

BDIC2005J Probability and Statistics 2019-2020

Question 1:

Vacancy (Each blank 3 marks)

1. There are two events A and B . $P(A) + P(B) = 0.7$, $P(AB) = 0.3$.

Then, $P(A \cup B) = \underline{0.4}$, $P(A\bar{B}) + P(\bar{A}B) = \underline{0.1}$.

Solution:

Given

$$P(A) + P(B) = 0.7, \quad P(AB) = 0.3.$$

We have

$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.7 - 0.3 = 0.4.$$

We known that

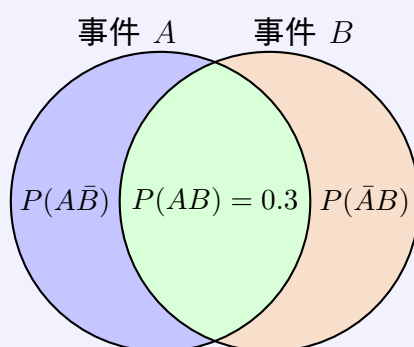
$$P(A\bar{B}) = P(A) - P(AB), \quad P(\bar{A}B) = P(B) - P(AB).$$

Hence,

$$P(A\bar{B}) + P(\bar{A}B) = P(A) + P(B) - 2P(AB) = 0.7 - 2 \times 0.3 = 0.1.$$

解题方法：画图解决

1. 几何方法：利用韦恩图分析该问题



2. 事件与韦恩图的对应关系解释

- **并集 $P(A \cup B)$** : 对应图中所有的颜色区域。公式 $P(A) + P(B) - P(AB)$ 的意思是: A 圆加 B 圆时, 中间的重叠部分被算了两次, 所以要减去一次。
- **$P(A\bar{B}) + P(\bar{A}B)$** : 对应图中 A 独有和 B 独有的部分 (左右两个月牙)。公

式 $P(A) + P(B) - 2P(AB)$ 的意思是：从两个圆的总面积中把中间重叠的部分完全扣除。

3. 易错点

- 题目给出的条件是 $P(A) + P(B)$ 的“整体和”，不需要分别求出 $P(A)$ 或 $P(B)$ 的具体数值。
- 要区分“恰有一个发生”与“至少一个发生”的区别。

2. Suppose the random variable X follows the Poisson distribution with the parameter $\lambda = 1$, $P(X = 2) = \frac{e^{-1}}{2}$, $E(X^2) = 2$.

Solution:

If $X \sim \text{Poisson}(\lambda = 1)$.

Then

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Thus,

$$P(X = 2) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2}.$$

For a Poisson random variable,

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda.$$

Using

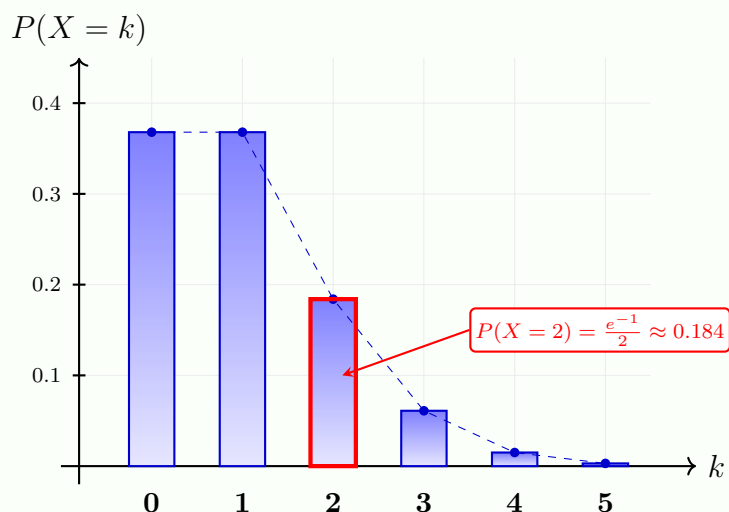
$$E(X^2) = \text{Var}(X) + [E(X)]^2,$$

We obtain

$$E(X^2) = 1 + 1^2 = 2.$$

要点：泊松分布

1. 概率分布图示 (当 $\lambda = 1$ 时)



2. 泊松分布的特征

- 形状：当 $\lambda = 1$ 时，泊松分布在 $k = 0$ 和 $k = 1$ 处取得最大概率（均为 e^{-1} ）。随着 k 的增大，概率呈阶乘级速度衰减。
- 二阶矩的意义： $E(X^2)$ 指的是把随机变量平方之后再取期望。

二阶矩与方差的关系：

随机变量 X 的方差定义为： $\text{Var}(X) = E[(X - E(X))^2]$ 。

对平方项展开，有 $(X - E(X))^2 = X^2 - 2XE(X) + [E(X)]^2$ 。

因此， $\text{Var}(X) = E(X^2 - 2XE(X) + [E(X)]^2)$ 。

利用期望的线性性质，展开得到 $\text{Var}(X) = E(X^2) - 2E(X)E(X) + [E(X)]^2$ 。

化简可得 $\text{Var}(X) = E(X^2) - [E(X)]^2$ 。

等价地，也可以写为 $E(X^2) = \text{Var}(X) + [E(X)]^2$ 。

即，二阶矩 = 方差 + 均值的平方

3. 技巧点

- 泊松分布的“期望等于方差”： $E(X) = \text{Var}(X) = \lambda$ 。
- 记住： $E(X^2) = \lambda^2 + \lambda$ 。

3. Suppose the random variables X and Y are mutually independent, $X \sim N(3, 3^2)$, $Y \sim N(1, 2^2)$, $Z = X - 2Y$, $Z \sim \underline{N(1, 25)}$, $P(-4 < Z < 6) = \underline{0.6826}$.
($\Phi(1) = 0.8413$)

Solution:

Since X and Y are independent and normally distributed, Z is also normally distributed.

$$E(Z) = E(X) - 2E(Y) = 3 - 2 \times 1 = 1,$$

$$\text{Var}(Z) = \text{Var}(X) + 4\text{Var}(Y) = 3^2 + 4 \times 2^2 = 9 + 16 = 25.$$

Therefore,

$$Z \sim N(1, 25).$$

Standardize Z :

$$P(-4 < Z < 6) = P\left(\frac{-4-1}{5} < \frac{Z-1}{5} < \frac{6-1}{5}\right) = P(-1 < Z_0 < 1),$$

where $Z_0 \sim N(0, 1)$.

Using $\Phi(1) = 0.8413$,

$$P(-1 < Z_0 < 1) = 2\Phi(1) - 1 = 2 \times 0.8413 - 1 = 0.6826.$$

提示：正态分布的线性性质与标准化

1. 正态分布的线性性质

若随机变量 X 与 Y 相互独立，且 $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ ，则线性组合仍然服从正态分布：

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2).$$

2. 关键点

- 均值计算： $E(X - 2Y) = E(X) - 2E(Y) = 3 - 2(1) = 1$ 。

- 方差计算： $\text{Var}(X - 2Y) = \text{Var}(X) + (-2)^2\text{Var}(Y)$ 。

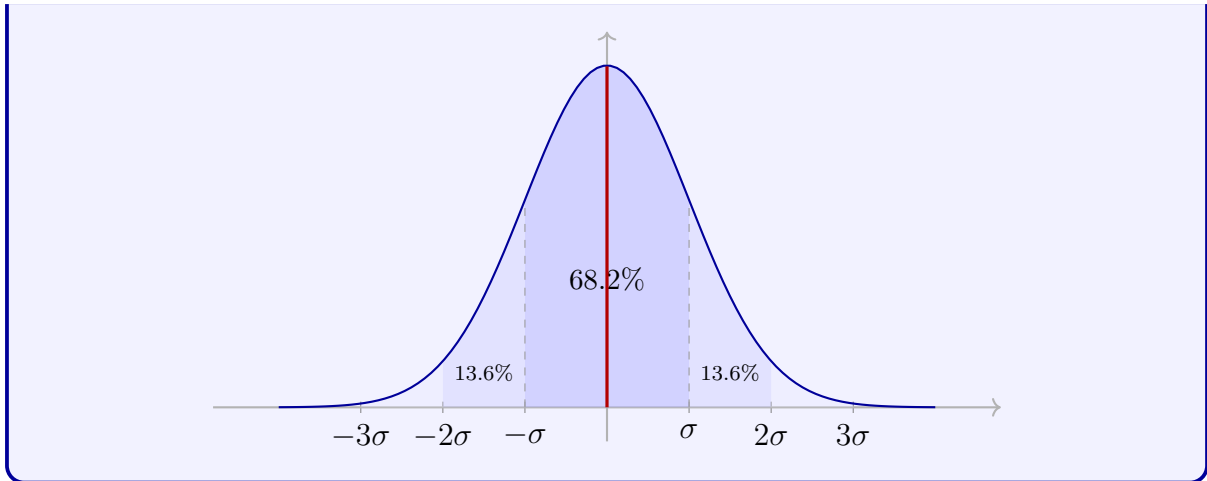
注意：中间是减号，但方差贡献是系数平方相加，即 $3^2 + 4 \times 2^2 = 25$ 。

3. 标准化

本题计算 $P(-4 < Z < 6)$ 等价于计算标准正态分布 Z_0 在 ± 1 个标准差之间的面积：

$$P\left(\frac{-4-1}{5} < Z_0 < \frac{6-1}{5}\right) = P(-1 < Z_0 < 1)$$

4. 标准正态分布图示



Question 2: (14 marks)

A company X has three factories F_1 , F_2 and F_3 , all of which produce similar automotive parts. Suppose that F_1 , F_2 and F_3 produce 20%, 45% and 35% of the total production, respectively. On the average, 2% of the parts produced by F_1 is unqualified. For F_2 and F_3 , the corresponding unqualified percent are 5% and 1%, respectively.

1. Find the probability that an automotive part produced by the company X is unqualified;
2. If one product was selected randomly and found to be unqualified, what is the probability that it was made by F_3 .

Solution:

1. Let U denote the event that a product is unqualified.

Define $P(F_i)$ the probability that the product was produced in F_i , then

$$P(F_1) = 0.20, \quad P(F_2) = 0.45, \quad P(F_3) = 0.35.$$

The corresponding unqualified rates are

$$P(U|F_1) = 0.02, \quad P(U|F_2) = 0.05, \quad P(U|F_3) = 0.01.$$

By the law of total probability,

$$\begin{aligned} P(U) &= P(F_1)P(U|F_1) + P(F_2)P(U|F_2) + P(F_3)P(U|F_3) \\ &= 0.20 \times 0.02 + 0.45 \times 0.05 + 0.35 \times 0.01 \\ &= 0.004 + 0.0225 + 0.0035 \\ &= 0.03. \end{aligned}$$

2. By Bayes' theorem,

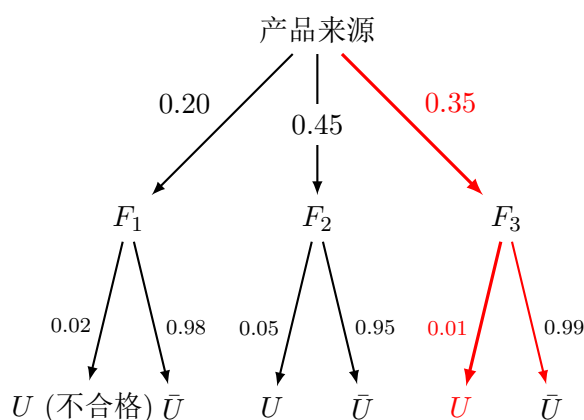
$$P(F_3|U) = \frac{P(F_3)P(U|F_3)}{P(U)} = \frac{0.35 \times 0.01}{0.03} = \frac{0.0035}{0.03} = \frac{7}{60}.$$

提示：全概率公式与贝叶斯定理

一、事件

- **完备事件组**：本题中工厂 F_1, F_2, F_3 构成了样本空间的一个划分（即产品只能由这三家工厂之一生产，且三者概率之和为 1）。
- **条件概率**：题目中给出的“某工厂生产的不合格率”即为条件概率 $P(U|F_i)$ ，而非联合概率。

二、概率树



三、全概率公式

- **适用场景**：当一个事件（如“不合格” U ）的发生受到多种不同背景（如不同工厂 F_i ）影响时，使用该公式求总概率。
- **直观理解**：全概率是各路径概率的加权平均。在本题中， $P(U)$ 是三家工厂不合格贡献的总和。

四、贝叶斯定理 (Bayes' Theorem)

- **核心逻辑**：贝叶斯定理用于“执果索因”。已知结果已经发生（已知产品不合格），反推它是由某个特定原因（ F_3 工厂）造成的概率。
- **计算技巧**：分子是全概率公式中的某一项（特定路径），分母是全概率公式计算出的总和。

五、注意

- **因果关系**：题目问“已知... 发现不合格，求... 的概率”，立刻联想到贝叶斯公式。

- 验证：各一级分枝概率之和必须等于 1 ($0.2 + 0.45 + 0.35 = 1$)，否则样本空间划分有误。
- 计算：将小数化为分数进行计算。

Question 3: (21 marks)

The joint probability density function of (X, Y) is

$$f(x, y) = \begin{cases} c(1 - y), & 0 \leq x \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find: 1. the constant c ; 2. the marginal density functions for X and Y ; 3. the correlation coefficient $\rho(X, Y)$.

Solution to Question 3

1. Since $f(x, y)$ is a joint density function,

$$\int_0^1 \int_0^y c(1 - y) dx dy = 1.$$

Compute the integral:

$$\begin{aligned} \int_0^1 \int_0^y c(1 - y) dx dy &= \int_0^1 c(1 - y)y dy \\ &= c \int_0^1 (y - y^2) dy \\ &= c \left(\frac{1}{2} - \frac{1}{3} \right) = \frac{c}{6}. \end{aligned}$$

Thus,

$$\frac{c}{6} = 1 \quad \Rightarrow \quad c = 6.$$

2. Marginal density of X :

$$\begin{aligned} f_X(x) &= \int_x^1 6(1 - y) dy \\ &= 6 \left[-\frac{(1 - y)^2}{2} \right]_x^1 \\ &= 3(1 - x)^2, \quad 0 \leq x \leq 1. \end{aligned}$$

Marginal density of Y :

$$\begin{aligned} f_Y(y) &= \int_0^y 6(1 - y) dx \\ &= 6y(1 - y), \quad 0 \leq y \leq 1. \end{aligned}$$

3. First compute expectations.

$$E(X) = \int_0^1 x f_X(x) dx = \int_0^1 3x(1-x)^2 dx = \frac{1}{4}.$$

$$E(Y) = \int_0^1 y f_Y(y) dy = \int_0^1 6y^2(1-y) dy = \frac{1}{2}.$$

$$E(XY) = \int_0^1 \int_0^y xy \cdot 6(1-y) dx dy = \frac{3}{20}.$$

Thus,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{3}{20} - \frac{1}{4} \times \frac{1}{2} = \frac{1}{40}.$$

Next compute variances.

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{3}{80}, \quad \text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{1}{20}.$$

Therefore,

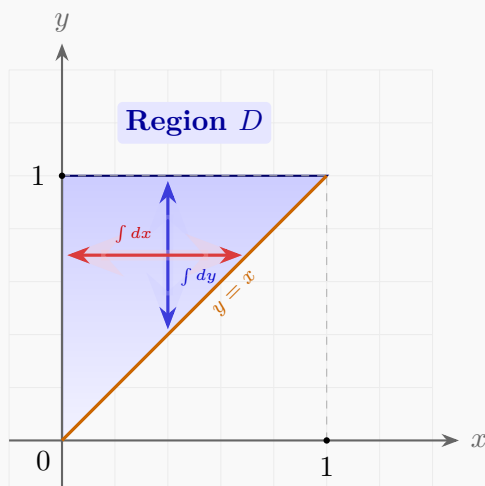
$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{1/40}{\sqrt{(3/80)(1/20)}} = \frac{1}{\sqrt{3}}.$$

提示：二维随机变量与相关性分析

一、积分区域的确定 (Integration Region)

- **核心难点：** 本题的积分区域是 $0 \leq x \leq y \leq 1$ 。在平面直角坐标系中，这是一个由直线 $y = x$ 、 $y = 1$ 以及 y 轴围成的三角形区域。
- **积分限写法：**
 - 对 x 求边缘分布时，固定 y ，让 x 从 0 变到 y （水平扫描）。
 - 对 y 求边缘分布时，固定 x ，让 y 从 x 变到 1 （垂直扫描）。

二、图示



二、边缘概率密度 (Marginal Density)

- 公式: $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ 以及 $f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$ 。
- 注意: 求出的边缘分布必须注明自变量的取值范围 (本题均为 $[0, 1]$), 否则无法求出正确结果

三、相关系数 $\rho(X, Y)$ 的计算步骤

1. 期望: 分别求出 $E(X)$ 、 $E(Y)$ 以及混合矩 $E(XY) = \iint xyf(x, y) dxdy$ 。
2. 协方差: $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ 。
3. 方差: 计算 $\text{Var}(X)$ 和 $\text{Var}(Y)$ 。注意使用公式 $\text{Var}(X) = E(X^2) - [E(X)]^2$ 。
4. 标准化: 相关系数是将协方差除以两个标准差的乘积, 其值一定在 $[-1, 1]$ 之间。

四、需要注意的地方

- 计算过程: 积分计算易错, 建议在求 $E(X^n)$ 时利用已知的边缘分布计算, 比直接用联合分布进行二重积分更简便。
- 相关性判断: 若 $\rho = 0$, 称 X 与 Y 不相关; 本题 $\rho = 1/\sqrt{3} \approx 0.577$, 说明两者存在正的线性相关关系。

Question 4: (7 marks)

Let X_1, \dots, X_n constitute a random sample from the probability density function given by

$$f_{\theta}(x) = \begin{cases} \frac{2}{\theta^2}(\theta - x), & 0 \leq x \leq \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

Find an estimator for θ by using the method of moments.

Solution to Question 4

First, calculate $E(X)$:

$$\begin{aligned}
 E(X) &= \int_0^\theta x \cdot \frac{2}{\theta^2}(\theta - x) dx \\
 &= \frac{2}{\theta^2} \int_0^\theta (\theta x - x^2) dx \\
 &= \frac{2}{\theta^2} \left[\frac{\theta x^2}{2} - \frac{x^3}{3} \right]_0^\theta \\
 &= \frac{2}{\theta^2} \left(\frac{\theta^3}{2} - \frac{\theta^3}{3} \right) \\
 &= \frac{\theta}{3}.
 \end{aligned}$$

Then equate $E(X)$ with the sample mean \bar{X} :

$$\bar{X} = \frac{\theta}{3}.$$

Finally, we get the result:

$$\hat{\theta}_{\text{MM}} = 3\bar{X}.$$

提示：矩估计法 (Method of Moments)

一、矩估计法的核心思想

- **基本原理：**利用样本的 k 阶矩（如样本均值 \bar{X} ）作为总体 k 阶矩（如期望 $E(X)$ ）的估计量。
- **理论依据：**大数定律。当样本量 n 足够大时，样本平均值依概率收敛于总体期望。

二、解题步骤

1. **计算总体矩：**根据给定的概率密度函数 $f_\theta(x)$ ，计算期望 $E(X) = \int_{-\infty}^{+\infty} x f_\theta(x) dx$ 。此时结果通常是关于未知参数 θ 的函数。
2. **建立方程：**令总体期望等于样本均值，即 $E(X) = \bar{X}$ 。
3. **反解参数：**从方程中解出 θ ，得到的表达式即为矩估计量 $\hat{\theta}_{\text{MM}}$ 。

三、注意事项

- **积分区间：**在计算 $E(X)$ 时，务必注意分段函数的有效区间（本题为 0 到 θ ），区间外的积分为 0。

- **多参数情况:** 如果分布中有两个未知参数 (如 μ 和 σ^2), 则需要利用二阶矩, 建立方程组:

$$\begin{cases} E(X) = \bar{X} \\ E(X^2) = \frac{1}{n} \sum X_i^2 \end{cases}$$

- **记号规范:** 最终结果写成 $\hat{\theta}$, 并在下标注明 MM (Method of Moments) 以示区分。

Question 5: (14 marks)

The Rayleigh density function is given by

$$f_{\theta}(x) = \begin{cases} \frac{2x}{\theta} e^{-x^2/\theta}, & x \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

1. First show that X^2 has an exponential distribution with mean θ ;
2. If X_1, \dots, X_n denote a random sample from a Rayleigh distribution, find the MLE for θ .

Solution to Question 5

1. Let $Y = X^2$. For $y > 0$,

$$\begin{aligned} P(Y \leq y) &= P(X \leq \sqrt{y}) = \int_0^{\sqrt{y}} \frac{2x}{\theta} e^{-x^2/\theta} dx \\ &= \left[-e^{-x^2/\theta} \right]_0^{\sqrt{y}} \\ &= 1 - e^{-y/\theta}. \end{aligned}$$

Thus, the density of Y is

$$f_Y(y) = \frac{d}{dy}(1 - e^{-y/\theta}) = \frac{1}{\theta} e^{-y/\theta}, \quad y > 0,$$

which is an exponential distribution with mean θ .

2. Let X_1, \dots, X_n be a random sample. The likelihood function is

$$L(\theta) = \prod_{i=1}^n \frac{2X_i}{\theta} e^{-X_i^2/\theta}.$$

The log-likelihood is

$$\ell(\theta) = n \ln 2 + \sum_{i=1}^n \ln X_i - n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n X_i^2.$$

Differentiate with respect to θ :

$$\frac{d\ell}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n X_i^2.$$

Setting this equal to zero gives

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

提示：变量代换与极大似然估计

一、变量代换法 (Transformation Method)

- 分布函数法：求 $Y = g(X)$ 的分布时，先求其累积分布函数 $F_Y(y) = P(g(X) \leq y)$ ，再求导得到密度函数 $f_Y(y)$ 。
- 识别分布：得到 $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ 后，应能立即识别出这是均值为 θ 的指数分布 $\text{Exp}(1/\theta)$ 。

二、极大似然估计 (MLE) 的标准步骤

1. 写出似然函数 $L(\theta)$ ：即所有观测值联合密度的乘积 $\prod_{i=1}^n f(x_i; \theta)$ 。
2. 取对数 $\ell(\theta)$ ：指数族分布（包含 e 的函数）取对数后可以把乘法变加法，极大地简化求导过程。
3. 求导并令其为零：解似然方程 $\frac{d\ell}{d\theta} = 0$ 。
4. 求出估计量 $\hat{\theta}$ ：通常最后需要用样本观察值（如 $\sum X_i^2$ ）来表达 θ 。

三、注意事项

- 求导细节：注意 $\frac{d}{d\theta}(-n \ln \theta) = -\frac{n}{\theta}$ 以及 $\frac{d}{d\theta}(-\frac{k}{\theta}) = \frac{k}{\theta^2}$ 的符号变化。
- 结果检查：MLE 的结果通常与样本的某种“平均”形式相关。在本题中，由于 X^2 的均值是 θ ，根据大数定律， θ 的估计量自然就是 X^2 的样本均值 $\frac{1}{n} \sum X_i^2$ 。

Question 6: (14 marks)

The specifications for a certain kind of ribbon call for a mean breaking strength of 185 pounds and a standard deviation of 5 pounds. Ten pieces randomly selected from different rolls have breaking strengths of

171.6, 191.8, 178.3, 184.9, 189.1, 181.5, 176.4, 187.3, 190.7, 180.2

pounds.

Under the assumption that the breaking strength is normally distributed as $N(\mu, \sigma^2)$, at a level of significance 0.05, please determine whether the specifications for the mean and standard deviation are conformed.

Solution to Question 6

The specifications are

$$\mu_0 = 185, \quad \sigma_0 = 5.$$

From the data above,

$$\bar{X} = 183.37, \quad S^2 = 39.56, \quad n = 10$$

1. Test for the mean

Hypotheses: $H_0 : \mu = 185, \quad H_1 : \mu \neq 185$.

The test statistic is

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{183.37 - 185}{\sqrt{39.56/10}} = -0.82.$$

At significance level 0.05, the critical value is $t_{0.025,9} = 2.262$.

Since $|t| < 2.262$, we fail to reject H_0 . At the 5% significance level, the mean specifications are conformed.

2. Test for the variance

Hypotheses: $H_0 : \sigma^2 = 25, \quad H_1 : \sigma^2 \neq 25$.

The test statistic is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9 \times 39.56}{25} = 14.24.$$

The critical values are $\chi_{0.025,9}^2 = 2.70$ and $\chi_{0.975,9}^2 = 19.02$.

Since $2.70 < 14.24 < 19.02$, we fail to reject H_0 . At the 5% significance level, the standard deviation specifications are conformed.

第 6 题：正态总体均值与方差的双侧检验

【已知条件】

规格要求： $\mu_0 = 185, \sigma_0 = 5$ 。样本量 $n = 10$ 。

计算样本指标：平均值 $\bar{X} = 183.37$ ，样本方差 $S^2 = 39.56$ 。

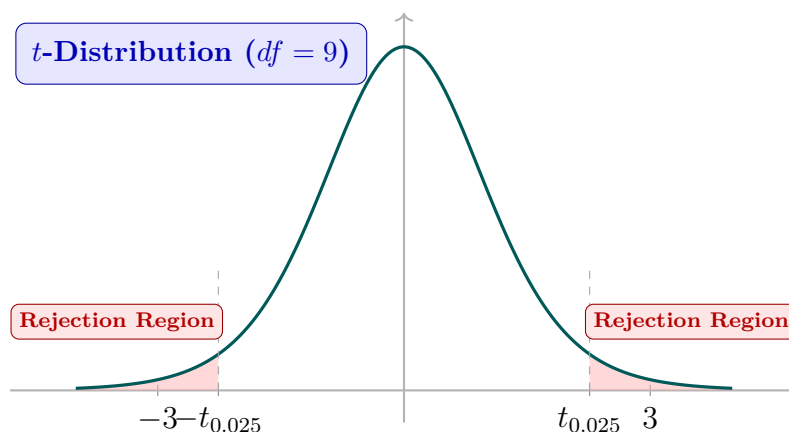
一、均值检验 (t 检验)

- 建立假设: $H_0: \mu = 185$ vs $H_1: \mu \neq 185$

- 计算统计量:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{183.37 - 185}{\sqrt{39.56/10}} \approx -0.82$$

- 确定判定准则: 显著性水平 $\alpha = 0.05$, 自由度 $df = 9$, 查表得临界值 $t_{0.025}(9) = 2.262$ 。



结论: 因为 $|t| = 0.82 < 2.262$, 观测值落入接受域, 故不拒绝 H_0 。在 5% 显著水平下, 均值符合规格。

二、方差检验 (χ^2 检验)

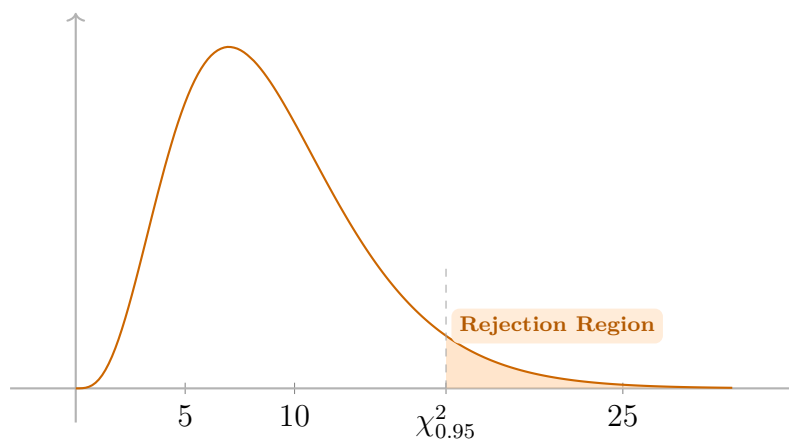
- 建立假设: $H_0: \sigma^2 = 25$ vs $H_1: \sigma^2 \neq 25$

- 计算统计量:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9 \times 39.56}{25} \approx 14.24$$

- 确定判定准则: 显著性水平 $\alpha = 0.05$ 。查表得临界值:

左侧 $\chi_{0.975}^2(9) = 2.70$, 右侧 $\chi_{0.025}^2(9) = 19.02$ 。



结论：因为 $2.70 < 14.24 < 19.02$ ，统计量落入接受域，故不拒绝 H_0 。方差亦符合规格。

BDIC2005J Probability and Statistics 2018-2019

Question 1:

Vacancy (Each blank 3 marks)

(1). There are two events A and B . $P(A) = 0.1, P(A \cup B) = 0.4$.

If A and B are mutually exclusive, $P(B) =$.

If A and B are mutually independent, $P(B) =$.

Solution:

For mutually exclusive events,

$$P(A \cup B) = P(A) + P(B).$$

Thus,

$$0.4 = 0.1 + P(B),$$

which gives

$$P(B) = 0.3$$

For independent events,

$$P(A \cap B) = P(A)P(B).$$

Using the union formula,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

we have

$$0.4 = 0.1 + P(B) - 0.1P(B).$$

Hence,

$$0.4 = 0.1 + 0.9P(B),$$

$$0.9P(B) = 0.3,$$

and therefore

$$P(B) = \frac{1}{3}$$

考点：“互斥”和“独立”

1. 核心考点：两种关系的本质区别

- **互斥 (Mutually Exclusive)**: 两个事件“有你没我”，不能同时发生，所以交集 $P(AB) = 0$ 。计算并集时直接求和即可： $P(A \cup B) = P(A) + P(B)$ 。
- **独立 (Independent)**: 指的是两个事件“互不干涉”，一个发不发生不影响另一个。此时交集的概率等于概率的乘积： $P(AB) = P(A)P(B)$ 。

2. 易错点

- **减项**: 在处理独立事件时，牢记并集公式 $P(A \cup B) = P(A) + P(B) - P(AB)$ 后面那个减项。只有互斥时那个减项才为 0。
- **代数计算**: 独立情况下的方程 $0.4 = 0.1 + P(B) - 0.1P(B)$ ，可以看作是 $0.3 = (1 - 0.1)P(B)$ 。解方程时别在小数点上栽跟头。

3. 解题技巧

- 看到“互斥”， $P(AB) = 0$ 。
- 看到“独立”， $P(AB) = P(A)P(B)$ 。
- 无论哪种情况，通用的并集公式 $P(A \cup B) = P(A) + P(B) - P(AB)$ 。

(2). Suppose the random variable $X \sim U(a, b)$. $E(X) = 5, D(X) = 3, a = ?, b = ?$.

Solution:

For a uniform distribution,

$$E(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}.$$

From the expectation,

$$\frac{a+b}{2} = 5 \Rightarrow a+b = 10.$$

From the variance,

$$\frac{(b-a)^2}{12} = 3 \Rightarrow (b-a)^2 = 36,$$

which implies

$$b-a = 6.$$

Solving the system,

$$\begin{cases} a + b = 10, \\ b - a = 6, \end{cases}$$

we obtain

$$a = 2, \quad b = 8$$

考点：均匀分布

1. 核心公式

- 均值 (Expectation): $E(X) = \frac{a+b}{2}$ 。直观理解就是区间的“中点”。
- 方差 (Variance): $D(X) = \frac{(b-a)^2}{12}$ 。容易记混的地方：分母是 12，分子是区间长度的平方。

2. 易错点

- 开方：计算 $(b-a)^2 = 36$ 时，虽然代数上有 ± 6 两种可能，但根据均匀分布定义，上界 b 一定大于下界 a ，所以直接取 $b-a=6$ 即可。
- 公式：别把方差的分母记成 2 或者 4。

3. 解题技巧

- 这类题本质上是“解二元一次方程组”。
- 第一步：根据均值算出 $a+b$ 。
- 第二步：根据方差算出 $b-a$ 。

(3). Let $X_1, X_2, \dots, X_n (n > 2)$ be a sample from $N(\mu, \sigma^2)$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

$$\sqrt{n}(\bar{X} - \mu)/\sqrt{S^2} \sim, \quad (n-1)S^2/\sigma^2 \sim.$$

Solution:

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

and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

考点：数理统计的三大分布

1. 核心考点：由正态分布衍生的分布

- 卡方分布 (χ^2): 用来描述“方差”的分布。公式 $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ 。这里的自由度是 $n-1$ ，因为在计算 S^2 时掉了一个自由度（用掉了样本均值 \bar{X} ）。
- t 分布：当总体方差 σ^2 未知，我们被迫用样本方差 S^2 来代替它时，原本的正态分布就变成了 t 分布。公式为 $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ 。

2. 易错点

- 对应关系：记住“方差比方差”通常对应卡方；“均值比标准差”通常对应 t 分布。
- 自由度：这两个统计量的自由度都是 $n-1$ ，千万不要手抖写成 n 。
- 符号：注意 t 分布公式里分母是 S （样本标准差），而卡方公式里是 S^2 （样本方差）。

3. 解题技巧

- \sqrt{n} 出现在分子上，且分母有 S 的，是 t 分布。
- 有 $(n-1)S^2$ 这种的，是 χ^2 分布。
- 记忆： t 分布是胖的正态分布”，随着 n 变大，它会越来越像标准正态分布 $N(0, 1)$ 。

(4). Let X_1, \dots, X_{10} be a sample from $N(\mu, \sigma^2)$. We can calculate the sample mean 5 and sample variance 0.25 are respectively. μ and σ^2 are unknown parameters.

The confidence interval of μ with confidence level 95% is $[\cdot, \cdot]$.

The confidence interval of σ^2 with confidence level 95% is $[\cdot, \cdot]$.

Solution:

Since σ^2 is unknown, the 95% confidence interval for μ is

$$\bar{X} \pm t_{0.975}(n-1) \frac{S}{\sqrt{n}}.$$

With $t_{0.975}(9) = 2.262$ and $S = 0.5$,

$$\frac{S}{\sqrt{n}} = \frac{0.5}{\sqrt{10}}.$$

Thus, the confidence interval is

$$[4.643, 5.357]$$

The 95% confidence interval for σ^2 is

$$\left[\frac{(n-1)S^2}{\chi_{0.975}^2(n-1)}, \frac{(n-1)S^2}{\chi_{0.025}^2(n-1)} \right].$$

Using $\chi_{0.975}^2(9) = 19.023$ and $\chi_{0.025}^2(9) = 2.700$,

$$(n-1)S^2 = 9 \times 0.25 = 2.25.$$

Therefore, the confidence interval is

$$[0.118, 0.833]$$

核心要点：单正态总体的区间估计

1. 均值 μ 的置信区间 (σ^2 未知)

- 核心公式： $\left[\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right]$ 。
- 关键点：因为总体方差 σ^2 未知，必须使用 t 分布。注意分母使用的是样本标准差 S 而非方差 S^2 。

2. 方差 σ^2 的置信区间

- 核心公式： $\left[\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \right]$ 。
- 注意：卡方分布是非对称的，因此区间的左端点对应较大的分位数（分母大，数值小），右端点对应较小的分位数。

3. 易错点提醒

- 查表分位数：置信水平为 95% 时，对应双侧显著性水平 $\alpha = 0.05$ ，查表时应找 $\alpha/2 = 0.025$ 或 $1 - \alpha/2 = 0.975$ 对应的数值。
- 样本标准差与方差：题目给出的“样本方差 0.25”是 S^2 ，计算均值区间时要记得开方取 $S = 0.5$ 。
- 自由度 $n - 1$ ：本题样本量 $n = 10$ ，所有查表操作均需基于自由度 $df = 9$ 。

Question 2: (14 marks)

5 of the 8 guns have been calibrated, 3 have not been calibrated. The probability of a shooter hitting a target with a calibrated gun is 0.8; The probability of the shooter hitting target using an uncalibrated gun is 0.3.

- (a) Find the probability of the shooter not hitting target;
 (b) Now take one of the 8 guns and shoot, the result is that the shooter hits the target. Find the probability that the gun used to shoot is a calibrated gun.

Solution:

There are 8 guns in total, among which 5 are calibrated and 3 are uncalibrated. The probability that the shooter hits the target using

- a calibrated gun is 0.8,
- an uncalibrated gun is 0.3.

(a) Probability that the shooter does not hit the target

First, compute the probability of choosing each type of gun:

$$P(C) = \frac{5}{8}, \quad P(U) = \frac{3}{8}.$$

The probability of missing the target given each gun type is

$$P(M | C) = 1 - 0.8 = 0.2, \quad P(M | U) = 1 - 0.3 = 0.7.$$

By the law of total probability,

$$P(M) = P(M | C)P(C) + P(M | U)P(U).$$

Thus,

$$P(M) = 0.2 \cdot \frac{5}{8} + 0.7 \cdot \frac{3}{8} = \frac{1}{8} + \frac{2.1}{8} = \frac{3.1}{8}.$$

$$P(\text{not hit}) = \frac{3.1}{8}$$

(b) Probability that the gun is calibrated given a hit

Let H denote the event that the target is hit.

First compute $P(H)$:

$$P(H) = P(H | C)P(C) + P(H | U)P(U) = 0.8 \cdot \frac{5}{8} + 0.3 \cdot \frac{3}{8} = \frac{4.9}{8}.$$

Using Bayes' theorem,

$$P(C | H) = \frac{P(H | C)P(C)}{P(H)} = \frac{0.8 \cdot \frac{5}{8}}{\frac{4.9}{8}} = \frac{4}{4.9}.$$

$$P(C | H) = \frac{4}{4.9}$$

核心要点：全概率公式与贝叶斯推断

1. 知识点定位：分阶段概率模型

- **全概率公式**：用于计算“结果”的总概率。当一个事件（如“没击中”）可以通过多种途径（如“用校准枪”或“用未校准枪”）发生时，我们需要加权求和。
- **贝叶斯定理**：用于“由果溯因”。在已知结果发生的条件下，计算它是由某个特定原因引起的概率。

2. 解题逻辑梳理

- **第一步：理清概率树。**
 - 第一层：选枪。 $P(C) = 5/8$, $P(U) = 3/8$ 。
 - 第二层：射击结果。注意“击中”与“未击中”互为对立事件。
- **第二步：对立事件的应用。**在第 (a) 问中，已知击中概率为 0.8，则未击中概率 $P(M|C) = 1 - 0.8 = 0.2$ 。
- **第三步：分母是对子集的汇总。**在第 (b) 问使用贝叶斯公式时，分母 $P(H)$ 实际上是所有能导致“击中”的路径概率之和。

3. 易错点提醒

- **事件定义模糊**：建议在草稿纸上明确列出 C （校准枪）、 U （未校准枪）、 H （击中）等符号，避免计算时将概率放错位置。
- **忽视互斥性**：全概率公式的前提是各种原因（校准与未校准）必须是互斥且构成完备事件组的（即枪只有这两类）。
- **计算陷阱**：最后的计算结果（如 $4/4.9$ ）可以化简为 $40/49$ 。在概率题中，分数形式通常比小数更精确。

Question 3: (14 marks)

Suppose that the probability density function of the continuous random variable X is

$$f(x) = \begin{cases} \frac{1}{c^2 + x^2}, & x \in (0, \frac{\pi}{4}); \\ 0, & \text{elsewhere.} \end{cases}$$

Find:

- (a) The value of constant c ;
- (b) $P\left(\frac{\sqrt{3}\pi}{12} < X \leq \frac{\pi}{3}\right)$;
- (c) $E(X)$.

Solution:**(a) Determination of constant c**

Since $f(x)$ is a probability density function, it must satisfy

$$\int_0^{\pi/4} \frac{1}{c^2 + x^2} dx = 1.$$

Recall that

$$\int \frac{1}{c^2 + x^2} dx = \frac{1}{c} \arctan\left(\frac{x}{c}\right).$$

Therefore,

$$\frac{1}{c} \left[\arctan\left(\frac{x}{c}\right) \right]_0^{\pi/4} = 1,$$

which gives

$$\frac{1}{c} \arctan\left(\frac{\pi}{4c}\right) = 1.$$

Hence,

$$\arctan\left(\frac{\pi}{4c}\right) = c$$

(b) Probability $P\left(\frac{\sqrt{3}\pi}{12} < X \leq \frac{\pi}{3}\right)$

Since the support of X is $(0, \frac{\pi}{4})$, note that

$$\frac{\pi}{3} > \frac{\pi}{4}.$$

Thus,

$$P\left(\frac{\sqrt{3}\pi}{12} < X \leq \frac{\pi}{3}\right) = P\left(\frac{\sqrt{3}\pi}{12} < X \leq \frac{\pi}{4}\right).$$

Compute the probability:

$$\int_{\frac{\sqrt{3}\pi}{12}}^{\frac{\pi}{4}} \frac{1}{c^2 + x^2} dx = \frac{1}{c} \left[\arctan\left(\frac{x}{c}\right) \right]_{\frac{\sqrt{3}\pi}{12}}^{\frac{\pi}{4}}.$$

Thus,

$$P = \frac{1}{c} \left[\arctan\left(\frac{\pi}{4c}\right) - \arctan\left(\frac{\sqrt{3}\pi}{12c}\right) \right]$$

(c) Expectation $E(X)$

The expectation of X is

$$E(X) = \int_0^{\pi/4} x f(x) dx = \int_0^{\pi/4} \frac{x}{c^2 + x^2} dx.$$

Using the substitution $u = c^2 + x^2$, we obtain

$$\int \frac{x}{c^2 + x^2} dx = \frac{1}{2} \ln(c^2 + x^2).$$

Therefore,

$$E(X) = \frac{1}{2} [\ln(c^2 + x^2)]_0^{\pi/4} = \frac{1}{2} \ln\left(\frac{c^2 + (\frac{\pi}{4})^2}{c^2}\right).$$

$$E(X) = \frac{1}{2} \ln\left(1 + \frac{\pi^2}{16c^2}\right)$$

核心要点：连续型随机变量的微积分应用

1. 知识点定位：密度函数

- **归一化**：PDF 在整个实数域上的积分必须等于 1。在解决第 (a) 问时，这是确定待定系数 c 的唯一途径。
- **概率是面积**：连续型随机变量在某个区间内的概率，等于该区间上密度函数的定积分。
- **期望的定义**： $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ 。可以形象地理解为所有可能取值 x 与其对应概率密度 $f(x)$ 的加权平均积分。

2. 微积分

- **反三角函数积分**：本题涉及 $\int \frac{1}{c^2 + x^2} dx = \frac{1}{c} \arctan(\frac{x}{c}) + C$ 。这是考试中涉及 PDF 性质时的常见积分形式。

- **换元积分法：**在计算期望时，分子出现了 x ，此时应联想到使用对数积分形式： $\int \frac{x}{c^2+x^2} dx = \frac{1}{2} \ln(c^2+x^2) + C$ 。

3. 易错点

- **有效区间判定：**在第 (b) 问中，题目给出的上限 $\pi/3$ 超过了密度函数的定义域上限 $\pi/4$ 。切记：在定义域之外 $f(x) = 0$ ，积分时只需计算到 $\pi/4$ 即可。
- **参数 c 的隐函数形式：**第 (a) 问解出的 c 满足 $\arctan(\frac{\pi}{4c}) = c$ 。在没有具体数值表的情况下，后续步骤通常保留 c 的表达式，不需要强行求出数值解。
- **期望计算漏掉 x ：**计算 $E(X)$ 时，积分表达式是 $x \cdot f(x)$ ，不要直接对 $f(x)$ 求积分（那样结果永远是 1）。

Question 4: (14 marks)

Suppose that the joint probability density of two random variables X and Y are given by

$$f(x, y) = \begin{cases} c \cdot y^3(2-x), & 0 \leq x \leq 1, 0 \leq y \leq x, \\ 0, & \text{elsewhere.} \end{cases}$$

Find:

- The value of constant c ;
- The marginal densities $f_X(x), f_Y(y)$;
- Whether the two random variables X and Y are independent or not? (Please give your reason);
- $E(XY)$

Solution:

The joint probability density function of (X, Y) is

$$f(x, y) = \begin{cases} c y^3(2-x), & 0 \leq x \leq 1, 0 \leq y \leq x, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determination of constant c

Since $f(x, y)$ is a joint pdf,

$$\int_0^1 \int_0^x c y^3(2-x) dy dx = 1.$$

Compute the inner integral:

$$\int_0^x y^3 dy = \frac{x^4}{4}.$$

Thus,

$$c \int_0^1 (2-x) \frac{x^4}{4} dx = 1.$$

$$\frac{c}{4} \int_0^1 (2x^4 - x^5) dx = \frac{c}{4} \left(\frac{2}{5} - \frac{1}{6} \right) = \frac{c}{4} \cdot \frac{7}{30}.$$

Hence,

$$\frac{7c}{120} = 1 \Rightarrow c = \frac{120}{7}$$

(b) Marginal densities

Marginal density of X :

$$f_X(x) = \int_0^x c y^3 (2-x) dy = c(2-x) \frac{x^4}{4}, \quad 0 \leq x \leq 1.$$

$$f_X(x) = \frac{30}{7} x^4 (2-x), \quad 0 \leq x \leq 1$$

Marginal density of Y :

For fixed y , x ranges from y to 1:

$$f_Y(y) = \int_y^1 c y^3 (2-x) dx.$$

$$= c y^3 \left[2x - \frac{x^2}{2} \right]_y^1 = c y^3 \left(\frac{3}{2} - 2y + \frac{y^2}{2} \right).$$

$$f_Y(y) = \frac{120}{7} y^3 \left(\frac{3}{2} - 2y + \frac{y^2}{2} \right), \quad 0 \leq y \leq 1$$

(c) Independence of X and Y

If X and Y were independent, then

$$f(x, y) = f_X(x) f_Y(y).$$

However, the support of (X, Y) satisfies $0 \leq y \leq x$, which is not a rectangular region. Therefore, X and Y are **not independent**.

X and Y are not independent

(d) Computation of $E(XY)$

$$\begin{aligned}
 E(XY) &= \int_0^1 \int_0^x xy c y^3 (2-x) dy dx. \\
 &= c \int_0^1 x(2-x) \int_0^x y^4 dy dx = c \int_0^1 x(2-x) \frac{x^5}{5} dx. \\
 &= \frac{c}{5} \int_0^1 (2x^6 - x^7) dx = \frac{c}{5} \left(\frac{2}{7} - \frac{1}{8} \right) = \frac{c}{5} \cdot \frac{9}{56}.
 \end{aligned}$$

Substitute $c = \frac{120}{7}$:

$$E(XY) = \frac{108}{245}$$

核心要点：二维随机变量的综合分析

1. 知识点定位：从“面”到“线”的转化

- **联合密度的归一性**：在二维平面区域上，联合概率密度函数 $f(x, y)$ 的二重积分必须等于 1。这是求常数 c 的基本依据。
- **边缘分布 (Marginal Density)**：求 X 的边缘分布时，“积掉” y ；求 Y 的边缘分布时，“积掉” x 。
- **独立性判定**：
 - **代数标准**：看是否满足 $f(x, y) = f_X(x) \cdot f_Y(y)$ 。
 - **几何直观**：如果变量的取值范围（支撑集）不是一个 ** 矩形区域 **（如本题中 y 的范围依赖于 x ），那么这两个变量通常不独立。

2. 关键难点：确定积分上下限

- **求 $f_X(x)$** ：固定 x ，观察 y 的变动范围。根据题目条件 $0 \leq y \leq x$ ，积分限即为 0 到 x 。
- **求 $f_Y(y)$** ：固定 y ，观察 x 的变动范围。从区域图中可以看出，对于确定的 y ， x 是从直线 $x = y$ 开始变到 $x = 1$ 为止。因此积分限是 y 到 1。这是本题最易出错的地方。

3. 计算建议与提醒

- **分步积分**：处理 $E(XY)$ 这种复杂的二重积分时，先对其中一个变量求内层积分，化简后再对第二个变量积分，避免算式过于臃肿。

- **结果自检：**边缘分布 $f_X(x)$ 和 $f_Y(y)$ 本身也是概率密度函数，它们在各自区间上的积分也必须等于 1。如果时间充裕，可以通过这个方法校验计算是否正确。

Question 5: (14 marks)

Suppose $X_1, X_2 \dots X_n$ is a sample from X , and the probability density function of X is

$$f(x; \theta) = \begin{cases} \theta^2 x e^{-\theta x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

where $\theta > 0$. Find:

- The moment estimator of θ ;
- The maximum likelihood estimator of θ .

Solution:

(a) Moment estimator of θ

Compute the expectation:

$$E(X) = \int_0^{\infty} x \theta^2 x e^{-\theta x} dx = \theta^2 \int_0^{\infty} x^2 e^{-\theta x} dx.$$

Using $\int_0^{\infty} x^2 e^{-\theta x} dx = \frac{2}{\theta^3}$,

$$E(X) = \frac{2}{\theta}.$$

Equating $E(X)$ with the sample mean \bar{X} ,

$$\bar{X} = \frac{2}{\theta} \Rightarrow \hat{\theta}_{\text{MM}} = \frac{2}{\bar{X}}$$

(b) Maximum likelihood estimator of θ

The likelihood function is

$$L(\theta) = \prod_{i=1}^n \theta^2 X_i e^{-\theta X_i} = \theta^{2n} \left(\prod X_i \right) e^{-\theta \sum X_i}.$$

The log-likelihood is

$$\ell(\theta) = 2n \ln \theta + \sum \ln X_i - \theta \sum X_i.$$

Differentiate and set equal to zero:

$$\frac{d\ell}{d\theta} = \frac{2n}{\theta} - \sum X_i = 0.$$

Thus,

$$\hat{\theta}_{\text{MLE}} = \frac{2n}{\sum X_i}$$

核心要点：矩估计与极大似然估计

1. 知识点定位：参数估计的两种方法

- **矩估计法 (Method of Moments)**: 核心思想是用“样本矩”去估计“总体矩”。通常第一步是计算总体均值 $E(X)$ (即一阶原点矩), 然后令 $E(X) = \bar{X}$, 解出未知参数。
- **极大似然估计法 (MLE)**: 核心思想是寻找一个参数值 θ , 使得当前观测到的样本序列出现的概率 (似然度) 达到最大。

2. 解题步骤与技巧

- **矩估计中的积分计算**: 计算 $E(X)$ 时, 常涉及形如 $\int_0^\infty x^n e^{-ax} dx$ 的积分, 这与 Γ 函数相关。记住结论 $\int_0^\infty x^n e^{-\theta x} dx = \frac{n!}{\theta^{n+1}}$ 可以大幅提高计算速度。
- **极大似然的标准三部曲**:
 1. 写出似然函数 $L(\theta)$ (所有样本密度的乘积)。
 2. 取对数得到 $\ln L(\theta)$ (将连乘变连加, 方便求导)。
 3. 求导并令导数为 0, 解方程得到 $\hat{\theta}$ 。

3. 易错点

- **指数项处理**: 在极大似然估计中, $\prod e^{-\theta X_i}$ 应该化简为 $e^{-\theta \sum X_i}$ 。对数化后, 它会变成 $-\theta \sum X_i$, 求导后只剩下 $-\sum X_i$ 。
- **参数位置**: 注意本题中参数 θ 在底数和指数位置同时出现。在对数似然函数中, $\ln(\theta^{2n})$ 应写成 $2n \ln \theta$ 。
- **合理性检查**: 在大多数简单分布中, 矩估计和极大似然估计的结果往往形式相似 (如本题中均与 $2/\bar{X}$ 相关), 如果两个结果差异巨大, 建议回头检查计算。

Question 6: (14 marks)

The capacities (in ampere-hours) of 10 batteries were recorded as follows:

102, 97, 101, 103, 101, 98, 99, 104, 103, 98

Under the assumption that the capacity is normal distributed as $N(\mu, \sigma^2)$.

Question: at level of significance 0.05,

(a) test $H_0 : \mu = 100 \leftrightarrow H_1 : \mu \neq 100$, (b) test $H_0 : \sigma^2 = 2.5 \leftrightarrow H_1 : \sigma^2 > 2.5$.

The t distribution table and the χ^2 distribution table

$t_9(0.025) = 2.2622$	$t_9(0.05) = 1.8331$	$t_{10}(0.025) = 2.2281$	$t_{10}(0.05) = 1.8125$
$\chi_9^2(0.025) = 19.023$	$\chi_9^2(0.05) = 16.919$	$\chi_9^2(0.975) = 2.700$	$\chi_9^2(0.95) = 3.325$
$\chi_{10}^2(0.025) = 20.483$	$\chi_{10}^2(0.05) = 18.307$	$\chi_{10}^2(0.975) = 3.247$	$\chi_{10}^2(0.95) = 3.940$

Solution:

The observed data are

102, 97, 101, 103, 101, 98, 99, 104, 103, 98.

Sample size $n = 10$.

$$\bar{X} = 100.6, \quad S^2 = 5.16.$$

(a) Test of $H_0 : \mu = 100$ vs $H_1 : \mu \neq 100$

Test statistic:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{100.6 - 100}{\sqrt{5.16}/\sqrt{10}} = 0.835.$$

Critical value:

$$t_{0.025}(9) = 2.2622.$$

Since $|t| < 2.2622$, we fail to reject H_0 .

Fail to reject H_0

(b) Test of $H_0 : \sigma^2 = 2.5$ vs $H_1 : \sigma^2 > 2.5$

Test statistic:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9 \times 5.16}{2.5} = 18.58.$$

Critical value:

$$\chi_{0.05}^2(9) = 16.919.$$

Since $18.58 > 16.919$, we reject H_0 .

Reject H_0

核心要点：单正态总体的假设检验

1. 知识点：推断统计的决策

- 均值检验 (σ^2 未知): 当总体方差未知时, 使用 t 检验。统计量为 $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ 。
- 方差检验: 检验总体波动情况时使用卡方检验。统计量为 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ 。

2. 关键步骤：双侧检验 vs 单侧检验

- 双侧检验 (如第 a 问): 备择假设 $H_1: \mu \neq 100$ 。拒绝域分布在两端, 查表时需要取 $\alpha/2 = 0.025$ 。
- 单侧检验 (如第 b 问): 备择假设 $H_1: \sigma^2 > 2.5$ 。拒绝域只在右侧 (大值端), 查表时直接取 $\alpha = 0.05$ 。

3. 易错点

- 数据预处理: 在给出原始数据 (如 10 个容量值) 时, 第一步必须准确算出样本均值 \bar{X} 和样本方差 S^2 。这是后续所有计算的基石。
- 自由度的选择: 样本量 $n = 10$, 对应的自由度 $df = n - 1 = 9$ 。查表时千万不要看成 10 那一行。
- 结论判定标准:
 - 如果 $|\text{观察值}| > \text{临界值}$, 则 ** 拒绝 ** 原假设 (落在拒绝域)。
 - 如果 $|\text{观察值}| \leq \text{临界值}$, 则 ** 接受 (不拒绝) ** 原假设 (落在接受域)。

BDIC2005J Probability and Statistics 2018-2019

Question 1: Vacancy (Each blank 3 marks)

(1) There are two events A and B , $P(A) = 0.6$, $P(B) = 0.3$. If A and B are mutually independent, $P(A \cup B) =$, $P(A - B) =$.

Solution:

Given $P(A) = 0.6$, $P(B) = 0.3$, and A, B are mutually independent.

1. To find $P(A \cup B)$: Since A and B are independent, we have:

$$P(AB) = P(A)P(B) = 0.6 \times 0.3 = 0.18.$$

Using the addition rule:

$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.6 + 0.3 - 0.18 = 0.72.$$

2. To find $P(A - B)$: By the definition of the difference of events:

$$P(A - B) = P(A) - P(AB) = 0.6 - 0.18 = 0.42.$$

(Alternatively, since A and \bar{B} are also independent: $P(A - B) = P(A)P(\bar{B}) = 0.6 \times (1 - 0.3) = 0.42$.)

核心要点：独立事件的概率计算

1. 知识点：概率基本性质与独立性

- **独立性的定义：**若事件 A, B 相互独立，则其同时发生的概率满足乘法公式 $P(AB) = P(A)P(B)$ 。这是解决本题的关键突破口。
- **和事件公式：**对于任意两事件， $P(A \cup B) = P(A) + P(B) - P(AB)$ 。

2. 关键步骤：差事件的理解

- **语义转化：** $A - B$ 表示“事件 A 发生且事件 B 不发生”，即 $A \cap \bar{B}$ 。
- **独立性的性质延伸：**如果 A 与 B 相互独立，则 A 与 \bar{B} 、 \bar{A} 与 B 、 \bar{A} 与 \bar{B} 均相互独立。

3. 易错点

- **混淆概念：**务必区分“相互独立”与“互不相容（互斥）”。

- 若 ** 相互独立 **, 则 $P(AB) = P(A)P(B) = 0.18$ 。
- 若 ** 互不相容 **, 则 $P(AB) = 0$ 。
- **计算精度:** 在处理小数乘法时需细心, 避免将 0.6×0.3 误算为 1.8 或 0.018。

(2) $X \sim P(\lambda), P(X = 0) = 1/2, \lambda = \underline{\hspace{2cm}}$.

Solution:

Given X follows a Poisson distribution with parameter λ , denoted as $X \sim P(\lambda)$. The probability mass function (PMF) of X is:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

Given $P(X = 0) = 1/2$, substitute $k = 0$ into the formula:

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} = \frac{1}{2}.$$

Taking the natural logarithm on both sides:

$$-\lambda = \ln\left(\frac{1}{2}\right) = -\ln 2.$$

Thus:

$$\lambda = \ln 2 \approx 0.693.$$

核心要点：泊松分布的性质

1. 知识点定位：离散型随机变量

- **泊松分布定义:** 泊松分布 $P(\lambda)$ 常用于描述单位时间内随机事件发生的次数。参数 $\lambda > 0$ 既是该分布的均值也是方差。
- **分布律公式:** 掌握 $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ 是解题的基础。

2. 关键步骤：特殊值的代入

- **$k = 0$ 的特殊性:** 在泊松分布中, $P(X = 0)$ 总是等于 $e^{-\lambda}$ 。这是因为 $\lambda^0 = 1$ 且 $0! = 1$ 。
- **对数运算:** 方程 $e^{-\lambda} = a$ 的解为 $\lambda = -\ln a$ 。本题中 $a = 1/2$, 利用对数性质 $\ln(1/2) = -\ln 2$ 简化结果。

3. 易错点排查与提醒

- **阶乘计算：**容易误记 $0! = 0$ ，实际上 $0! = 1$ 。
- **参数范围：**泊松分布的参数 λ 必须大于 0。如果计算出负值，需检查对数运算符号。
- **结果形式：**通常填空题保留精确解 $\ln 2$ 即可，除非题目要求保留小数。

(3) Let $X_1, X_2, \dots, X_n (n > 2)$ be a sample from $N(\mu, \sigma^2)$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

$$\sqrt{n}(\bar{X} - \mu)/\sigma \sim \underline{\hspace{2cm}}, \quad \sqrt{n}(\bar{X} - \mu)/\sqrt{S^2} \sim \underline{\hspace{2cm}},$$

$$(n-1)S^2/\sigma^2 \sim \underline{\hspace{2cm}}.$$

Solution:

Based on the sampling distribution theory for a normal population $N(\mu, \sigma^2)$:

1. **Distribution of \bar{X} :** Since $X_i \sim N(\mu, \sigma^2)$, the sample mean $\bar{X} \sim N(\mu, \sigma^2/n)$. Standardizing \bar{X} gives:

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

2. **Distribution of S^2 :** The ratio of the sample variance to the population variance follows a Chi-square distribution:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

3. **Student's t-distribution:** By definition, $t = \frac{Z}{\sqrt{Y/k}}$, where $Z \sim N(0, 1)$ and $Y \sim \chi^2(k)$ are independent. Substituting $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ and $Y = \frac{(n-1)S^2}{\sigma^2}$ with $k = n-1$:

$$\frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1).$$

核心要点：正态总体的抽样分布

1. 知识点定位：三大抽样分布

- **标准正态分布** $N(0, 1)$: 当总体方差 σ^2 已知时, 对样本均值进行标准化得到的结果。
- **卡方分布** $\chi^2(n-1)$: 用于描述样本方差 S^2 与总体方差 σ^2 之间的关系。
- **t 分布** $t(n-1)$: 当总体方差 σ^2 未知, 用样本标准差 S 代替 σ 时, 均值的标准化统计量服从 t 分布。

2. 关键点：自由度 (Degrees of Freedom)

- 注意在 S^2 相关的分布中, 由于计算方差时消耗了一个自由度 (用来估计 \bar{X}), 所以 χ^2 和 t 分布的自由度均为 $n-1$, 而非 n 。

3. 易错点排查与提醒

- **分清已知与未知**: 看到分母是 σ 时, 选正态分布; 看到分母是 S (或 $\sqrt{S^2}$) 时, 选 t 分布。
- **符号辨析**: \bar{X} 是均值, S^2 是无偏样本方差。公式 $\frac{(n-1)S^2}{\sigma^2}$ 也可以写成 $\sum \frac{(X_i - \bar{X})^2}{\sigma^2}$ 。
- **独立性前提**: 在正态总体中, \bar{X} 与 S^2 是相互独立的, 这是构造 t 统计量的理论基础。

(4) Let X_1, \dots, X_n be a sample from $N(\mu, \sigma^2)$. μ and σ^2 are unknown parameters.

The confidence interval of μ with confidence level 95% is $[\cdot, \cdot]$.

The confidence interval of σ^2 with confidence level 95% is $[\cdot, \cdot]$.

Solution:

For a normal population with both μ and σ^2 unknown:

1. **Confidence Interval for μ** : Using the pivot $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, the $1-\alpha$ confidence interval is:

$$\bar{X} \pm t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}$$

For a 95% level, $\alpha = 0.05$, the interval is $\left[\bar{X} - t_{0.025}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{0.025}(n-1) \frac{S}{\sqrt{n}} \right]$.

2. **Confidence Interval for σ^2 :** Using the pivot $\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, we have:

$$P\left(\chi^2_{1-\alpha/2}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2}(n-1)\right) = 1 - \alpha$$

Rearranging for σ^2 gives the interval:

$$\left[\frac{(n-1)S^2}{\chi^2_{0.025}(n-1)}, \frac{(n-1)S^2}{\chi^2_{0.975}(n-1)} \right]$$

核心要点：单正态总体的区间估计

1. 知识点：参数估计

- **置信区间定义：**在 $1 - \alpha$ 的置信水平下，由样本确定的包含待估参数的随机区间。
- **枢轴量法：**构造一个包含待估参数、但不包含其他未知参数，且分布已知的统计量（枢轴量）。

2. 关键公式：选择正确的分布

- **均值 μ ：**当 σ^2 未知时，必须使用 t 分布。区间具有对称性，形式为 $\bar{X} \pm$ 误差限。
- **方差 σ^2 ：**使用 χ^2 分布。由于卡方分布是非对称的，置信区间的上下界由两个不同的临界值决定。

3. 易错点

- **卡方区间的上下界：**注意在计算 σ^2 区间时，分母越大数值越小。因此，左端点（下界）对应的分母是较大的临界值 $\chi^2_{\alpha/2}$ ，右端点（上界）对应的分母是较小的临界值 $\chi^2_{1-\alpha/2}$ 。
- **分位数下标：**不同的教科书对 t_α 和 χ^2_α 的定义（左尾或右尾）可能不同，考试时需参考给出的附表说明。
- **自由度：**所有涉及 S 或 S^2 的估计，自由度一律为 $n - 1$ 。

Question 2 (Version 2): (14 marks)

In our warehouse, there are 10 boxes of products provided by three factories. Exactly 2 boxes are provided by factory A; 3 boxes are provided by factory B; the others are provided by factory C. And we also know that the qualification rates for the three factories are 95%, 90% and 96% separately.

Find:

- (1) The qualification rate of the products in our warehouse;
- (2) Now we randomly take a product from a randomly chosen box. And we can see the selected product is qualified. What is the probability that it is provided by factory A? What is the probability that it is provided by factory B? What is the probability that it is provided by factory C?

Solution:

Let H_1, H_2, H_3 be the events that the product is provided by factory A, B, and C, respectively. Let D be the event that a selected product is qualified.

From the problem description:

$$P(H_1) = \frac{2}{10} = 0.2, \quad P(H_2) = \frac{3}{10} = 0.3, \quad P(H_3) = \frac{5}{10} = 0.5.$$

The conditional probabilities (qualification rates) are:

$$P(D|H_1) = 0.95, \quad P(D|H_2) = 0.90, \quad P(D|H_3) = 0.96.$$

(1) Total qualification rate $P(D)$: By the Law of Total Probability:

$$\begin{aligned} P(D) &= P(H_1)P(D|H_1) + P(H_2)P(D|H_2) + P(H_3)P(D|H_3) \\ &= 0.2 \times 0.95 + 0.3 \times 0.90 + 0.5 \times 0.96 \\ &= 0.19 + 0.27 + 0.48 = 0.94. \end{aligned}$$

The qualification rate in the warehouse is 0.94.

(2) Posterior probabilities using Bayes' Theorem: The probability that the product came from factory A given it is qualified:

$$P(H_1|D) = \frac{P(H_1)P(D|H_1)}{P(D)} = \frac{0.19}{0.94} = \frac{19}{94} \approx 0.2021.$$

The probability that it came from factory B:

$$P(H_2|D) = \frac{P(H_2)P(D|H_2)}{P(D)} = \frac{0.27}{0.94} = \frac{27}{94} \approx 0.2872.$$

The probability that it came from factory C:

$$P(H_3|D) = \frac{P(H_3)P(D|H_3)}{P(D)} = \frac{0.48}{0.94} = \frac{48}{94} = \frac{24}{47} \approx 0.5106.$$

核心要点：全概率公式与贝叶斯公式

1. 知识点：原因与结果的推断

- **全概率公式 (Total Probability):** 已知各“原因”(厂家)发生的概率及各原因导致“结果”(合格)的概率, 求结果发生的总概率。这是“由因导果”。
- **贝叶斯公式 (Bayes' Theorem):** 在已知结果发生的条件下, 反推是由某个特定原因导致的概率。这是“由果索因”。

2. 关键步骤

- **划分完备事件组:** 首先确定 H_1, H_2, H_3 构成了样本空间的一个划分 (概率之和为 1), 这是使用上述公式的前提。
- **分母的意义:** 在计算第 (2) 问的贝叶斯概率时, 分母统一使用第 (1) 问算出的全概率 $P(D)$ 。

3. 易错点

- **数量转化:** 题目给出的是“箱数”, 需转化为选择各厂家的先验概率 $P(H_i)$ 。例如 C 厂虽然没直接说箱数, 但通过 $10 - 2 - 3 = 5$ 算出。
- **计算验证:** 在第 (2) 问中, 所有后验概率之和 $P(H_1|D) + P(H_2|D) + P(H_3|D)$ 必须等于 1。这可以作为检查计算准确性的有效手段。
- **有效数字:** 概率计算通常保留四位小数或给出精确分数, 需根据考试的具体要求而定。

Question 3 (Version 2): (14 marks)

Suppose, for a town, the electricity consumption per day X is a continuous random variable. And the density function of X is

$$f(x) = \begin{cases} cx(1-x)^2, & 0 < x < 1 \\ 0, & \text{else} \end{cases}$$

Find:

- (1) The value of constant c ;
- (2) $P(0.5 < X < 1.3)$;
- (3) $E(X)$.

Solution:

(1) Find the constant c : Using the normalization property of the probability density

function (PDF):

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \implies \int_0^1 cx(1-x)^2dx = 1$$

Expanding the integrand: $x(1-x)^2 = x(1-2x+x^2) = x-2x^2+x^3$.

$$c \int_0^1 (x-2x^2+x^3)dx = c \left[\frac{1}{2}x^2 - \frac{2}{3}x^3 + \frac{1}{4}x^4 \right]_0^1 = c \left(\frac{1}{2} - \frac{2}{3} + \frac{1}{4} \right) = c \left(\frac{6-8+3}{12} \right) = \frac{c}{12}.$$

Setting $\frac{c}{12} = 1$, we get $c = 12$.

(2) Find $P(0.5 < X < 1.3)$: Since $f(x) = 0$ for $x \geq 1$, we only integrate up to 1:

$$P(0.5 < X < 1.3) = \int_{0.5}^1 12(x-2x^2+x^3)dx = 12 \left[\frac{1}{2}x^2 - \frac{2}{3}x^3 + \frac{1}{4}x^4 \right]_{0.5}^1$$

Substituting the limits:

$$\begin{aligned} &= 12 \left(\frac{1}{12} - \left(\frac{1}{2}(0.5)^2 - \frac{2}{3}(0.5)^3 + \frac{1}{4}(0.5)^4 \right) \right) \\ &= 1 - 12 \left(\frac{1}{8} - \frac{1}{12} + \frac{1}{64} \right) = 1 - 12 \left(\frac{24-16+3}{192} \right) = 1 - 12 \left(\frac{11}{192} \right) = 1 - \frac{11}{16} = \frac{5}{16} = 0.3125. \end{aligned}$$

(3) Find $E(X)$: By definition of mathematical expectation:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x \cdot 12x(1-x)^2dx = 12 \int_0^1 (x^2 - 2x^3 + x^4)dx \\ E(X) &= 12 \left[\frac{1}{3}x^3 - \frac{2}{4}x^4 + \frac{1}{5}x^5 \right]_0^1 = 12 \left(\frac{1}{3} - \frac{1}{2} + \frac{1}{5} \right) = 12 \left(\frac{10-15+6}{30} \right) = 12 \left(\frac{1}{30} \right) = \frac{2}{5} = 0.4. \end{aligned}$$

核心要点：连续型随机变量及其数字特征

1. 知识点定位：密度函数的性质

- **归一性**：所有可能的取值概率之和为 1，即 $\int f(x)dx = 1$ 。这是求待定常数 c 的标准方法。
- **概率计算**：落在某个区间内的概率等于密度函数在该区间上的定积分。

2. 关键步骤：积分上下限的确定

- **分段积分**：在计算概率（如第 2 问）时，必须注意变量的有效定义域。虽然题目要求 $0.5 < X < 1.3$ ，但由于 $x > 1$ 时 $f(x) = 0$ ，实际积分只需从 0.5 积到 1。

- **多项式展开:** 对于 $(1-x)^n$ 形式的密度函数, 先展开再逐项积分通常比分部积分法更快捷。

3. 易错点提醒

- **计算陷阱:** 多项式积分容易在通分 (如 $\frac{1}{2} - \frac{2}{3} + \frac{1}{4}$) 时出错, 建议每一步都进行复算。
- **期望公式:** 注意 $E(X)$ 是 $\int xf(x)dx$, 很多同学会漏掉公式里的那个额外的 x , 导致算成了归一化积分。
- **贝塔函数关系:** 本题实质上与 Beta 分布有关, 即 $X \sim \text{Beta}(2, 3)$ 。熟练的学生可以直接利用性质检验答案: $c = \frac{1}{B(2,3)} = \frac{\Gamma(5)}{\Gamma(2)\Gamma(3)} = \frac{24}{1 \times 2} = 12$ 。

Question 4: (14 marks)

Suppose that the joint probability density of two random variables X and Y are given by

$$f(x, y) = \begin{cases} c, & x^2 \leq y \leq x \\ 0, & \text{else} \end{cases}$$

Find:

- The value of constant c ;
- The marginal densities $f_X(x)$, $f_Y(y)$;
- Whether the two random variables X and Y are independent or not? (Please give your reason).
- $E(XY)$

Solution:

First, determine the region of integration. The inequalities $x^2 \leq y \leq x$ imply that $x^2 \leq x$, which means $0 \leq x \leq 1$. The region D is bounded by the parabola $y = x^2$ and the line $y = x$ for $x \in [0, 1]$.

(a) **Value of constant c :** Using the normalization property $\iint_D f(x, y) dx dy = 1$:

$$\int_0^1 \int_{x^2}^x c dy dx = c \int_0^1 (x - x^2) dx = c \left[\frac{1}{2}x^2 - \frac{1}{3}x^3 \right]_0^1 = c \left(\frac{1}{6} \right) = 1$$

Thus, $c = 6$.

(b) **Marginal densities:** For $f_X(x)$, integrate with respect to y :

$$f_X(x) = \int_{x^2}^x 6 dy = 6(x - x^2), \quad 0 < x < 1$$

(Otherwise $f_X(x) = 0$).

For $f_Y(y)$, integrate with respect to x . Since $x^2 \leq y \leq x \implies y \leq x \leq \sqrt{y}$ for $y \in [0, 1]$:

$$f_Y(y) = \int_y^{\sqrt{y}} 6 dx = 6(\sqrt{y} - y), \quad 0 < y < 1$$

(Otherwise $f_Y(y) = 0$).

(c) Independence: Two variables are independent if $f(x, y) = f_X(x)f_Y(y)$ for all x, y . Checking the product: $6(x - x^2) \cdot 6(\sqrt{y} - y) \neq 6$ in the region D . Also, the domain D is not a rectangle, which is a sufficient condition for dependence. **Result:** X and Y are not independent.

(d) Mathematical Expectation $E(XY)$:

$$\begin{aligned} E(XY) &= \iint_D xyf(x, y)dxdy = \int_0^1 \int_{x^2}^x 6xy dydx \\ &= 6 \int_0^1 x \left[\frac{1}{2}y^2 \right]_{x^2}^x dx = 3 \int_0^1 x(x^2 - x^4)dx = 3 \int_0^1 (x^3 - x^5)dx \\ &= 3 \left[\frac{1}{4}x^4 - \frac{1}{6}x^6 \right]_0^1 = 3 \left(\frac{1}{4} - \frac{1}{6} \right) = 3 \left(\frac{3-2}{12} \right) = \frac{3}{12} = \frac{1}{4} = 0.25. \end{aligned}$$

核心要点：二维连续型随机变量

1. 知识点定位：联合分布与边缘分布

- **归一化系数：**通过二重积分 $\iint f(x, y)dA = 1$ 确定常数 c 。
- **边缘密度：**计算 $f_X(x)$ 时，积分限是针对 y 的函数；反之亦然。务必根据图形确定正确的积分上下限。

2. 关键判定：独立性的快速检查

- **几何法：**如果联合密度的非零定义域区域不是一个边平行于坐标轴的矩形（本题是曲边形），则 X 和 Y 一定不独立。
- **代数法：**验证 $f(x, y) = f_X(x)f_Y(y)$ 是否恒成立。

3. 易错点

- **积分限反转：**在计算 $f_Y(y)$ 时，需要从 y 的函数反解出 x 的范围（即 $x^2 \leq y \implies x \leq \sqrt{y}$ 且 $y \leq x$ ）。

- **定义域书写：**在给出边缘分布结果时，必须附带其成立的取值范围（如 $0 < x < 1$ ），否则结果是不完整的。
- **二重积分计算：** $E(XY)$ 的计算通常有两条路径，选择先积 y 还是先积 x 应以计算简便为准，本题先积 y 较为直观。

Question 5: (14 marks)

Suppose X_1, X_2, \dots, X_n is a sample from X , and the probability density function of X is

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x \geq 0 \\ 0, & \text{else} \end{cases}$$

Find:

- (1) The moment estimator $\hat{\theta}$ of θ , is it unbiased?;
- (2) The maximum likelihood estimator θ^* of θ .

Solution:

(1) Moment Estimator and Unbiasedness: First, calculate the first population moment (mean):

$$E(X) = \int_0^{\infty} x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$$

Let $t = x/\theta$, then $dx = \theta dt$:

$$E(X) = \theta \int_0^{\infty} t e^{-t} dt = \theta \cdot \Gamma(2) = \theta \cdot 1! = \theta.$$

Set the population mean equal to the sample mean \bar{X} :

$$\mu_1 = \bar{X} \implies \theta = \bar{X}.$$

Thus, the moment estimator is $\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

To check for unbiasedness:

$$E(\hat{\theta}) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n}(n\theta) = \theta.$$

Since $E(\hat{\theta}) = \theta$, the estimator is **unbiased**.

(2) Maximum Likelihood Estimator (MLE): The likelihood function for the sample X_1, \dots, X_n is:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \theta^{-n} \exp\left(-\frac{1}{\theta} \sum_{i=1}^n x_i\right).$$

Take the natural logarithm of the likelihood function:

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

Differentiate with respect to θ and set to zero:

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i = 0.$$

Multiplying by θ^2 gives:

$$-n\theta + \sum_{i=1}^n x_i = 0 \implies \theta = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Thus, the maximum likelihood estimator is $\theta^* = \bar{X}$.

核心要点：参数估计（矩估计与极大似然）

1. 知识点：点估计方法

- **矩估计法 (Method of Moments):** 利用“样本矩收敛于总体矩”的思想，通常令 $E(X) = \bar{X}$ 。
- **极大似然估计 (MLE):** 寻找使已发生的样本观测值出现概率最大的参数值。
- **无偏性 (Unbiasedness):** 如果估计量的数学期望等于参数真值，即 $E(\hat{\theta}) = \theta$ ，则称其为无偏估计。

2. 关键步骤：似然方程的求解

- **对数转换:** 似然函数通常是连乘形式，通过取对数变为求和形式，可以极大简化求导过程。
- **指数分布特征:** 本题中的 $f(x)$ 是参数为 $1/\theta$ 的指数分布，其均值为 θ 。记住常用分布的期望值能显著提高矩估计的做题速度。

3. 易错点排查与提醒

- **求导链式法则:** 在对 $\frac{1}{\theta} \sum x_i$ 关于 θ 求导时，注意负号。 $\frac{d}{d\theta}(\theta^{-1}) = -\theta^{-2}$ 。
- **符号区分:** 矩估计结果通常加尖角号 $\hat{\theta}$ ，极大似然结果有时加星号 θ^* ，并在结论中使用大写字母 \bar{X} 表示估计量。
- **无偏性证明:** 证明无偏性时，只需利用期望的线性性质 $E(\sum X_i) = \sum E(X_i)$ 即可，无需重新进行复杂的积分。

Question 6 (Version 2): (14 marks)

Suppose the nicotine content of a brand of cigarette follows a normal distribution. Now we randomly take 10 cigarettes, the average nicotine content $\bar{X} = 18.6$ milligram, the standard deviation $S = 2.4$ milligram. Take the significant level $\alpha = 0.05$.

- (a) Can we accept $\mu = 18$?
 (b) Can we accept $\sigma = 2$?

The t distribution table and the χ^2 distribution table

$t_9(0.025) = 2.2622$	$t_9(0.05) = 1.8331$	$t_{10}(0.025) = 2.2281$	$t_{10}(0.05) = 1.8125$
$\chi_9^2(0.025) = 19.023$	$\chi_9^2(0.05) = 16.919$	$\chi_9^2(0.975) = 2.700$	$\chi_9^2(0.95) = 3.325$
$\chi_{10}^2(0.025) = 20.483$	$\chi_{10}^2(0.05) = 18.307$	$\chi_{10}^2(0.975) = 3.247$	$\chi_{10}^2(0.95) = 3.940$

Solution:

Given $n = 10$, $\bar{X} = 18.6$, $S = 2.4$, and $\alpha = 0.05$. The degrees of freedom $df = n - 1 = 9$.

(a) Test $H_0 : \mu = 18$ vs $H_1 : \mu \neq 18$

Since the population variance σ^2 is unknown, we use the t -test statistic:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{18.6 - 18}{2.4/\sqrt{10}} = \frac{0.6}{0.7589} \approx 0.7906.$$

Critical value: For a two-tailed test with $\alpha = 0.05$ and $df = 9$, we check $t_{0.025}(9) = 2.2622$.

Since $|t| = 0.7906 < 2.2622$, the test statistic falls in the acceptance region. **Conclusion:** We **accept** H_0 . There is no significant evidence that the mean differs from 18.

(b) Test $H_0 : \sigma = 2$ vs $H_1 : \sigma \neq 2$ (or equivalently $\sigma^2 = 4$ vs $\sigma^2 \neq 4$)

Using the Chi-square test statistic for variance:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9 \times (2.4)^2}{2^2} = \frac{9 \times 5.76}{4} = \frac{51.84}{4} = 12.96.$$

Critical values: For a two-tailed test with $\alpha = 0.05$ and $df = 9$, the rejection region is outside $[\chi_{0.975}^2(9), \chi_{0.025}^2(9)]$. From the table: $\chi_{0.975}^2(9) = 2.700$ and $\chi_{0.025}^2(9) = 19.023$.

Since $2.700 < 12.96 < 19.023$, the test statistic falls in the acceptance region. **Conclusion:** We **accept** H_0 . There is no significant evidence that the standard deviation differs from 2.

核心要点：单正态总体的假设检验

1. 知识点定位：双侧检验的决策

- 均值检验 (σ^2 未知): 当总体方差未知时, 使用 $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ 。
- 方差检验: 检验总体波动 σ^2 时, 使用 $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ 。注意题目给的是 S 和 σ , 计算时需平方。

2. 关键步骤：查表与自由度

- 自由度确认: 样本量 $n = 10$, 所有查表操作必须基于 $df = 9$ 。
- 双侧临界值: 对于“能否接受”类问题, 默认进行双侧检验。此时显著性水平 α 需分配到两端 (各 $\alpha/2 = 0.025$)。

3. 易错点排查与提醒

- 标准差与方差的陷阱: 在卡方检验中, 公式使用的是 S^2 和 σ_0^2 。本题给出的是 2.4 和 2, 代入公式前务必进行平方运算。
- 卡方分布的不对称性: 与对称的 t 分布不同, 卡方检验的两端临界值 $\chi_{0.975}^2$ 和 $\chi_{0.025}^2$ 需要分别查表获取, 不能简单加负号。
- 结论表述: 若统计量在区间内, 结论为“接受 H_0 ”或“不能拒绝 H_0 ”; 若在区间外, 则“拒绝 H_0 ”。