+1}{3}$

真题 4.8：21-22 春夏, 3

某大型计算机在任何长度为 t（单位：天）的时间段内故障的次数 N_t 服从参数为 $\dfrac{t}{72}$ 的泊松分布。则：

(1) 72 天内至少发生 2 次故障的概率为 ________；

(2) 从年初一开始到第一次发生故障的时间为 T（天），则当 $t>0$ 时，$P(T>t) =$ ________；

(3) $E(T) =$ ________。

解题思路：

(1) 72 天内故障次数 $N_{72} \sim P(1)$，计算 $P(N_{72} \geq 2)$；

(2) T 为首次故障时间，服从指数分布，$P(T > t) = P(N_t = 0)$；

(3) 指数分布的期望为参数的倒数。

详细解答：

(1) $N_{72} \sim P\left(\frac{72}{72}\right) = P(1)$，

$P(N_{72} \geq 2) = 1 - P(N_{72} = 0) - P(N_{72} = 1) = 1 - e^{-1} - e^{-1} = 1 - 2e^{-1}$；

(2) $P(T > t) = P(N_t = 0) = \dfrac{\left(\frac{t}{72}\right)^0 e^{-\frac{t}{72}}}{0!} = e^{-\frac{t}{72}}$；

(3) 由 (2) 知 $T \sim Exp\left(\frac{1}{72}\right)$，故 $E(T) = 72$。

故答案依次为 $1 - 2e^{-1}$，$e^{-\frac{t}{72}}$，72

真题 4.9：21-22 春夏, 4

设 $(X, Y) \sim N(1, 1, 1, 1, 0.5)$（二元正态分布）。则：

(1) $P(X > 0) =$ ________；

(2) $X - Y$ 与 $X + Y$ ___（独立/不独立）；

(3) 从总体抽取样本 (X_i, Y_i)（$i = 1, \ldots, n$），则 $\sum\limits_{i=1}^{n}(X_i - Y_i)^2$ 服从 ________ 分布（写出参数）；

(4) 当 $n \to \infty$ 时，$\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - Y_i)^2$ 依概率收敛到 ________。

解题思路：

(1) 边缘分布 $X \sim N(1, 1)$，标准化后用标准正态分布计算；

(2) 二元正态变量独立的充要条件是协方差为 0；

(3) 先求 $X - Y$ 的分布，再利用卡方分布的构造；

(4) 用辛钦大数定律，收敛到总体方差。

详细解答：

(1) $X \sim N(1, 1)$，标准化 $Z = X - 1 \sim N(0, 1)$，

$P(X > 0) = P(Z > -1) = 1 - \Phi(-1) = \Phi(1) = 0.841$；

(2) $Cov(X - Y, X + Y) = Var(X) - Var(Y) = 1 - 1 = 0$，二元正态变量协方差为 0 则独立，

故独立；

(3) $X-Y \sim N(0, Var(X-Y))$，$Var(X-Y) = Var(X)+Var(Y)-2Cov(X,Y) = 1+1-2\times 0.5 = 1$，故 $X_i - Y_i \sim N(0,1)$，平方和服从 $\chi^2(n)$ 分布；

(4) 由辛钦大数定律，样本均值依概率收敛到总体期望 $E[(X-Y)^2] = Var(X-Y) + [E(X-Y)]^2 = 1+0 = 1$。

故答案依次为 (1) 0.841；(2) 独立；(3) $\chi^2(n)$；(4) 1

? 真题 4.10：20-21 秋冬，2

设 $(X,Y) \sim N(2,1,4,9,0.5)$。则：

(1) $P(2 < X < 4) =$ ________；

(2) $Var(2X-Y) =$ ________。

解题思路：

(1) 边缘分布 $X \sim N(2,4)$，标准化后计算；(2) 利用方差性质，结合协方差公式。

详细解答：

(1) $X \sim N(2,4)$，标准化 $Z = \dfrac{X-2}{2} \sim N(0,1)$，

$P(2 < X < 4) = P(0 < Z < 1) = \Phi(1) - 0.5 = 0.8413 - 0.5 = 0.3413$；

(2) $Cov(X,Y) = \rho\sigma_X\sigma_Y = 0.5 \times 2 \times 3 = 3$，

$Var(2X-Y) = 4Var(X) + Var(Y) - 4Cov(X,Y) = 4\times 4 + 9 - 4\times 3 = 13$。

故答案依次为 0.3413，13

? 真题 4.11：20-21 秋冬，3(2)

设随机变量 X 的分布函数 $F(x) = \begin{cases} 0, & x < -1, \\ 0.4, & -1 \le x < 0, \\ 0.6, & 0 \le x < 1, \\ 1, & x \ge 1. \end{cases}$ 则：

(1) $P(X=0) = \underline{0.2}$；

(2) 对 X 独立重复观测 3 次，至多有一次观测到“0”的概率为 ________。

详细解答：

(2) $Y \sim B(3, 0.2)$，$P(Y \le 1) = P(Y=0) + P(Y=1) = 0.8^3 + C_3^1 \times 0.2 \times 0.8^2 = 0.896$。

故答案为 0.896

真题 4.12：20-21 秋冬，4

设 X 与 Y 独立，X 服从参数为 1 的指数分布，Y 服从区间 $(0,2)$ 上的均匀分布。则 $Var(XY)=$ ________。

解题思路：

利用方差公式 $Var(XY)=E[(XY)^2]-[E(XY)]^2$，结合独立性 $E(XY)=E(X)E(Y)$。

详细解答：

$E(X)=1$，$E(Y)=1$，

$E(X^2)=Var(X)+[E(X)]^2=2$，$E(Y^2)=Var(Y)+[E(Y)]^2=\dfrac{4}{12}+1=\dfrac{4}{3}$，

$E(XY)=1\times 1=1$，$E[(XY)^2]=E(X^2)E(Y^2)=2\times\dfrac{4}{3}=\dfrac{8}{3}$，

$Var(XY)=\dfrac{8}{3}-1=\dfrac{5}{3}$。

故答案为 $\dfrac{5}{3}$

小结：

本章内容难度不大，分布类型也已知，主要是对公式的记忆（建议自己多推导推导），看到这发现还有公式不熟练的，可以回头再看看 SAVIA 大佬讲义中整理的公式。虽然把公式汇总在这下面是一句 prompt 的事情，但笔者是懒鬼。

5　随机变量函数的概率分布

如你所见，本章有一种题型：随机变量函数分布求解。通常不会单独考察，可能是某个填空题的一空，或者计算题的一小问，难度也相对较大。

5.1　题型九：随机变量函数分布求解

真题 5.1：24-25 秋冬, 8(4)

设连续型随机变量 X 的分布函数为

$$F_X(x)=\begin{cases} ae^x, & x<0, \\ b-ae^{-3x}, & x\geq 0. \end{cases}$$

(1) 求 a,b；

(2) 求 X 的概率密度函数 $f_X(x)$；

(3) 求 $E(X)$；

(4) 设 $Y=e^X$，求 $f_Y(y)$。

解题思路：

(1) 利用分布函数的性质：$\lim_{x\to-\infty}F_X(x)=0$，$\lim_{x\to+\infty}F_X(x)=1$，连续性 $F_X(0^-)=F_X(0^+)$；

(2) 密度函数为分布函数的导数；

(3) 利用期望定义积分计算；

(4) 用分布函数法求 Y 的密度。

详细解答：

前面部分见题型五，这里重点解答 (4)

(1) $a=\dfrac{1}{2}$，$b=1$；

(2)

$$f_X(x)=\begin{cases} \dfrac{1}{2}e^x, & x<0, \\ \dfrac{3}{2}e^{-3x}, & x\geq 0. \end{cases}$$

(3) $E(X)=-\dfrac{1}{3}$。

(4) 求 $f_Y(y)$：

- 当 $x<0$ 时，$Y=e^x\in(0,1)$，$F_Y(y)=P(Y\le y)=P(X\le \ln y)=\frac{1}{2}e^{\ln y}=\frac{y}{2}$，$f_Y(y)=\frac{1}{2}$；
- 当 $x\ge 0$ 时，$Y=e^x\in[1,+\infty)$，$F_Y(y)=P(Y\le y)=P(X\le \ln y)=1-\frac{1}{2}e^{-3\ln y}=1-\frac{1}{2y^3}$，$f_Y(y)=\frac{3}{2y^4}$；

即

$$f_Y(y)=\begin{cases}\frac{1}{2}, & 0<y<1,\\ \frac{3}{2y^4}, & y\ge 1,\\ 0, & else.\end{cases}$$

答案： (1) $a=\frac{1}{2}$，$b=1$；(2) 见上述密度函数；(3) $-\frac{1}{3}$；(4) 见上述 $f_Y(y)$。

真题 5.2：21-22 春夏，7(1)

假设 $X_1,X_2,\dots,X_5$ 为相互独立的随机变量，均服从均值为 1 的指数分布。Y 服从 $B(1,\frac{1}{3})$，记 $Z_i=X_iY$（$i=1,\dots,5$）。

(1) 求 $W=\max(X_1,X_2,\dots,X_5)$ 的分布函数和概率密度函数；

解题思路：

(1) 指数分布的分布函数为 $F_X(x)=1-e^{-x}$（$x>0$），最大值的分布函数为 $F_W(w)=[F_X(w)]^n$；

详细解答：

(1) $X_i\sim Exp(1)$，分布函数 $F_X(x)=\begin{cases}0, & x\le 0,\\ 1-e^{-x}, & x>0.\end{cases}$

W 的分布函数：$F_W(w)=P(\max X_i\le w)=[F_X(w)]^5=\begin{cases}0, & w\le 0,\\ (1-e^{-w})^5, & w>0.\end{cases}$

概率密度函数：$f_W(w)=F'_W(w)=\begin{cases}0, & w\le 0,\\ 5(1-e^{-w})^4e^{-w}, & w>0.\end{cases}$

答案： (1) 见上述 $F_W(w)$ 和 $f_W(w)$；

真题 5.3：20-21 秋冬，9(1)

设随机变量 X 的概率密度函数 $f(x)=\begin{cases}3x^2, & 0<x<1,\\ 0, & else.\end{cases}$ 令 $Y=X^2$，对 X 独立重复观测

240 次，结果记为 $X_1, X_2, \ldots, X_{240}$。

(1) 求 (X, Y) 的联合分布函数 $F(x, y)$，并判断 X 与 Y 是否独立；

解题思路：

(1) 联合分布函数 $F(x,y) = P(X \le x, Y \le y) = P(X \le x, X^2 \le y)$，需分区间讨论 x 和 y 的取值；独立性通过验证 $f(x,y) = f_X(x) f_Y(y)$ 判断（先求 Y 的边际密度）；

详细解答：

(1) 分区间讨论 $F(x, y)$：

当 $x \le 0$ 或 $y \le 0$ 时，$F(x, y) = 0$；

当 $0 < x < 1$ 且 $0 < y < x^2$ 时，$F(x,y) = P(X \le x, X \le \sqrt{y}) = P(X \le \sqrt{y}) = \int_0^{\sqrt{y}} 3t^2 dt = y$；

当 $0 < x < 1$ 且 $y \ge x^2$ 时，$F(x,y) = P(X \le x) = \int_0^x 3t^2 dt = x^3$；

当 $x \ge 1$ 且 $0 < y < 1$ 时，$F(x,y) = P(X \le \sqrt{y}) = y$；

当 $x \ge 1$ 且 $y \ge 1$ 时，$F(x, y) = 1$。

综上：

$$F(x,y) = \begin{cases} 0, & x \le 0 \text{ 或} y \le 0, \\ y, & 0 < x < 1, 0 < y < x^2 \text{ 或} x \ge 1, 0 < y < 1, \\ x^3, & 0 < x < 1, y \ge x^2, \\ 1, & x \ge 1, y \ge 1. \end{cases}$$

判断独立性：

先求 Y 的边际密度 $f_Y(y)$：当 $0 < y < 1$ 时，$f_Y(y) = \dfrac{d}{dy}F(+\infty, y) = \dfrac{d}{dy}y = 1$；其他区域 $f_Y(y) = 0$；

联合密度 $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$，当 $0 < y < x < 1$ 时，$f(x,y) = 0$（与 $f_X(x) f_Y(y) = 3x^2 \times 1 = 3x^2$ 不相等），故 X 与 Y 不独立。

答案： (1) 联合分布函数见上述表达式，X 与 Y 不独立；

注意：

这道题也许有些复杂，过程有些奇怪，但笔者精力有限，不想纠结于此题，欢迎读者向笔者提供更优美的解答。

真题 5.4：24-25 秋冬, 7

设随机变量 X_i 满足 $P(X_i=0)=P(X_i=1)=\frac{1}{2}$（$i=1,2,3$），且相互独立。令 $Y_1=\min\{X_1,X_2\}$，$Y_2=\min\{X_1,X_2,X_3\}$。

(1) 求 $Cov(X_1X_2,X_1X_2X_3)$；

(2) 求 (Y_1,Y_2) 的联合分布律；

(3) 求 $Y_1=1$ 时，Y_2 的条件分布律。

解题思路：

(1) 利用协方差公式 $Cov(U,V)=E(UV)-E(U)E(V)$，枚举所有取值计算期望；

(2) 确定 Y_1,Y_2 的可能取值（0 或 1），计算联合概率；

(3) 条件概率公式 $P(Y_2=k|Y_1=1)=\dfrac{P(Y_1=1,Y_2=k)}{P(Y_1=1)}$。

详细解答：

(1) 计算协方差：

设 $U=X_1X_2$，$V=X_1X_2X_3$，则 $UV=X_1^2X_2^2X_3=X_1X_2X_3=V$（因 $X_i^2=X_i$）。

$E(U)$：$U=1$ 当且仅当 $X_1=X_2=1$，概率 $\frac{1}{4}$，故 $E(U)=\frac{1}{4}$；

$E(V)$：$V=1$ 当且仅当 $X_1=X_2=X_3=1$，概率 $\frac{1}{8}$，故 $E(V)=\frac{1}{8}$；

$E(UV)=E(V)=\frac{1}{8}$；

协方差：$Cov(U,V)=\frac{1}{8}-\frac{1}{4}\times\frac{1}{8}=\frac{3}{32}$。

(2) 联合分布律（$Y_1,Y_2\in\{0,1\}$）：

$P(Y_1=1,Y_2=1)=P(X_1=X_2=X_3=1)=\frac{1}{8}$；

$P(Y_1=1,Y_2=0)=P(X_1=X_2=1,X_3=0)=\frac{1}{8}$；

$P(Y_1=0,Y_2=1)=0$（$Y_2=1$ 则 $X_1=X_2=X_3=1$，Y_1 必为 1）；

$P(Y_1=0,Y_2=0)=1-\frac{1}{8}-\frac{1}{8}=\frac{3}{4}$。

$Y_1\setminus Y_2$	0	1
0	$\frac{3}{4}$	0
1	$\frac{1}{8}$	$\frac{1}{8}$

(3) 条件分布律：$P(Y_1=1)=\frac{1}{4}$，$P(Y_2=1|Y_1=1)=\dfrac{\frac{1}{8}}{\frac{1}{4}}=\frac{1}{2}$，$P(Y_2=0|Y_1=1)=\dfrac{\frac{1}{8}}{\frac{1}{4}}=\frac{1}{2}$。

答案：（1）$\frac{3}{32}$；(2) 见上述联合分布律；(3) 见上述条件分布律。

Y_2	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

真题 5.5：20-21 秋冬, 7

设 (X, Y) 的联合分布律如下表所示，令 $Z = \min\{X, Y\}$：

$X \setminus Y$	0	1	2
0	a	0	b
1	b	a	b
2	0	b	$2a$

(1) 求 $E(X^2)$；

(2) 求分布函数值 $F(1,1)$；

(3) 求 (X, Z) 的联合分布律；

(4) 若已知 $P(X > Y) = 0.2$，求 a, b 的值，并判断 X 与 Y 的相关性。

解题思路：

(1) 先由概率和为 1 得 a, b 的关系，再计算 $E(X^2)$；

(2) $F(1,1) = P(X \leq 1, Y \leq 1)$；

(3) $Z = \min\{X, Y\}$，列举所有可能取值；

(4) 计算 $P(X > Y)$ 得方程，求解 a, b，再计算协方差判断相关性。

详细解答：

(1) 概率和为 1：$a + b + b + a + b + b + 2a = 4a + 4b = 1 \implies a + b = 0.25$，$E(X^2) = 0^2(a + b) + 1^2(b + a + b) + 2^2(b + 2a) = 0 + (a + 2b) + 4(2a + b) = 9a + 6b = 3(3a + 2b)$。

(2) $F(1,1) = P(X \leq 1, Y \leq 1) = a + 0 + b + a = 2a + b$。

(3) Z 的可能取值为 0,1,2，联合分布律：

(4) $P(X > Y) = P(X = 1, Y = 0) + P(X = 2, Y = 0) + P(X = 2, Y = 1) = b + 0 + b = 2b = 0.2 \implies b = 0.1$，故 $a = 0.25 - 0.1 = 0.15$，

$E(X) = 1 \times (0.1 + 0.15 + 0.1) + 2 \times (0.1 + 0.3) = 1.15$，

$E(Y) = 1 \times (0.15 + 0.1) + 2 \times (0.1 + 0.1 + 0.3) = 1.25$，

$X\setminus Z$	0	1	2
0	$a+b$	0	0
1	b	$a+b$	0
2	0	b	$2a$

$E(XY)=0+1\times 0.1+2\times 0.1+1\times 0.15+2\times 0.1+4\times 0.3=1.75$，

$Cov(X,Y)=1.75-1.15\times 1.25=0.3125\neq 0$，故正相关。

答案： (1) $3(3a+2b)$；(2) $2a+b$；(3) 见上述联合分布律；(4) $a=0.15$，$b=0.1$，正相关。

? 真题 5.6：24-25 秋冬, 9(3)

设二元随机变量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}12y^{-5}, & 1<x<y,\\ 0, & .\end{cases}$$

(1) 求 $f_X(x)$，$f_Y(y)$，并判断 X,Y 是否独立；

(2) 求 $P(1<X<1.5|Y=2)$；

(3) 设 $Z=X+Y$，求 Z 的概率密度函数 $f_Z(z)$。

解题思路：

(1) (2) 略

(3) 用卷积公式或分布函数法求 $Z=X+Y$ 的密度。

详细解答：

(1)(2) 略

(3) $Z=X+Y$ 的密度 $f_Z(z)$：

第一步：先求分布函数 $F_Z(z)=P(X+Y\leq z)$

当 $z\leq 2$ 时，由 $1<x<y$ 可知 $x+y>2$，故 $F_Z(z)=0$；

当 $z>2$ 时，积分区域满足 $1<x<y$ 且 $x+y\leq z$，即 $1<x<\frac{z}{2}$，$x<y\leq z-x$，故：

$$F_Z(z)=\int_1^{\frac{z}{2}}\int_x^{z-x}12y^{-5}dydx$$

不难算出（二重积分过程略去）

$$F_Z(z) = 1 + \frac{1}{(z-1)^3} - \frac{16}{z^3}$$

第二步：对 $F_Z(z)$ 求导得密度函数

当 $z > 2$ 时：

$$f_Z(z) = F_Z'(z) = \frac{d}{dz}\left(1 + \frac{1}{(z-1)^3} - \frac{16}{z^3}\right) = \frac{48}{z^4} - \frac{3}{(z-1)^4}$$

当 $z \le 2$ 时，$f_Z(z) = 0$，故：

$$f_Z(z) = \begin{cases} \dfrac{48}{z^4} - \dfrac{3}{(z-1)^4}, & z > 2, \\ 0, & . \end{cases}$$

答案：(1) $f_X(x) = 3x^{-4}(x > 1)$，$f_Y(y) = 12y^{-5}(y-1)(y > 1)$，不独立；(2) 0.5；

? 真题 5.7：21-22 秋冬, 5

设 X, Y 相互独立，$P(X = -1) = \dfrac{1}{2}$，$P(X = 1) = \dfrac{1}{2}$，Y 为参数为 1 的指数分布。

(1) 求 $W = XY$ 的数学期望与方差；

(2) 求 $W = XY$ 的分布函数与概率密度函数。

解题思路：

(1) 利用独立性和指数分布的期望、方差公式；

(2) 分 $w < 0$ 和 $w > 0$ 计算分布函数，再求导得密度。

详细解答：

(1) $E(W) = E(XY) = E(X)E(Y) = 0 \times 1 = 0$；

$Var(W) = E(W^2) - [E(W)]^2 = E(X^2Y^2) = E(X^2)E(Y^2)$，

$E(X^2) = 1$，$E(Y^2) = Var(Y) + [E(Y)]^2 = 1 + 1 = 2$，故 $Var(W) = 2$。

(2) 当 $w < 0$ 时，$F_W(w) = P(XY \le w) = P(X = -1, Y \ge -w) = \dfrac{1}{2}e^{w}$；

当 $w \ge 0$ 时，$F_W(w) = P(XY \le w) = P(X = -1, Y \ge -w) + P(X = 1, Y \le w) = \dfrac{1}{2} \times 1 + \dfrac{1}{2}(1 - e^{-w}) = 1 - \dfrac{1}{2}e^{-w}$；

概率密度 $f_W(w) = F_W'(w) = \begin{cases} \dfrac{1}{2}e^{w}, & w < 0, \\ \dfrac{1}{2}e^{-w}, & w \ge 0. \end{cases}$（拉普拉斯分布）。

答案：(1) $E(W) = 0$，$Var(W) = 2$；

(2) 分布函数 $F_W(w)=\begin{cases}\frac{1}{2}e^{w}, & w<0,\\ 1-\frac{1}{2}e^{-w}, & w\geq 0\end{cases}$，密度 $f_W(w)=\begin{cases}\frac{1}{2}e^{w}, & w<0,\\ \frac{1}{2}e^{-w}, & w\geq 0\end{cases}$。

真题 5.8：21-22 春夏，7(2)

假设 $X_1,X_2,\ldots,X_5$ 为相互独立的随机变量，均服从均值为 1 的指数分布。Y 服从 $B(1,\frac{1}{3})$，记 $Z_i=X_iY$（$i=1,\ldots,5$）。

(1) 求 $W=\max(X_1,X_2,\ldots,X_5)$ 的分布函数和概率密度函数；

(2) 求 $M=\max(Z_1,Z_2,\ldots,Z_5)$ 的分布函数和 $P(M=0)$。

解题思路：

(1) 指数分布的分布函数为 $F_X(x)=1-e^{-x}$（$x>0$），最大值的分布函数为 $F_W(w)=[F_X(w)]^n$；

(2) $Z_i=X_iY$，$Y=0$ 时 $Z_i=0$，$Y=1$ 时 $Z_i=X_i$，利用全概率公式求分布函数。

详细解答：

(1) $X_i\sim Exp(1)$，分布函数 $F_X(x)=\begin{cases}0, & x\leq 0,\\ 1-e^{-x}, & x>0.\end{cases}$

W 的分布函数：$F_W(w)=P(\max X_i\leq w)=[F_X(w)]^5=\begin{cases}0, & w\leq 0,\\ (1-e^{-w})^5, & w>0.\end{cases}$

概率密度函数：$f_W(w)=F_W'(w)=\begin{cases}0, & w\leq 0,\\ 5(1-e^{-w})^4e^{-w}, & w>0.\end{cases}$

(2) 先求单个 Z_i 的分布函数 $F_Z(z)$

$P(Y=0)=\frac{2}{3}$，$P(Y=1)=\frac{1}{3}$

当 $z<0$ 时：$Z_i=X_iY\geq 0$（$X_i>0$，Y 非负），故 $P(Z_i\leq z)=0$，即 $F_Z(z)=0$；

当 $z\geq 0$ 时：

$$F_Z(z)=P(Z_i\leq z)=P(Y=0)P(Z_i\leq z|Y=0)+P(Y=1)P(Z_i\leq z|Y=1)$$

条件概率 1：$Y=0$ 时，$Z_i=0$，故 $P(Z_i\leq z|Y=0)=1$；

条件概率 2：$Y=1$ 时，$Z_i=X_i$，$X_i\sim Exp(1)$ 的分布函数为 $P(X_i\leq z)=1-e^{-z}$；

代入得：

$$F_Z(z)=\frac{2}{3}\times 1+\frac{1}{3}(1-e^{-z})=1-\frac{1}{3}e^{-z}$$

综上，Z_i 的分布函数：

$$F_Z(z) = \begin{cases} 0, & z < 0, \\ 1 - \dfrac{1}{3}e^{-z}, & z \geq 0. \end{cases}$$

因 $Z_1, \dots, Z_5$ 相互独立（X_i 独立，Y 与所有 X_i 独立），最大值的分布函数满足：

$$F_M(m) = P(M \leq m) = P(Z_1 \leq m, Z_2 \leq m, \dots, Z_5 \leq m) = \prod_{i=1}^{5} F_Z(m)$$

代入 $F_Z(m)$ 的表达式：当 $m \geq 0$ 时：

$$F_M(m) = \left(1 - \frac{1}{3}e^{-m}\right)^5$$

综上，M 的分布函数：

$$F_M(m) = \begin{cases} 0, & m < 0, \\ \left(1 - \frac{1}{3}e^{-m}\right)^5, & m \geq 0. \end{cases}$$

因 $Z_i \geq 0$，故 $M = \max(Z_i) \geq 0$，$P(M = 0) = P(M \leq 0) = F_M(0)$：

$$P(M = 0) = \left(1 - \frac{1}{3}e^0\right)^5 = \left(1 - \frac{1}{3}\right)^5 = \left(\frac{2}{3}\right)^5 = \frac{32}{243}$$

答案： (1) 见上；(2) 分布函数见上，$P(M = 0) = \dfrac{32}{243}$。

小结：

简单讲，本章内容就是通过“因变量”与“自变量”函数关系转化，求解随机变量函数的概率（密度）分布。求解方法主要有两种：

(1) 利用随机变量的分布函数，通过函数关系转化求解；

(2) 利用随机变量的分布密度，通过积分求解。

详细的解题思路可参考 SAVIA 讲义第 5 讲。

6 大数定律与中心极限定理

本章可能考察一道填空题，计算题的一个小问，可分为两种题型：

(1) 求解依概率收敛

(2) 利用 CLT 求解近似分布

6.1 题型十：求解依概率收敛

真题 6.1：24-25 春夏, 3(2)

设 $X \sim N(5,9)$，$Y \sim N(0,4)$，且 $Cov(X,Y)=0$。则 $\left(\dfrac{2X-3Y-10}{2X+3Y-10}\right)^2$ 服从 ________ 分布（写出参数）；依概率收敛于 $\dfrac{1}{2n}\sum_{i=n+1}^{2n} 2X_i$ 的值为 ________。

解题思路：

(1) 正态分布的线性组合仍为正态分布，由 $Cov(X,Y)=0$ 知 $2X-3Y-10$ 与 $2X+3Y-10$ 独立，且均服从 $N(0,72)$，标准化后比值的平方服从 $F(1,1)$ 分布；

(2) 由大数定律，样本均值依概率收敛于总体期望，化简求和式即可。

详细解答：

1. 分布判断：

先计算线性组合的期望和方差：

$E(2X-3Y-10)=2E(X)-3E(Y)-10=2\times5-0-10=0$，

$Var(2X-3Y-10)=4Var(X)+9Var(Y)=4\times9+9\times4=72$，

同理 $E(2X+3Y-10)=0$，$Var(2X+3Y-10)=72$。

因 $Cov(X,Y)=0$，两线性组合独立，均服从 $N(0,72)$。标准化后：$\dfrac{2X-3Y-10}{\sqrt{72}} \sim N(0,1)$，$\dfrac{2X+3Y-10}{\sqrt{72}} \sim N(0,1)$，故其比值的平方服从 $F(1,1)$ 分布。

2. 化简求和式：$\dfrac{1}{2n}\sum_{i=n+1}^{2n} 2X_i = \dfrac{1}{n}\sum_{i=n+1}^{2n} X_i$，由大数定律，样本均值依概率收敛于总体期望 $E(X)=5$。

故答案依次为 $F(1,1)$，5

真题 6.2：24-25 秋冬, 4

设随机变量 X 满足 $P(X=2)=\dfrac{3}{4}$，$P(X=\dfrac{1}{2})=\dfrac{1}{4}$。$X_1,\cdots,X_n$ 是独立且与 X 同分布的随机变量，则当 $n\to+\infty$ 时：

(1) $\dfrac{X_1+\cdots+X_n}{n} \xrightarrow{P}$ ________；

(2) $\sqrt[n]{X_1X_2\cdots X_n} \xrightarrow{P}$ ________；

(3) $X_1, X_2, \cdots, X_{4800}$ 中有 3650 个随机变量大于 1 的概率近似为 ________。

解题思路：

(1) 利用辛钦大数定律，样本均值依概率收敛于总体期望；

(2) 取对数转化为样本均值，再用辛钦大数定律；

(3) 设“$X_i > 1$”为成功事件，服从二项分布，用正态近似计算。

详细解答：

1. 总体期望计算（辛钦大数定律）：$E(X) = 2\times\dfrac{3}{4}+\dfrac{1}{2}\times\dfrac{1}{4} = \dfrac{3}{2}+\dfrac{1}{8} = \dfrac{13}{8}$，故 $\dfrac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \dfrac{13}{8}$。

2. 乘积均值转化（取对数）：

$\ln\left(\sqrt[n]{\prod_{i=1}^{n} X_i}\right) = \dfrac{1}{n}\sum_{i=1}^{n}\ln X_i$，

$E(\ln X) = \ln 2\times\dfrac{3}{4} + \ln\dfrac{1}{2}\times\dfrac{1}{4} = \dfrac{3}{4}\ln 2 - \dfrac{1}{4}\ln 2 = \dfrac{1}{2}\ln 2 = \ln\sqrt{2}$，

由辛钦大数定律，对数均值依概率收敛于 $\ln\sqrt{2}$，故原表达式收敛于 $\sqrt{2}$。

3. 二项分布正态近似：

设 $Y = \sum_{i=1}^{4800} I(X_i > 1)$，则 $Y \sim B(4800, p)$，其中 $p = P(X>1) = P(X=2) = \dfrac{3}{4}$。

均值 $np = 4800\times\dfrac{3}{4} = 3600$，方差 $np(1-p) = 4800\times\dfrac{3}{4}\times\dfrac{1}{4} = 900$，标准差 $\sqrt{900} = 30$。

近似 $Y \sim N(3600, 30^2)$，故：$P(Y > 3650) = 1-\Phi\left(\dfrac{3650-3600}{30}\right) = 1-\Phi(\dfrac{5}{3})$。

故答案依次为 $\dfrac{13}{8}$，$\sqrt{2}$，$1-\Phi(\dfrac{5}{3})$

? 真题 6.3：23-24 春夏, 4(2)

设正态总体 $X \sim N(\mu, \sigma^2)$（$\mu\in\mathbb{R}$，$\sigma^2 > 0$ 未知），$X_1, \ldots, X_n$ 为来自该总体的简单随机样本（$n \ge 4$），$\overline{X}$ 为样本均值。则：

(2) 当 $n\to\infty$ 时，$\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 依概率收敛到 ________；

解题思路：

(2) 利用辛钦大数定律和样本方差的相合性；

详细解答：

(2) $\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \dfrac{n-1}{n}S^2$，其中 S^2 为样本方差，依概率收敛到 σ^2，故该式收敛到 σ^2。

故答案为 (2) σ^2

真题 6.4：21-22 春夏，4(4)

设 $(X,Y)\sim N(1,1,1,1,0.5)$（二元正态分布）。则：

(4) 当 $n\to\infty$ 时，$\frac{1}{n}\sum_{i=1}^{n}(X_i-Y_i)^2$ 依概率收敛到 ________。

详细解答：

(4) 由辛钦大数定律，样本均值依概率收敛到总体期望 $E[(X-Y)^2]=Var(X-Y)+[E(X-Y)]^2=1+0=1$。

故答案为 (4) 1

真题 6.5：21-22 秋冬，3(1)

设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$。则：

(1) $\frac{1}{n}\sum_{i=1}^{n}(X_i+Y_i)^2$ 依概率收敛为 ________；

解题思路：

(1) 用辛钦大数定律，收敛到 $E[(X+Y)^2]$；

详细解答：

(1) $E[(X+Y)^2]=Var(X+Y)+[E(X+Y)]^2=(\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)+(\mu_1+\mu_2)^2$，故收敛到该值；

故答案为 (1) $(\mu_1+\mu_2)^2+\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2$；

6.2 题型十一：利用 CLT 求解近似分布

当题目出现“样本数量较多”的类似说法时，可考虑用正态分布近似。

真题 6.6：24-25 春夏，10

已知一架飞机上有 200 个座位，购票的乘客有 10% 的概率不乘飞机，故售票时会多卖票。请问最多卖多少张票，会使航空公司有至少 95% 的概率能保证所有购票的乘客都有位置坐？

解题思路：

设售卖票数为 n，实际乘机人数 $X\sim B(n,0.9)$（二项分布）。因 n 较大，用正态分布近似 $N(np,np(1-p))$，利用标准化变换和置信水平 0.95 对应的分位数 $z_{0.05}=1.645$ 建立不等式求解。

详细解答：

设售卖票数为 n，实际乘机人数 $X \sim B(n, 0.9)$（成功概率为乘机概率 0.9）。

因 n 较大，用正态近似：$X \sim N(np, np(1-p)) = N(0.9n, 0.09n)$。

要求 $P(X \leq 200) \geq 0.95$，标准化得：$P\left(\frac{X-0.9n}{\sqrt{0.09n}} \leq \frac{200-0.9n}{\sqrt{0.09n}}\right) \geq 0.95$。

由标准正态分布，$P(Z \leq 1.645) = 0.95$，故：$\frac{200-0.9n}{\sqrt{0.09n}} \geq 1.645$。

整理不等式：$200 - 0.9n \geq 1.645 \times 0.3\sqrt{n}$，$0.9n + 0.4935\sqrt{n} - 200 \leq 0$。

令 $t = \sqrt{n}$（$t > 0$），得 $0.9t^2 + 0.4935t - 200 \leq 0$，解二次方程：$t \approx 14.65$。

故 $n = t^2 \approx 214.6$，因 n 为整数，最多售卖 214 张票。

故答案为 214

? 真题 6.7：23-24 春夏, 8(2)

已知总体的概率密度函数为

$$f(x) = \begin{cases} 3\theta^{-3}x^2, & 0 < x < \theta, \\ 0, & . \end{cases}$$

$X_1, \ldots, X_n$ 是从总体中抽取的 n 个样本。

(1) 求 θ 的矩估计 $\hat{\theta}_1$，并判断是否是相合估计；

(2) 当 n 足够大时，求 $\hat{\theta}_1$ 的分布；

详细解答：

略，详见【题型十五/十六（真题 8.7）】

? 真题 6.8：22-23 秋冬, 2(2)

X 服从参数为 1.5 的泊松分布。则：

(1) $Var(2X-4) =$ ________；

(2) $X_1, \ldots, X_{150}$ 是从总体简单随机抽样的样本，则 $\frac{1}{15}\sum_{i=1}^{150} X_i$ 服从 ________ 分布。

解题思路：

(1) 泊松分布 $Var(X) = \lambda$，方差性质 $Var(aX+b) = a^2 Var(X)$；

(2) 用中心极限定理，大样本下样本均值近似正态分布。

详细解答：

(1) 方差计算：$Var(X) = 1.5$，故 $Var(2X-4) = 4 \times 1.5 = 6$。

(2) CLT：样本均值 $\overline{X}=\dfrac{1}{150}\sum_{i=1}^{150}X_i$，$E(\overline{X})=1.5$，$Var(\overline{X})=\dfrac{1.5}{150}=0.01$，

故 $\dfrac{1}{15}\sum_{i=1}^{150}X_i=10\overline{X}\sim N(15,1)$（近似）。

故答案依次为 6，$N(15,1)$（近似）

? 真题 6.9：20-21 秋冬，9(2)

设随机变量 X 的概率密度函数 $f(x)=\begin{cases}3x^2, & 0<x<1,\\ 0, & else.\end{cases}$ 令 $Y=X^2$，对 X 独立重复观测 240 次，结果记为 $X_1,X_2,\ldots,X_{240}$。

(2) 求 $P\left(\sum_{i=1}^{240}X_i>177\right)$ 的近似值。

解题思路：

(2) 利用中心极限定理，大样本下样本和近似服从正态分布，先计算 $E(X)$ 和 $Var(X)$，再标准化计算概率。

详细解答：

(2) $E(X)=\int_0^1 x\cdot 3x^2dx=3\int_0^1 x^3dx=\dfrac{3}{4}$，

$E(X^2)=\int_0^1 x^2\cdot 3x^2dx=3\int_0^1 x^4dx=\dfrac{3}{5}$，

$Var(X)=E(X^2)-[E(X)]^2=\dfrac{3}{5}-\left(\dfrac{3}{4}\right)^2=\dfrac{3}{5}-\dfrac{9}{16}=\dfrac{3}{80}$；

中心极限定理近似：令 $S_n=\sum_{i=1}^{240}X_i$，则 $E(S_n)=240\times\dfrac{3}{4}=180$，$Var(S_n)=240\times\dfrac{3}{80}=9$，故 $S_n\sim N(180,3^2)$，标准化得 $Z=\dfrac{S_n-180}{3}\sim N(0,1)$；

$$P(S_n>177)=P\left(Z>\frac{177-180}{3}\right)=P(Z>-1)=1-\Phi(-1)=\Phi(1)=0.8413$$

答案： (2) $P\left(\sum_{i=1}^{240}X_i>177\right)\approx 0.8413$。

小结：

本章还有一个知识点是“切比雪夫不等式”，但近几年的卷子也没考察到。

本章内容难度不大，依概率收敛内容理解起来也相对容易，其本质上是求极限、求均值方差；而中心极限定理则是利用正态分布近似样本和的分布，本质上是正态分布内容的延伸。

7 统计量与抽样分布

7.1 题型十二：求解统计量的性质

12.1 样本均值与样本方差

真题 7.1：20-21 秋冬, 5(2)

设总体 $X \sim N(\mu, \sigma^2)$，$X_1, \ldots, X_n$ 为样本，$\overline{X}, S^2$ 分别为样本均值和样本方差。则：

(2) 若 $\mu = 0$，则 $Var(S^2 - \overline{X}^2) =$ ________。

解题思路：

(2) 利用卡方分布和正态分布的方差公式。

详细解答：

(2) $\mu = 0$ 时，$n\overline{X}^2 \sim \chi^2(1)$，$(n-1)S^2 \sim \chi^2(n-1)$，$Var(\overline{X}^2) = \dfrac{2\sigma^4}{n^2}$，$Var(S^2) = \dfrac{2\sigma^4}{n-1}$，$Var(S^2 - \overline{X}^2) = \dfrac{2\sigma^4}{n-1} + \dfrac{2\sigma^4}{n^2}$。

故答案为 (2) $\dfrac{2\sigma^4}{n-1} + \dfrac{2\sigma^4}{n^2}$

12.2 计算协方差与相关系数

真题 7.2：24-25 春夏, 7

设随机变量序列 $X_1, X_2, \ldots, X_{n+1}$ 相互独立，X_i 的分布律为 $P(X_i = 0) = \dfrac{1}{i}$，$P(X_i = 1) = 1 - \dfrac{1}{i}$（$i = 1, 2, \ldots, n+1$），令 $S = X_1X_2 + X_2X_3 + \cdots + X_nX_{n+1}$。则 $Cov(X_2, S) =$ ________，$Cov(X_{n+1}, S) =$ ________。

解题思路：

利用协方差的线性性质拆分 $Cov(X_k, S)$，仅当 X_k 与乘积项有公共因子时协方差非零，其余项因独立性为 0。结合 $E(X_i) = 1 - \dfrac{1}{i}$，$Var(X_i) = \dfrac{1}{i} - \dfrac{1}{i^2}$ 计算。

详细解答：

1. 由协方差线性性质，S 中仅 X_1X_2 和 X_2X_3 与 X_2 相关，其余项独立（协方差为 0）：$Cov(X_2, S) = Cov(X_2, X_1X_2) + Cov(X_2, X_2X_3)$。

其中:$Cov(X_2, X_1X_2) = E(X_1)Var(X_2) = (1-\dfrac{1}{1})\times(\dfrac{1}{2}-\dfrac{1}{4}) = 0$,$Cov(X_2, X_2X_3) = E(X_3)Var(X_2) = (1 - \dfrac{1}{3}) \times \dfrac{1}{4} = \dfrac{2}{3} \times \dfrac{1}{4} = \dfrac{1}{6}$，故 $Cov(X_2, S) = \dfrac{1}{6}$。

2. S 中仅 X_nX_{n+1} 与 X_{n+1} 相关：$Cov(X_{n+1},S)=Cov(X_{n+1},X_nX_{n+1})=E(X_n)Var(X_{n+1})$。

代入期望和方差：$E(X_n)=\dfrac{n-1}{n}$，$Var(X_{n+1})=\dfrac{1}{n+1}-\dfrac{1}{(n+1)^2}=\dfrac{n}{(n+1)^2}$，

故 $Cov(X_{n+1},S)=\dfrac{n-1}{n}\times\dfrac{n}{(n+1)^2}=\dfrac{n-1}{(n+1)^2}$。

故答案依次为 $\dfrac{1}{6}$，$\dfrac{n-1}{(n+1)^2}$

? 真题 7.3：24-25 秋冬，6(3)(4)

设正态总体 $X\sim N(\mu,\sigma^2)$，$\overline{X}=\dfrac{1}{4}(X_1+X_2+X_3+X_4)$，$S^2=\dfrac{1}{3}\sum_{i=1}^4(X_i-\overline{X})^2$。则：

(3) 若 $\dfrac{\overline{X}-X_1-X_2+a}{bS}\sim t(m)$，则 $(a,b,m)=$ ______；

(4) 设 $\overline{X^*}=\dfrac{1}{6}(X_1+X_2+\cdots+X_6)$，则 $\overline{X}$ 与 $\overline{X^*}$ 的相关系数为 ______。

解题思路：

(3) 构造 t 分布需分子正态（期望为 0）、分母为根号下卡方除以自由度；

(4) 利用协方差公式计算相关系数。

详细解答：

(3) 求 (a,b,m)：

分子期望 $E(\overline{X}-X_1-X_2+a)=0\implies\mu-\mu-\mu+a=0\implies a=\mu$（结合正态总体期望）；

计算分子方差：(错误计算)$Var(\overline{X}-X_1-X_2)=\dfrac{\sigma^2}{4}+\sigma^2+\sigma^2=\dfrac{9\sigma^2}{4}$，标准差 $\dfrac{3\sigma}{2}$；

注意到$\overline{X}$ 与 X_1,X_2 相关，重新计算方差：

$$Var(\overline{X}-X_1-X_2)=Var(-\frac{3}{4}X_1-\frac{3}{4}X_2+\frac{1}{4}X_3+\frac{1}{4}X_4)=\left(\frac{9}{16}+\frac{9}{16}+\frac{1}{16}+\frac{1}{16}\right)\sigma^2=\frac{5\sigma^2}{4},$$

标准差为 $\dfrac{\sqrt{5}}{2}\sigma$；分母 bS 需为 σ 量级，故 $b=\dfrac{\sqrt{5}}{2}$；自由度 $m=3$（S^2 对应的自由度），即 $(a,b,m)=(\mu,\dfrac{\sqrt{5}}{2},3)$。

(4) 相关系数：

$$Cov(\overline{X},\overline{X^*})=Cov\left(\frac{1}{4}\sum_{i=1}^4X_i,\frac{1}{6}\sum_{j=1}^6X_j\right)=\frac{1}{24}\sum_{i=1}^4Var(X_i)=\frac{4\sigma^2}{24}=\frac{\sigma^2}{6},$$

$Var(\overline{X})=\dfrac{\sigma^2}{4}$，$Var(\overline{X^*})=\dfrac{\sigma^2}{6}$，

相关系数 $\rho=\dfrac{\frac{\sigma^2}{6}}{\sqrt{\frac{\sigma^2}{4}}\cdot\sqrt{\frac{\sigma^2}{6}}}=\dfrac{\sqrt{6}}{3}$。

故答案依次为 (3) $(\mu,\dfrac{\sqrt{5}}{2},3)$；(4) $\dfrac{\sqrt{6}}{3}$。

12.3 计算事件概率

? 真题 7.4：24-25 春夏, 4(1)

设 $X \sim t(n)$，$Y \sim F(1, n)$。已知 $P(X > c) = 0.1$，则 $P(Y > c^2) =$ ________。

解题思路：

利用 t 分布与 F 分布的关系：若 $X \sim t(n)$，则 $X^2 \sim F(1, n)$。将 $P(Y > c^2)$ 转化为 $P(X^2 > c^2)$，再利用 t 分布的对称性计算。

详细解答：

由 t 分布与 F 分布的关系，知 $Y = X^2$，故：$P(Y > c^2) = P(X^2 > c^2) = P(X > c$ 或$X < -c)$。

因 t 分布对称，$P(X < -c) = P(X > c) = 0.1$，故：$P(Y > c^2) = 2 \times 0.1 = 0.2$。

故答案为 0.2

7.2 题型十三：判断统计量函数的分布

? 真题 7.5：24-25 春夏, 3(1)

设 $X \sim N(5, 9)$，$Y \sim N(0, 4)$，且 $Cov(X, Y) = 0$。则 $\left(\frac{2X - 3Y - 10}{2X + 3Y - 10}\right)^2$ 服从 ________ 分布（写出参数）；依概率收敛于 $\frac{1}{2n}\sum_{i=n+1}^{2n} 2X_i$ 的值为 ________。

解题思路：

(1) 正态分布的线性组合仍为正态分布，由 $Cov(X, Y) = 0$ 知 $2X - 3Y - 10$ 与 $2X + 3Y - 10$ 独立，且均服从 $N(0, 72)$，标准化后比值的平方服从 $F(1, 1)$ 分布；

(2) 由大数定律，样本均值依概率收敛于总体期望，化简求和式即可。

详细解答：

1. 分布判断：

先计算线性组合的期望和方差：

$E(2X - 3Y - 10) = 2E(X) - 3E(Y) - 10 = 2 \times 5 - 0 - 10 = 0$，

$Var(2X - 3Y - 10) = 4Var(X) + 9Var(Y) = 4 \times 9 + 9 \times 4 = 72$，

同理 $E(2X + 3Y - 10) = 0$，$Var(2X + 3Y - 10) = 72$。

因 $Cov(X, Y) = 0$，两线性组合独立，均服从 $N(0, 72)$。标准化后：

$\dfrac{2X-3Y-10}{\sqrt{72}} \sim N(0,1)$，$\dfrac{2X+3Y-10}{\sqrt{72}} \sim N(0,1)$，

故其比值的平方服从 $F(1,1)$ 分布。

2. 依概率收敛值：

化简求和式：$\dfrac{1}{2n}\sum_{i=n+1}^{2n} 2X_i = \dfrac{1}{n}\sum_{i=n+1}^{2n} X_i$，

由大数定律，样本均值依概率收敛于总体期望 $E(X)=5$。

故答案依次为 $F(1,1)$，5

? 真题 7.6：23-24 春夏，4(1)

设正态总体 $X \sim N(\mu,\sigma^2)$（$\mu \in \mathbb{R}$，$\sigma^2>0$ 未知），$X_1,\dots,X_n$ 为来自该总体的简单随机样本（$n \geq 4$），$\overline{X}$ 为样本均值。则：

(1) $\left(\dfrac{X_1+X_2-X_3-\mu}{X_2+X_3-X_4-\mu}\right)^2$ 服从 ________ 分布（写出参数）；若 $\dfrac{(\overline{X}-X_1)^2}{a} \sim \chi^2(k)$，则 $(a,k)=$ ________；

解题思路：

(1) 正态变量线性组合仍为正态，独立正态变量比值的平方服从 F 分布；$\overline{X}-X_1$ 为正态变量，平方后除以方差服从卡方分布；

详细解答：

(1) 令 $U=X_1+X_2-X_3-\mu \sim N(0,3\sigma^2)$，$V=X_2+X_3-X_4-\mu \sim N(0,3\sigma^2)$，且 U,V 独立，故 $\dfrac{U}{\sqrt{3}\sigma} \sim N(0,1)$，$\dfrac{V}{\sqrt{3}\sigma} \sim N(0,1)$，比值的平方服从 $F(1,1)$ 分布；

$\overline{X}-X_1=\dfrac{-n+1}{n}X_1+\dfrac{1}{n}\sum_{i=2}^{n} X_i$，方差 $Var(\overline{X}-X_1)=\dfrac{(n-1)^2\sigma^2}{n^2}+\dfrac{n-1}{n^2}\sigma^2=\dfrac{(n-1)\sigma^2}{n}$，

故 $a=\dfrac{(n-1)\sigma^2}{n}$，$k=1$，即 $(a,k)=\left(\dfrac{n-1}{n}\sigma^2,1\right)$。

故答案为 (1) $F(1,1)$，$\left(\dfrac{n-1}{n}\sigma^2,1\right)$；

? 真题 7.7：21-22 春夏，4(3)

设 $(X,Y) \sim N(1,1,1,1,0.5)$（二元正态分布）。则：

(3) 从总体抽取样本 (X_i,Y_i)（$i=1,\dots,n$），则 $\sum\limits_{i=1}^{n}(X_i-Y_i)^2$ 服从 ________ 分布（写出参数）；

详细解答：

(3) $X-Y \sim N(0,Var(X-Y))$，$Var(X-Y)=Var(X)+Var(Y)-2Cov(X,Y)=1+1-2\times 0.5=1$，故 $X_i-Y_i \sim N(0,1)$，平方和服从 $\chi^2(n)$ 分布；

答案为 (3) $\chi^2(n)$

小结：

本章重点在于理解和区分随机变量和统计量、样本和总体。三种抽样分布特征也相对明显，读者不妨自己总结一下其中奥妙。这章内容非常重要，是学习统计学内容的基础。

破题关键在于抽样样本所属“正态总体”是否为标准正态分布。

8 点估计

通常本章会有一道计算题，可分成两种（或三种）题型：

(1) 根据估计量的评价标准求参（填空题偏多）

(2) 求解点估计（矩估计和极大似然估计）并评价

8.1 题型十四：估计量评价求参

共有四种评价准则：无偏性，有效性，均方误差，相合性。

? 真题 8.1：24-25 春夏, 5

设 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, \ldots, X_n$ 为样本，$\overline{X}$ 为样本均值，S^2 为样本方差，令 $T = \overline{X}^2 - \dfrac{S^2}{n}$。则 T ___（是/不是）μ^2 的无偏估计；若 $\mu = 0$，$\sigma^2 = 1$，$n = 5$，则 $Var(T) =$ ________。

解题思路：

(1) 无偏估计的判定：验证 $E(T) = \mu^2$，利用 $E(\overline{X}^2) = Var(\overline{X}) + [E(\overline{X})]^2$ 和 $E(S^2) = \sigma^2$；

(2) 计算方差时，利用 $\overline{X} \sim N(0, 1/5)$ 推导 $5\overline{X}^2 \sim \chi^2(1)$，及 $(n-1)S^2 \sim \chi^2(n-1)$，结合卡方分布的方差公式。

详细解答：

1. 无偏性判定：$E(\overline{X}^2) = Var(\overline{X}) + [E(\overline{X})]^2 = \dfrac{\sigma^2}{n} + \mu^2$，又 $E(S^2) = \sigma^2$，故：

$E(T) = E(\overline{X}^2) - \dfrac{1}{n}E(S^2) = \dfrac{\sigma^2}{n} + \mu^2 - \dfrac{\sigma^2}{n} = \mu^2$，因此 T 是 μ^2 的无偏估计。

2. 方差计算（$\mu = 0$，$\sigma^2 = 1$，$n = 5$）：

对 $\overline{X}$：$\overline{X} \sim N(0, 1/5)$，故 $5\overline{X}^2 \sim \chi^2(1)$，

对 S^2：$4S^2 \sim \chi^2(4)$，故 $Var(S^2) = Var\left(\dfrac{1}{4}\chi^2(4)\right) = \dfrac{1}{16} \times 8 = \dfrac{1}{2}$；

因 $\overline{X}$ 与 S^2 独立，故：$Var(T) = Var(\overline{X}^2) + Var\left(\dfrac{S^2}{5}\right) = \dfrac{2}{25} + \dfrac{1}{25} \times \dfrac{1}{2} = \dfrac{1}{10}$。

故答案依次为是，$\dfrac{1}{10}$

? 真题 8.2：21-22 秋冬, 3(2)

设 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$。则：

(1) $\dfrac{1}{n}\sum_{i=1}^{n}(X_i + Y_i)^2$ 依概率收敛为 ________；

(2) 若 $\mu_1 = \mu_2 = 2\theta$，$\sigma_1 = \sigma_2 = \theta$，$\rho = \dfrac{1}{2}$，则 $a\sum_{i=1}^{n}(X_i + Y_i)^2$ 为 θ^2 无偏估计的充要条件为 $a =$ ________。

解题思路：

(1) 用辛钦大数定律，收敛到 $E[(X+Y)^2]$；

(2) 无偏估计要求 $E[a\sum(X_i+Y_i)^2]=\theta^2$，计算期望并求解 a。

详细解答：

(1) $E[(X+Y)^2]=Var(X+Y)+[E(X+Y)]^2=(\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)+(\mu_1+\mu_2)^2$；

(2) $E(X+Y)=2\theta+2\theta=4\theta$，$Var(X+Y)=\theta^2+\theta^2+2\times\frac{1}{2}\times\theta\times\theta=3\theta^2$，

$E[(X+Y)^2]=3\theta^2+(4\theta)^2=19\theta^2$，无偏性要求 $a\times n\times 19\theta^2=\theta^2$，故 $a=\frac{1}{19n}$。

故答案依次为 (1) $(\mu_1+\mu_2)^2+\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2$；(2) $\frac{1}{19n}$

真题 8.3：20-21 秋冬, 5

设总体 $X\sim N(\mu,\sigma^2)$，$X_1,\dots,X_n$ 为样本，$\overline{X},S^2$ 分别为样本均值和样本方差。则：

(1) 当 $k=$ ________ 时，$\overline{X}^2-kS^2$ 是 μ^2 的无偏估计量；

(2) 若 $\mu=0$，则 $Var(S^2-\overline{X}^2)=$ ________。

解题思路：

(1) 无偏性要求 $E(\overline{X}^2-kS^2)=\mu^2$，计算期望求解 k；

(2) 利用卡方分布和正态分布的方差公式。

详细解答：

(1) $E(\overline{X}^2)=\frac{\sigma^2}{n}+\mu^2$，$E(S^2)=\sigma^2$，故 $\frac{\sigma^2}{n}+\mu^2-k\sigma^2=\mu^2\implies k=\frac{1}{n}$；

(2) $\mu=0$ 时，$n\overline{X}^2\sim\chi^2(1)$，$(n-1)S^2\sim\chi^2(n-1)$，

$Var(\overline{X}^2)=\frac{2\sigma^4}{n^2}$，$Var(S^2)=\frac{2\sigma^4}{n-1}$，$Var(S^2-\overline{X}^2)=\frac{2\sigma^4}{n-1}+\frac{2\sigma^4}{n^2}$。

故答案依次为 (1) $\frac{1}{n}$；(2) $\frac{2\sigma^4}{n-1}+\frac{2\sigma^4}{n^2}$

8.2 题型十五/六：求解矩估计/极大似然估计（并评价）

真题 8.4：24-25 秋冬, 10

设随机变量 X 的概率密度函数为

$$f(x;\mu)=\begin{cases}2e^{-2(x-\mu)}, & x\ge\mu,\\ 0, & \text{其它}.\end{cases}$$

$X_1, \cdots, X_n$ 为简单随机样本。

(1) 求参数 μ 的矩估计 $\hat{\mu}_1$，并判断其是否为无偏估计；

(2) 求参数 μ 的极大似然估计 $\hat{\mu}_2$，并判断其是否为相合估计；

(3) 用均方误差准则说明 $\hat{\mu}_1$ 和 $\hat{\mu}_2$ 哪个更有效。

解题思路：

(1) 矩估计用一阶原点矩，计算 $E(X)$ 并令其等于样本均值；

(2) 构造似然函数，利用单调性求 MLE，相合性通过切比雪夫不等式验证；

(3) 均方误差 $MSE(\hat{\mu}) = Var(\hat{\mu}) + [E(\hat{\mu}) - \mu]^2$，比较两者大小。

详细解答：

(1)$E(X) = \int_{\mu}^{+\infty} x \cdot 2e^{-2(x-\mu)} dx = \mu + \dfrac{1}{2}$（令 $t = x - \mu$ 积分）；

矩估计：$\overline{X} = \hat{\mu}_1 + \dfrac{1}{2} \implies \hat{\mu}_1 = \overline{X} - \dfrac{1}{2}$；

无偏性：$E(\hat{\mu}_1) = E(\overline{X}) - \dfrac{1}{2} = \mu + \dfrac{1}{2} - \dfrac{1}{2} = \mu$，故为无偏估计。

(2) 似然函数：$L(\mu) = \prod_{i=1}^{n} 2e^{-2(X_i-\mu)} = 2^n e^{-2\sum_{i=1}^{n}(X_i-\mu)}$（$X_i \geq \mu$）；

单调性：$L(\mu)$ 随 μ 增大而增大，故 MLE 为 $\hat{\mu}_2 = \min\{X_1, \cdots, X_n\}$；

相合性：$P(|\hat{\mu}_2 - \mu| > \varepsilon) = P(\min X_i > \mu + \varepsilon) = \prod_{i=1}^{n} P(X_i > \mu + \varepsilon) = e^{-2n\varepsilon} \to 0$（$n \to \infty$），故为相合估计。

(3) 均方误差比较：

$MSE(\hat{\mu}_1) = Var(\hat{\mu}_1) = \dfrac{Var(X)}{n} = \dfrac{1}{4n}$（$Var(X) = \dfrac{1}{4}$）；

$E(\hat{\mu}_2) = \mu + \dfrac{1}{2n}$，$Var(\hat{\mu}_2) = \dfrac{1}{4n^2}$，故 $MSE(\hat{\mu}_2) = \dfrac{1}{4n^2} + \left(\dfrac{1}{2n}\right)^2 = \dfrac{1}{2n^2}$；

当 $n \geq 1$ 时，$\dfrac{1}{2n^2} < \dfrac{1}{4n}$，故 $\hat{\mu}_2$ 更有效。

答案： (1) $\hat{\mu}_1 = \overline{X} - \dfrac{1}{2}$，无偏；(2) $\hat{\mu}_2 = \min\{X_1, \cdots, X_n\}$，相合；(3) $\hat{\mu}_2$ 更有效。

真题 8.5：24-25 春夏, 11

设总体 X 服从区间 $[0, \theta]$ 上的均匀分布，$X_1, X_2, \ldots, X_n$ 为样本，令 $T_c = cX_{(n)}$（$X_{(n)} = \max\{X_1, \ldots, X_n\}$，$c$ 为常数）。

(1) 求 $Var(X)$ 的最大似然估计；

(2) 当 T_c 为 θ 的无偏估计时，求 c 的大小；

(3) 以 T_c 为估计量，求当其均方误差取最小值时 c 的大小。

解题思路：

(1) 先求 θ 的最大似然估计（均匀分布的 MLE 为 $X_{(n)}$），再利用方差公式 $Var(X)=\dfrac{\theta^2}{12}$ 和 MLE 的不变性；

(2) 无偏估计要求 $E(T_c)=\theta$，需先求 $X_{(n)}$ 的期望；

(3) 均方误差 $MSE(T_c)=E[(T_c-\theta)^2]$，转化为关于 c 的二次函数，求最小值点。

详细解答：

(1) 均匀分布的似然函数 $L(\theta)=\dfrac{1}{\theta^n}$（$0\le X_i\le\theta$），$\theta$ 越小似然函数越大，

故 θ 的 MLE 为 $\hat{\theta}=X_{(n)}$。

由 $Var(X)=\dfrac{\theta^2}{12}$，根据 MLE 的不变性，$Var(X)$ 的 MLE 为：$Va\hat{r}(X)=\dfrac{\hat{\theta}^2}{12}=\dfrac{X_{(n)}^2}{12}$。

(2) 无偏估计的 c 值：

先求 $X_{(n)}$ 的密度函数：$F_{X_{(n)}}(t)=[P(X\le t)]^n=\left(\dfrac{t}{\theta}\right)^n$（$0\le t\le\theta$），$f_{X_{(n)}}(t)=\dfrac{nt^{n-1}}{\theta^n}$。

期望：$E(X_{(n)})=\int_0^\theta t\cdot\dfrac{nt^{n-1}}{\theta^n}dt=\dfrac{n}{\theta^n}\cdot\dfrac{\theta^{n+1}}{n+1}=\dfrac{n\theta}{n+1}$。

无偏性要求 $E(T_c)=cE(X_{(n)})=\theta$，故：$c=\dfrac{n+1}{n}$。

(3) 均方误差最小的 c 值：先求 $E(X_{(n)}^2)$：$E(X_{(n)}^2)=\int_0^\theta t^2\cdot\dfrac{nt^{n-1}}{\theta^n}dt=\dfrac{n\theta^2}{n+2}$。

均方误差：$MSE(T_c)=E[(cX_{(n)}-\theta)^2]=c^2E(X_{(n)}^2)-2c\theta E(X_{(n)})+\theta^2$。

代入期望表达式，整理为关于 c 的二次函数：$MSE(T_c)=\theta^2\left(\dfrac{n}{n+2}c^2-2\cdot\dfrac{n}{n+1}c+1\right)$。

二次函数 ac^2+bc+d 最小值在 $c=-\dfrac{b}{2a}$，故：$c=\dfrac{2\cdot\dfrac{n}{n+1}}{2\cdot\dfrac{n}{n+2}}=\dfrac{n+2}{n+1}$。

答案： (1) $\dfrac{X_{(n)}^2}{12}$；(2) $\dfrac{n+1}{n}$；(3) $\dfrac{n+2}{n+1}$

真题 8.6：23-24 春夏, 8

已知总体的概率密度函数为

$$f(x)=\begin{cases}3\theta^{-3}x^2, & 0<x<\theta,\\ 0, & .\end{cases}$$

$X_1,\dots,X_n$ 是从总体中抽取的 n 个样本。

(1) 求 θ 的矩估计 $\hat{\theta}_1$，并判断是否是相合估计；

(2) 当 n 足够大时，求 $\hat{\theta}_1$ 的分布；

(3) 求 θ 的极大似然估计 $\hat{\theta}_2$，并判断是否是无偏估计。

解题思路：

(1) 矩估计用一阶原点矩，相合性通过切比雪夫不等式验证；

(2) 用中心极限定理近似；

(3) 构造似然函数，求最大值点得 MLE，再验证无偏性。

详细解答：

(1) 一阶原点矩：$E(X) = \int_0^\theta x \cdot 3\theta^{-3}x^2dx = 3\theta^{-3} \cdot \frac{\theta^4}{4} = \frac{3\theta}{4}$；

矩估计：$\overline{X} = \frac{3\hat{\theta}_1}{4} \implies \hat{\theta}_1 = \frac{4\overline{X}}{3}$；

相合性：$E(\hat{\theta}_1) = \theta$，$Var(\hat{\theta}_1) = \frac{16Var(X)}{9n}$，其中 $Var(X) = \frac{3\theta^2}{80}$，故 $Var(\hat{\theta}_1) \to 0$（$n \to \infty$），为相合估计。

(2) 由中心极限定理，$\sqrt{n}(\overline{X} - \frac{3\theta}{4}) \xrightarrow{L} N(0, \frac{3\theta^2}{80})$，

故 $\sqrt{n}(\hat{\theta}_1 - \theta) = \frac{4}{3}\sqrt{n}(\overline{X} - \frac{3\theta}{4}) \xrightarrow{L} N(0, \frac{16}{9} \times \frac{3\theta^2}{80}) = N(0, \frac{\theta^2}{15})$，即 n 足够大时，$\hat{\theta}_1 \sim N(\theta, \frac{\theta^2}{15n})$。

(3) 似然函数：$L(\theta) = \prod_{i=1}^n 3\theta^{-3}x_i^2 = 3^n\theta^{-3n}\prod_{i=1}^n x_i^2$（$0 < x_i < \theta$）；

MLE：$L(\theta)$ 随 θ 增大而减小，故 $\hat{\theta}_2 = \max\{X_1, \ldots, X_n\}$；

无偏性：$F_{\hat{\theta}_2}(t) = \left(\frac{t}{\theta}\right)^n$（$0 << t < \theta$），密度 $f_{\hat{\theta}_2}(t) = \frac{nt^{n-1}}{\theta^n}$，$E(\hat{\theta}_2) = \int_0^\theta t \cdot \frac{nt^{n-1}}{\theta^n}dt = \frac{n\theta}{n+1} \neq \theta$，故不是无偏估计。

答案：(1) $\hat{\theta}_1 = \frac{4\overline{X}}{3}$，是相合估计；(2) 近似服从 $N(\theta, \frac{\theta^2}{15n})$；(3) $\hat{\theta}_2 = \max\{X_1, \ldots, X_n\}$，不是无偏估计。

真题 8.7：22-23 秋冬，8

设总体的概率密度函数为

$$f(x) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x < \theta, \\ 0, & else. \end{cases}$$

$X_1, \ldots, X_n$ 是简单随机抽取的样本。

(1) 求矩估计量 $\hat{\theta}_1$ 并求 $E(\hat{\theta}_1)$；

(2) 判断 $\hat{\theta}_1$ 是否为无偏估计量、是否为相合估计量；

(3) 求 θ 的极大似然估计量 $\hat{\theta}_2$。

解题思路：

(1) 矩估计用一阶原点矩；

(2) 无偏性看 $E(\hat{\theta}_1)=\theta$，相合性看方差趋于 0；

(3) 构造似然函数，求最大值点。

详细解答：

(1) 一阶原点矩：$E(X)=\int_0^\theta x\cdot\frac{2x}{\theta^2}dx=\frac{2\theta}{3}$；

矩估计：$\overline{X}=\frac{2\hat{\theta}_1}{3}\implies\hat{\theta}_1=\frac{3\overline{X}}{2}$；

期望：$E(\hat{\theta}_1)=\frac{3}{2}\times\frac{2\theta}{3}=\theta$。

(2) 无偏性：$E(\hat{\theta}_1)=\theta$，故为无偏估计；

相合性：$Var(\hat{\theta}_1)=\frac{9Var(X)}{4n}$，$Var(X)=\frac{\theta^2}{18}$，故 $Var(\hat{\theta}_1)=\frac{\theta^2}{8n}\to 0$（$n\to\infty$），为相合估计。

(3) 极大似然估计：

似然函数：$L(\theta)=\prod_{i=1}^n\frac{2x_i}{\theta^2}=2^n\theta^{-2n}\prod_{i=1}^n x_i$（$0<x_i<\theta$）；

MLE：$L(\theta)$ 随 θ 增大而减小，故 $\hat{\theta}_2=\max\{X_1,\ldots,X_n\}$。

答案： (1) $\hat{\theta}_1=\frac{3\overline{X}}{2}$，$E(\hat{\theta}_1)=\theta$；(2) 无偏，相合；(3) $\hat{\theta}_2=\max\{X_1,\ldots,X_n\}$。

真题 8.8：21-22 春夏, 8

假设总体 $X\sim N(\mu,\sigma^2)$，μ,σ^2 均为未知参数，$X_1,\ldots,X_n$ 为样本。

(1) 分别求 μ,σ^2 的极大似然估计；

(2) 求参数函数 $g(\mu,\sigma^2)=3\mu+4\sigma^2$ 的极大似然估计 T；

(3) 求 $E(T)$，并判断 T 是不是 $g(\mu,\sigma^2)$ 的无偏估计、相合估计？说明理由；

(4) 求 $Var(T)$。

解题思路：

(1) 构造似然函数，求导找最大值点；

(2) 利用极大似然估计的不变性；

(3) 无偏性看 $E(T)=g(\mu,\sigma^2)$，相合性看方差趋于 0；

(4) 利用样本均值和样本方差的独立性计算方差。

详细解答：

(1) 似然函数 $L(\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\mu)^2}$，

对 μ 求导得 $\hat{\mu}_{MLE}=\overline{X}$；

对 σ^2 求导得 $\hat{\sigma}^2_{MLE}=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2$。

(2) 由不变性，$T = 3\hat{\mu}_{MLE} + 4\hat{\sigma}^2_{MLE} = 3\overline{X} + \dfrac{4}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$。

(3)$E(T) = 3E(\overline{X}) + 4E\left(\dfrac{1}{n}\sum(X_i - \overline{X})^2\right) = 3\mu + 4 \times \dfrac{n-1}{n}\sigma^2$，

因 $E(T) \neq 3\mu + 4\sigma^2$，故不是无偏估计；

相合性：$\overline{X} \xrightarrow{P} \mu$，$\dfrac{1}{n}\sum(X_i - \overline{X})^2 \xrightarrow{P} \sigma^2$，

由依概率收敛的性质，$T \xrightarrow{P} 3\mu + 4\sigma^2$，故是相合估计。

(4)$Var(T) = Var(3\overline{X}) + Var\left(4 \times \dfrac{n-1}{n}S^2\right)$（$S^2$ 为样本方差），

$Var(3\overline{X}) = 9 \times \dfrac{\sigma^2}{n}$，$Var\left(\dfrac{n-1}{n}S^2\right) = \left(\dfrac{n-1}{n}\right)^2 \times \dfrac{2\sigma^4}{n-1} = \dfrac{2(n-1)\sigma^4}{n^2}$，

故 $Var(T) = \dfrac{9\sigma^2}{n} + 16 \times \dfrac{2(n-1)\sigma^4}{n^2} = \dfrac{9\sigma^2}{n} + \dfrac{32(n-1)\sigma^4}{n^2}$。

答案：(1) $\hat{\mu}_{MLE} = \overline{X}$，$\hat{\sigma}^2_{MLE} = \dfrac{1}{n}\sum(X_i - \overline{X})^2$；(2) $T = 3\overline{X} + \dfrac{4}{n}\sum(X_i - \overline{X})^2$；(3) $E(T) = 3\mu + \dfrac{4(n-1)}{n}\sigma^2$，不是无偏估计，是相合估计；(4) $Var(T) = \dfrac{9\sigma^2}{n} + \dfrac{32(n-1)\sigma^4}{n^2}$。

真题 8.9：20-21 秋冬, 10

设 X 的概率密度函数为 $f(x,\theta) = \begin{cases} \dfrac{2\theta^2}{x^3}, & x \geq \theta, \\ 0, & x < \theta \end{cases}$，其中 $\theta > 0$ 是未知参数，$X_1, X_2, \dots, X_n$ 是总体 X 的简单随机样本。

(1) 求 θ 的矩估计量 $\hat{\theta}_1$，并判断其是否为 θ 的相合估计量；

(2) 求 θ 的极大似然估计量 $\hat{\theta}_2$，并判断其是否为 θ 的无偏估计量。

解题思路：

(1) 矩估计通过一阶原点矩建立方程（先求 $E(X)$）；相合估计量的判定：若 $E(\hat{\theta}_1) \to \theta$ 且 $Var(\hat{\theta}_1) \to 0$（$n \to \infty$），则为相合估计；

(2) 极大似然估计通过构造似然函数，利用其单调性求最大值点；无偏估计的判定：验证 $E(\hat{\theta}_2) = \theta$。

详细解答：

(1) 求矩估计

$$E(X) = \int_{\theta}^{+\infty} x \cdot \frac{2\theta^2}{x^3}dx = 2\theta^2 \int_{\theta}^{+\infty} x^{-2}dx = 2\theta^2 \cdot \frac{1}{\theta} = 2\theta$$

矩估计量：令 $E(X) = \overline{X}$（样本均值），解得 $\hat{\theta}_1 = \dfrac{\overline{X}}{2}$；

相合性判定：$E(\hat{\theta}_1) = \dfrac{1}{2}E(\overline{X}) = \dfrac{1}{2}E(X) = \theta$（无偏）；

通过“依概率收敛的性质”：$\overline{X} \xrightarrow{P} E(X) = 2\theta$，故 $\hat{\theta}_1 = \dfrac{\overline{X}}{2} \xrightarrow{P} \theta$，因此 $\hat{\theta}_1$ 是相合估计量。

(2) 构造似然函数：当 $X_i \geq \theta$（$i = 1, \ldots, n$）时，$L(\theta) = \prod_{i=1}^{n} \dfrac{2\theta^2}{X_i^3} = 2^n \theta^{2n} \prod_{i=1}^{n} X_i^{-3}$；

求 MLE：$L(\theta)$ 是 θ 的增函数，故 θ 的最大值点为样本最小值，即 $\hat{\theta}_2 = \min\{X_1, X_2, \ldots, X_n\}$；

无偏性判定：

先求 $\hat{\theta}_2$ 的分布函数：$F_{\hat{\theta}_2}(t) = P(\min X_i \leq t) = 1 - P(\min X_i > t) = 1 - [P(X > t)]^n$，

其中 $P(X > t) = \int_t^{+\infty} \dfrac{2\theta^2}{x^3} dx = \dfrac{\theta^2}{t^2}$（$t \geq \theta$），故 $F_{\hat{\theta}_2}(t) = 1 - \left(\dfrac{\theta^2}{t^2}\right)^n = 1 - \dfrac{\theta^{2n}}{t^{2n}}$（$t \geq \theta$）；

概率密度：$f_{\hat{\theta}_2}(t) = F'_{\hat{\theta}_2}(t) = \dfrac{2n\theta^{2n}}{t^{2n+1}}$（$t \geq \theta$）；

计算期望：

$$E(\hat{\theta}_2) = \int_\theta^{+\infty} t \cdot \frac{2n\theta^{2n}}{t^{2n+1}} dt = 2n\theta^{2n} \int_\theta^{+\infty} t^{-2n} dt = 2n\theta^{2n} \cdot \frac{\theta^{-(2n-1)}}{2n-1} = \frac{2n}{2n-1}\theta \neq \theta$$

故 $\hat{\theta}_2$ 不是无偏估计量。

答案： (1) 矩估计量 $\hat{\theta}_1 = \dfrac{\overline{X}}{2}$，是相合估计量；

(2) 极大似然估计量 $\hat{\theta}_2 = \min\{X_1, \ldots, X_n\}$，不是无偏估计量。

小结：

笔者认为，这一章内容其实不难，算法相对固定，可能理解其中原理相对抽象，主要是对点估计（矩估计和极大似然估计）的理解和应用。就近几年试卷而言，考试肯定会考，可能是计算题，也可能是填空题，或者只是某计算题的一小问。

9 区间估计与假设检验

本章为考试必考内容，通常是最后一道计算题，可分为三种题型（第四种不考）：

(1) 求置信区间

(2) 求拒绝域并检验假设

(3) 求 P 值并检验假设

这一章节中，通常计算题会同时涉及求置信区间和假设检验，笔者精力有限，部分题目没有明确地分类相应的题型中，望读者谅解。

9.1 题型十七：求置信区间

真题 9.1：24-25 春夏, 6

已知 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, \ldots, X_{10}$ 为样本，样本均值 $\overline{X} = 5$，样本平方和 $\sum_{i=1}^{10} X_i^2 = 300$。当置信度为 0.95 时，$\mu$ 的单侧下限为 ________；当 μ 的置信区间为 $\left(5 - \frac{\sqrt{5}}{2}, 5 + \frac{\sqrt{5}}{2}\right)$ 时，$t_{\alpha/2}(9) =$ ________。

解题思路：

(1) 总体方差未知，构造枢轴量 $\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$，利用单侧置信下限公式推导；

(2) 置信区间长度与枢轴量的关系，结合已知区间反解 $t_{\alpha/2}(9)$。

详细解答：

1. 先求样本方差：$S^2 = \frac{1}{n-1}\left(\sum X_i^2 - n\overline{X}^2\right) = \frac{1}{9}(300 - 10 \times 25) = \frac{50}{9}$，故 $S = \frac{\sqrt{50}}{3}$。

置信度 0.95，单侧检验的分位数 $t_{0.05}(9) = 1.833$，单侧下限公式：

$\hat{\mu}_L = \overline{X} - t_\alpha(n-1)\frac{S}{\sqrt{n}} = 5 - 1.833 \times \frac{\sqrt{5/9}}{\sqrt{10}} \approx 3.635$。

2. 求 $t_{\alpha/2}(9)$：置信区间半长为 $\frac{\sqrt{5}}{2}$，而半长公式为 $t_{\alpha/2}(9)\frac{S}{\sqrt{n}}$，故：

$t_{\alpha/2}(9) = \frac{\sqrt{5}/2}{S/\sqrt{n}} = \frac{\sqrt{5}/2}{\sqrt{5/9}/\sqrt{10}} = \frac{3}{2} = 1.5$。

故答案依次为 3.635，1.5

真题 9.2：24-25 秋冬, 6(1)

设正态总体 $X \sim N(\mu, \sigma^2)$，$\overline{X} = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$，$S^2 = \frac{1}{3}\sum_{i=1}^{4}(X_i - \overline{X})^2$。则：

(1) 设 μ 未知，置信水平为 $1-\alpha$ 的 σ^2 的单侧置信下限为 _____；

解题思路：

(1) 总体正态、μ 未知，用 χ^2 分布构造置信区间，单侧下限公式推导；

详细解答：

(1) 枢轴量 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$（$n=4$），

由 $P\left(\dfrac{3S^2}{\sigma^2} \leq \chi^2_\alpha(3)\right) = \alpha$，得单侧置信下限：$\dfrac{3S^2}{\chi^2_\alpha(3)}$。

故答案为 (1) $\dfrac{3S^2}{\chi^2_\alpha(3)}$

真题 9.3：23-24 春夏, 9(1)

有 A、B 两种小麦，发芽时间分别服从 $N(\mu_1,\sigma_1^2)$，$N(\mu_2,\sigma_2^2)$。现分别从两种小麦中取 11 和 10 个样本，样本均值分别为 $\overline{x}=99.1$，$\overline{y}=98.9$数据存疑，样本方差分别为 $s_1^2=0.94$，$s_2^2=0.88$。

(1) 求 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 在置信水平为 0.95 下的双侧等尾置信区间，并判断 σ_1^2 是否等于 σ_2^2；

(2) 在 (1) 的基础上，检验 $H_0:\mu_1=\mu_2$；$H_1:\mu_1\neq\mu_2$（显著水平 $\alpha=0.05$）。

解题思路：

(1) 方差比置信区间用 F 分布，包含 1 则认为方差相等；

(2) 方差相等时用两样本 t 检验，构造统计量并对比临界值。

详细解答：

(1) 自由度 $df1=10$，$df2=9$，$\alpha=0.05$，

查表得 $F_{0.025}(10,9)=3.96$，$F_{0.975}(10,9)=\dfrac{1}{F_{0.025}(9,10)}\approx 0.255$；

置信区间：$\left(\dfrac{s_1^2}{s_2^2F_{\alpha/2}(df1,df2)},\dfrac{s_1^2}{s_2^2F_{1-\alpha/2}(df1,df2)}\right)=\left(\dfrac{0.94}{0.88\times3.96},\dfrac{0.94}{0.88\times0.255}\right)\approx(0.269,4.147)$；

因置信区间包含 1，故认为 $\sigma_1^2=\sigma_2^2$。

(2) 两样本 t 检验：

合并方差：$s_w^2=\dfrac{(11-1)s_1^2+(10-1)s_2^2}{11+10-2}=\dfrac{10\times0.94+9\times0.88}{19}\approx0.9126$，$s_w\approx0.955$；

检验统计量：$t=\dfrac{\overline{x}-\overline{y}}{s_w\sqrt{\dfrac{1}{11}+\dfrac{1}{10}}}=\dfrac{99.1-98.9}{0.955\times0.426}\approx0.498$；

临界值：$t_{0.025}(19)=2.093$，因 $|t|<2.093$，未落入拒绝域，接受 H_0。

答案： (1) 置信区间为 (0.269, 4.147)，$\sigma_1^2=\sigma_2^2$；(2) 接受原假设 H_0。

真题 9.4：22-23 秋冬, 9

设 $X\sim N(\mu,\sigma^2)$，16 个简单随机样本的样本均值 $\overline{X}=100$，样本方差 $S^2=36$。

(1) 求 σ^2 的 95% 的置信下限；

(2) 在 $\alpha = 0.05$ 下检验 $H_0 : \sigma^2 = 84$，$H_1 : \sigma^2 \neq 84$。

解题思路：

(1) 总体正态、μ 未知，用 χ^2 分布求单侧置信下限；

(2) 方差检验用 χ^2 统计量，对比临界值。

详细解答：

(1) 枢轴量 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(15)$，$1-\alpha = 0.95$，$\chi^2_{0.95}(15) = 7.261$；

置信下限：$\hat{\sigma}^2_L = \dfrac{(16-1)\times 36}{7.261} \approx 74.37$。

(2) 检验统计量：$\chi^2 = \dfrac{(16-1)\times 36}{84} \approx 6.429$；

临界值：$\chi^2_{0.025}(15) = 27.488$，$\chi^2_{0.975}(15) = 6.262$；

因 $6.262 < 6.429 < 27.488$，未落入拒绝域，接受 H_0。

答案： (1) 74.37（近似）；(2) 接受原假设 H_0。

? 真题 9.5：21-22 春夏, 5

设总体 $X \sim N(\mu, 1)$，抽取样本 $X_1, \ldots, X_{16}$，样本均值为 $\overline{X}$。则：

(1) 置信水平 0.95 的 μ 的置信区间为 $(\overline{X}-\delta, \overline{X}+\delta)$，则 $\delta =$ ________；

(2) 独立重复抽取 10 个容量为 16 的样本，得到 10 个置信区间，恰有一个区间不覆盖真值 μ 的概率为 ________。

解题思路：

(1) 总体方差已知，置信区间用正态分布，$\delta = z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$；

(2) 每个区间覆盖真值的概率为 0.95，服从二项分布 $B(10, 0.95)$。

详细解答：

(1) $\alpha = 0.05$，$z_{0.025} = 1.96$，$\sigma = 1$，$n = 16$，$\delta = 1.96 \times \dfrac{1}{\sqrt{16}} = 0.49$；

(2) 设 Y 为不覆盖真值的区间数，$Y \sim B(10, 0.05)$，概率为 $P(Y=1) = \mathrm{C}_{10}^1 \times 0.95^9 \times 0.05^1 = 10 \times 0.95^9 \times 0.05 \approx 0.315$。

故答案依次为 (1) 0.49；(2) 0.315。

注意：

记得带计算器！

9.2 题型十八：拒绝域与假设检验

真题 9.6：24-25 春夏, 12

为研究作者写作用词的特征，选取作家甲的八部作品（短单词比例 x）和作家乙的十部作品（短单词比例 y），计算得 $\overline{x} = 0.231$，$s_1^2 = 0.000212$，$\overline{y} = 0.209$，$s_2^2 = 0.000093$。

(1) 在显著性水平 $\alpha = 0.05$ 下，检验假设 $H_0 : \sigma_1^2 = \sigma_2^2$，$H_1 : \sigma_1^2 \neq \sigma_2^2$；

(2) 根据 (1) 的结论，求 $\mu_1 - \mu_2$ 在 95% 置信水平下的双侧置信区间（保留三位小数），并说明 μ_1 与 μ_2 是否有显著差异。

解题思路：

(1) 方差齐性检验，采用 F 检验，统计量 $F = \dfrac{S_1^2}{S_2^2}$，拒绝域为 $F > F_{\alpha/2}(n1-1, n2-1)$ 或 $F < F_{1-\alpha/2}(n1-1, n2-1)$；

(2) 方差相等时，构造两样本 t 置信区间，枢轴量为 $\dfrac{\overline{X} - \overline{Y} - (\mu1 - \mu2)}{S_w\sqrt{\dfrac{1}{n1} + \dfrac{1}{n2}}} \sim t(n1 + n2 - 2)$，其中 S_w^2 为合并方差。

详细解答：

(1) 方差齐性检验（F 检验）：检验统计量 $F = \dfrac{s_1^2}{s_2^2} = \dfrac{0.000212}{0.000093} \approx 2.28$。

自由度 $df1 = n1 - 1 = 7$，$df2 = n2 - 1 = 9$，显著性水平 $\alpha = 0.05$：

查表得 $F_{0.025}(7, 9) = 4.20$，$F_{0.975}(7, 9) = \dfrac{1}{F_{0.025}(9, 7)} \approx 0.238$。

拒绝域为 $F > 4.20$ 或 $F < 0.238$，而 $0.238 < 2.28 < 4.20$，未落入拒绝域，接受 H_0，即认为 $\sigma_1^2 = \sigma_2^2$。

(2) 合并方差：$S_w^2 = \dfrac{(n1-1)s1^2 + (n2-1)s2^2}{n1 + n2 - 2} = \dfrac{7 \times 0.000212 + 9 \times 0.000093}{16} = 0.000145$，$S_w = \sqrt{0.000145} \approx 0.01204$。

置信水平 95%，自由度 16，分位数 $t_{0.025}(16) = 2.120$，置信区间公式：

$$(\overline{x} - \overline{y}) \pm t_{\alpha/2}(16) \cdot S_w\sqrt{\frac{1}{n1} + \frac{1}{n2}}$$

代入数据：$\overline{x} - \overline{y} = 0.022$，$\sqrt{\dfrac{1}{8} + \dfrac{1}{10}} \approx 0.474$，边际误差：$2.120 \times 0.01204 \times 0.474 \approx 0.012$，置信区间：$(0.022 - 0.012, 0.022 + 0.012) = (0.010, 0.034)$。

因 $0 \notin (0.010, 0.034)$，故 $\mu1$ 与 $\mu2$ 有显著差异。

答案： (1) 接受 H_0；(2) 置信区间为 (0.010, 0.034)，有显著差异。

真题 9.7：23-24 春夏, 9(2)

有 A、B 两种小麦，发芽时间分别服从 $N(\mu_1,\sigma_1^2)$，$N(\mu_2,\sigma_2^2)$。现分别从两种小麦中取 11 和 10 个样本，样本均值分别为 $\overline{x}=99.1$，$\overline{y}=98.9$数据存疑，样本方差分别为 $s_1^2=0.94$，$s_2^2=0.88$。

(1) 求 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 在置信水平为 0.95 下的双侧等尾置信区间，并判断 σ_1^2 是否等于 σ_2^2；

(2) 在 (1) 的基础上，检验 $H_0:\mu_1=\mu_2$；$H_1:\mu_1\neq\mu_2$（显著水平 $\alpha=0.05$）。

解题思路：

(1) 方差比置信区间用 F 分布，包含 1 则认为方差相等；

(2) 方差相等时用两样本 t 检验，构造统计量并对比临界值。

详细解答：

(1) 自由度 $df1=10$，$df2=9$，$\alpha=0.05$，

查表得 $F_{0.025}(10,9)=3.96$，$F_{0.975}(10,9)=\dfrac{1}{F_{0.025}(9,10)}\approx 0.255$；

置信区间：$\left(\dfrac{s_1^2}{s_2^2F_{\alpha/2}(df1,df2)},\dfrac{s_1^2}{s_2^2F_{1-\alpha/2}(df1,df2)}\right)=\left(\dfrac{0.94}{0.88\times 3.96},\dfrac{0.94}{0.88\times 0.255}\right)\approx(0.269,4.147)$；

因置信区间包含 1，故认为 $\sigma_1^2=\sigma_2^2$。

(2) 两样本 t 检验：

合并方差：$s_w^2=\dfrac{(11-1)s_1^2+(10-1)s_2^2}{11+10-2}=\dfrac{10\times 0.94+9\times 0.88}{19}\approx 0.9126$，$s_w\approx 0.955$；

检验统计量：$t=\dfrac{\overline{x}-\overline{y}}{s_w\sqrt{\dfrac{1}{11}+\dfrac{1}{10}}}=\dfrac{99.1-98.9}{0.955\times 0.426}\approx 0.498$；

临界值：$t_{0.025}(19)=2.093$，因 $|t|<2.093$，未落入拒绝域，接受 H_0。

答案： (1) 置信区间为 (0.269, 4.147)，$\sigma_1^2=\sigma_2^2$；(2) 接受原假设 H_0。

真题 9.8：22-23 秋冬, 5

设 $X\sim N(\mu,4)$，从总体简单随机抽取 16 个样本，作出假设 $H_0:\mu=2$，$H_1:\mu<2$，拒绝域 $W=\{\overline{X}>3\}$。则：

(1) 犯第一类错误的概率为 ________；

(2) 当 $\mu=3.5$ 时，犯第二类错误的概率为 ________。

解题思路：

(1) 第一类错误概率即 H_0 成立时 $\overline{X}>3$ 的概率；

(2) 第二类错误概率即 H_1 成立时 $\overline{X}\leq 3$ 的概率。

详细解答：

(1) $\overline{X} \sim N(2, \frac{4}{16}) = N(2, 0.25)$，$P(\overline{X} > 3|H_0) = 1 - \Phi\left(\frac{3-2}{0.5}\right) = 1 - \Phi(2) = 0.0228$。

(2) $\overline{X} \sim N(3.5, 0.25)$，$P(\overline{X} \leq 3|\mu = 3.5) = \Phi\left(\frac{3-3.5}{0.5}\right) = \Phi(-1) = 0.1587$。

故答案依次为 0.0228，0.1587

注意：

这一题给出了拒绝域，求解第一类、第二类错误概率，广义上说也属于本类题型。

真题 9.9：21-22 春夏, 9

从两条糖果生产线上分别抽取容量为 10 的独立样本，计算得 A 线样本均值 $\overline{x}_1 = 100.55$，样本标准差 $s_1 = 5.44$；B 线样本均值 $\overline{x}_2 = 98.08$，样本标准差 $s_2 = 5.31$。假设 A 线服从 $N(\mu_1, \sigma_1^2)$，B 线服从 $N(\mu_2, \sigma_2^2)$，显著性水平 0.05。

(1) 检验 $H_0 : \sigma_1^2 = \sigma_2^2$，$H_1 : \sigma_1^2 \neq \sigma_2^2$；

(2) 假设方差相等，检验 $H_0 : \mu_1 = \mu_2$，$H_1 : \mu_1 \neq \mu_2$。

解题思路：

(1) 方差齐性检验用 F 检验，统计量 $F = \frac{S_1^2}{S_2^2}$；

(2) 方差相等时用两样本 t 检验，构造合并方差和 t 统计量。

详细解答：

(1) 检验统计量 $F = \frac{s_1^2}{s_2^2} = \frac{5.44^2}{5.31^2} \approx \frac{29.59}{28.19} \approx 1.05$；$H_0$ 为真时，$F \sim F(9, 9)$；

拒绝域：$F > F_{0.025}(9, 9) = 4.03$ 或 $F < F_{0.975}(9, 9) = 0.25$；

因 $0.25 < 1.05 < 4.03$，未落入拒绝域，接受 H_0，认为 $\sigma_1^2 = \sigma_2^2$。

(2) 合并方差 $s_w^2 = \frac{(10-1)s_1^2 + (10-1)s_2^2}{10+10-2} = \frac{9 \times (29.59 + 28.19)}{18} \approx 28.89$，$s_w \approx 5.38$；

检验统计量 $t = \frac{\overline{x}_1 - \overline{x}_2}{s_w\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{100.55 - 98.08}{5.38 \times 0.447} \approx \frac{2.47}{2.41} \approx 1.02$；$H_0$ 为真时，$t \sim t(18)$；

拒绝域：$|t| > t_{0.025}(18) = 2.10$；

因 $|t| < 2.10$，未落入拒绝域，接受 H_0，认为 $\mu_1 = \mu_2$。

答案： (1) 接受 H_0，认为方差相等；(2) 接受 H_0，认为均值相等。

真题 9.10：21-22 秋冬, 7

某钢丝标注的断裂强度为 3.5，五金店老板不信，亲自检验，取 16 根钢丝实验，样本均值 $\overline{x} = 3.25$。

假设钢丝断裂强度 $X \sim N(\mu, 0.4^2)$。

(1) 写出原假设、备择假设和拒绝域；

(2) 求老板犯错的概率；

(3) 要使老板犯错的概率 < 0.05，至少取多少根钢丝检验？

解题思路：

(1) 检验均值是否等于 3.5，单侧检验；

(2) 老板犯错的概率即第一类错误概率（H_0 为真时拒绝 H_0）；

(3) 按第一类错误概率要求求解样本量 n。

详细解答：

(1) 原假设 $H_0 : \mu = 3.5$，备择假设 $H_1 : \mu \neq 3.5$，

检验统计量：$Z = \dfrac{\overline{X} - 3.5}{0.4/\sqrt{16}} = 10(\overline{X} - 3.5)$

拒绝域：$W = \{|z| \geq Z_{\alpha/2}\} = \{10(|\overline{x} - 3.5|) \geq Z_{\alpha/2}\}$。

(2) 老板犯错即 H_0 为真，拒绝 H_0

$$P_{H_0}\{|Z| \geq 10(|\overline{x} - 3.5|)\} = 2(1 - \Phi(2.5)) = 0.0124$$

(3) 令 $P_{H_0}\{|Z| \geq \dfrac{|\overline{x} - 3.5|}{0.4/\sqrt{n}}\} < 0.05$，即 $2(1 - \Phi(0.625\sqrt{n})) < 0.05$,

解得 $n \geq (1.96/0.625)^2 = 9.83$，故至少 10 根。

答案： (1) $H_0 : \mu = 3.5$，$H_1 : \mu \neq 3.5$，拒绝域 $\{10(|\overline{x} - 3.5|) \geq Z_{\alpha/2}\}$；(2) 0.0124；(3) 10 根

注意：

这一题没有明确给出采用双边检验，还是单边检验。笔者是按照双边检验做的，单边检验的答案为 (1)$H_1 : \mu < 3.5$(2)0.0062, (3)7 根。

9.3 题型十九：P_ 值并检验假设

真题 9.11：24-25 秋冬, 6(2)

设正态总体 $X \sim N(\mu, \sigma^2)$，$\overline{X} = \dfrac{1}{4}(X_1 + X_2 + X_3 + X_4)$，$S^2 = \dfrac{1}{3}\sum_{i=1}^{4}(X_i - \overline{X})^2$。则：

(1) 设 μ 未知，置信水平为 $1 - \alpha$ 的 σ^2 的单侧置信下限为 ______；

(2) 若 (1) 中单侧置信下限 $P_- = 0.08$，$\alpha = 0.05$，是否接收假设 $H_0 : \mu = 0$？ ______；

解题思路：

(1) 总体正态、μ 未知，用 χ^2 分布构造置信区间，单侧下限公式推导；

(2) 置信下限与假设检验的关系；

详细解答：

(1) 枢轴量 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$（$n=4$），由 $P\left(\dfrac{3S^2}{\sigma^2} \le \chi^2_\alpha(3)\right) = \alpha$，得单侧置信下限：$\dfrac{3S^2}{\chi^2_\alpha(3)}$。

(2) 假设检验判断：$H_0: \mu = 0$ 时，总体 $X \sim N(0, \sigma^2)$，置信下限 0.08 大于 0，说明 σ^2 有下界，不影响 $\mu = 0$ 的合理性，故接收 H_0。

故答案依次为 (1) $\dfrac{3S^2}{\chi^2_\alpha(3)}$；(2) 是；

真题 9.12：23-24 春夏, 4(3)

设正态总体 $X \sim N(\mu, \sigma^2)$（$\mu \in \mathbb{R}$，$\sigma^2 > 0$ 未知），$X_1, \ldots, X_n$ 为来自该总体的简单随机样本（$n \ge 4$），$\overline{X}$ 为样本均值。则：

(3) 当 $\sigma = 2$ 已知时，若 $n = 36$，$\overline{x} = 1.5$，则检验 $H_0: \mu \ge 2$，$H_1: \mu < 2$ 的 p 值为 ________。

解题思路：

(3) 单侧检验的 p 值为标准正态分布下统计量左侧的概率。

详细解答：

(3) 检验统计量 $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{1.5-2}{2/\sqrt{36}} = -1.5$，$p$ 值 $= P(Z < -1.5) = \Phi(-1.5) = 0.0668$。

故答案为 (3) 0.0668 或 $\Phi(-1.5)$

真题 9.13：20-21 秋冬, 6

随机调查某地区 25 个乡的粮食产量，样本均值 1.2 千吨，样本标准差 0.3 千吨，设产量 $X \sim N(\mu, \sigma^2)$（μ, σ^2 未知）。则：

(1) 假设 $H_0: \mu \le 1.1$，$H_1: \mu > 1.1$，相应的 $p\_$ 值为 ________，在显著水平 0.05 下，___（拒绝/不拒绝）原假设；

(2) 总体标准差 σ 的置信度为 95% 的单侧置信上限为 ________。

解题思路：

(1) 用 t 检验，计算统计量，确定 p 值；

(2) 单侧置信上限用 χ^2 分布。

详细解答：

(1) 统计量观测值 $t = \dfrac{1.2-1.1}{0.3/\sqrt{25}} = \dfrac{0.1}{0.06} \approx 1.67$，

$p\_ = P(T \geq 1.67) = 0.054$，故 $p = 0.054 > 0.05$ ，不拒绝 H_0；

(2) 单侧置信上限：$\hat{\sigma}_U = \sqrt{\dfrac{(n-1)S^2}{\chi^2_{0.95}(24)}} = \sqrt{\dfrac{24 \times 0.09}{13.8}} \approx 0.396$。

故答案依次为 (1) 0.054，不拒绝；(2) 0.396

9.4 题型二十：拟合优度检验（略）

好像最近几次的都不在考试范围内了，最起码从笔者考试那年开始...

小结：

本章内容理解难度不大，但记忆的公式较多，注意理解并区分置信区间和假设检验之间的关系。几种枢轴量不要记混，注意题目要算的是单侧还是双侧。

看到这里，考试应该都十拿九稳了吧