Exam2023

q1

Let $X_1, X_2, \ldots, X_n$ denote a random sample from the following normal distribution: $N(\mu, 9)$ for $n \in \mathbb{N}$, and let $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ denote the associated sample mean. How large must n be so that $\bar{X}$ is no more than 0.7 from μ with 95% confidence? State your answer as an integer between 0 and 99.

Documentation: You are expected to demonstrate how you obtained the results either by supplying manual calculations or Python code.

The sample size must be at least 71 in order for the sample mean to be no more than 0.7 from μ with 95% confidence.

q2

Let X be a discrete stochastic variable with the following probability mass function:

$$p_X(x) = \begin{cases} 1/3 & \text{for } x \in \{-1, 0, 1\} \\ 0 & \text{else.} \end{cases}$$

For all questions in this assignment, state your inputs as integers between 0 and 99 such that all answers are given as either integers or irreducible fractions.

Documentation: You are expected to demonstrate how you obtained the results either by supplying manual calculations or Python code.

a. Find the expected value and variance of X.

$E(X) = \square$

Ch

$Var(X) = \frac{\square}{\square}$

Let Y denote a stochastic variable that is independent of X and has the same PMF as X, i.e.

$$p_Y(y) = \begin{cases} 1/3 & \text{for } y \in \{-1, 0, 1\} \\ 0 & \text{else.} \end{cases}$$

b. Find the values below.

$Var(X - Y) = \frac{4}{3}$

$Cov(X, Y + X) = \frac{\square}{\square}$

c. Find the values below.

$E(|X - 1|) = \square$

$E(|X - 1| \cdot |Y - 1|) = \square$

d. Find the probabilities below.

$$P(X \neq 0) = \frac{\square}{\square}$$

$$P(XY = 0) = \frac{5}{9}$$

q3

Let X and Y be two independent stochastic variables such that $X \sim \text{Binomial}(10, 0.5)$ and $Y \sim \text{Geometric}(1/5)$. Find the below probabilities. State your answer as an integer between 0 and 99 such that you supply four decimal precision.

Documentation: You are expected to demonstrate how you obtained the results either by supplying manual calculations or Python code.

$P(X > 5 \cup Y \leq 3)$ =0.6 81 ✔ 0

$P(X > 5 \mid X + Y = 3)$ =0.000 0 ✔

q4

Let X and Y be two jointly continuous random variables with joint PDF:

$$f_{XY}(x, y) = \begin{cases} \frac{x^2}{4} + \frac{y^2}{4} + \frac{xy}{6} & 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find $E\left[X^2 \mid Y = 1\right]$. State your input as a positive integer such that the answer is stated as an irreducible fraction.

Documentation: You are expected to demonstrate how you obtained the results either by supplying manual calculations or Python code.

$$\frac{21}{50}$$

q5

One way to design a spam filter is to look at the words in an email. In particular, some words are more frequent in spam emails. Suppose that we have the following information:

- 60% of emails are spam;
- 2% of spam emails contain the word "discount";
- 0.002% of non-spam emails contain the word "discount".

Suppose that an email is checked and found to contain the word "discount". Find the probabilities below. State all inputs as integers between 10 and 99 such that the answer is given as a decimal value with four decimal precision.

Documentation: You should supply the calculations leading up to the final result.

The probability that an email contains the word "discount": 0.0 12 ✔ 0

The probability that an email is spam given that it contains the word "discount": 0.9 99 ✔ 3 q6

The number of cars, X, passing the Storebælt Bridge over a certain time period, can reasonably be assumed to follow a Poisson distribution:

$$X \sim \text{poisson}(\lambda t)$$

Over a time period of t = 3 hours, x = 1140 cars pass the bridge. Based on this information, the estimate for the time parameter is:

$$\hat{\lambda} = \frac{x}{t} = \frac{1140}{3} = 380 \text{ cars per hour}$$

Documentation: In (a) no documentation is required. In (b) and (c) you are expected to demonstrate how you obtained the results either by supplying manual calculations or Python code.

a. An employee at the Storebælt Bridge claims that, on an annual basis and for a similar time period, an average of 400 cars pass per hour. Do the measured data (x = 1140 cars in t = 3 hours) agree with the employee's claim? To answer this question correctly, you need to conduct a hypothesis test. Identify the correct null hypothesis.

- [ ] $H_0 : \lambda_0 \neq 380$
- [ ] $H_0 : \lambda_0 \neq 1200$
- [ ] $H_0 : \lambda_0 \geq 1140$
- [ ] $H_0 : \lambda_0 \neq 400$
- [ ] $H_0 : \lambda_0 = 1140$
- [x] $H_0 : \lambda_0 = 400$
- [ ] $H_0 : \lambda_0 = 380$
- [ ] $H_0 : \lambda_0 \geq 380$
- [ ] $H_0 : \lambda_0 \neq 1140$
- [ ] $H_0 : \lambda_0 = 1200$

b. Determine the test statistic and the p-value for the test. State your inputs as integers between 10 and 99 so that the answer is displayed with four decimal precision.

Hint: Remember that the variance of a Poisson random variable is the same as the parameter.

Test statistic: -1.7 32 1

P-value: 0.0 83 3

Check answers

c. Assuming $\lambda = 400$ cars per hour, as stated by the employee, determine the probability that 380 or fewer cars pass over the bridge in one hour.

$0.1\boxed{64}9$

q7

Let P and Q be two independent $N(0,1)$ random variables and

$$R = 7 - P + P^2$$
$$S = 4 - PQ.$$

Find $\text{Cov}(R, S)$. State your answer as an integer between 0 and 99.

Documentation: You must supply your calculations leading up to the result.

$$Cov(R,S) = \boxed{0}$$

Problem

Let $X1, X2, \ldots, Xn$ denote a random sample from the following normal distribution: $N(\mu, 9)$ for $n \in N$, and let $X^{-} = 1n(X1 + X2 + \cdots + Xn)$ denote the associated sample mean. How large must n be so that X^{-} is no more than 0.7 from μ with 95% confidence? State your answer as an integer between 0 and 99.

ANSWER: The sample size must be at least … in order for the sample mean to be no more than 0.7 from μ with 95% confidence.

We are given that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution $N(\mu, 9)$.

We are asked to determine the **minimum sample size** n such that the sample mean $\bar{X}$ is **within 0.7 units** of the population mean μ with **95% confidence**.

Step-by-Step Solution

Step 1: Understand the Goal

We are solving a **confidence interval** problem. Specifically, we want the probability that the sample mean is within 0.7 of the true mean to be at least 95%. Mathematically, this is:

$$P(|\bar{X} - \mu| < 0.7) = 0.95$$

This condition is met by finding the required **sample size** n for a given **margin of error** $E = 0.7$, **confidence level** 95%, and **known population standard deviation**.

Step 2: Use the Margin of Error Formula

For a normally distributed population with known standard deviation:

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Where:

- $E = 0.7$ (margin of error)
- $Z_{\alpha/2} = 1.96$ for 95% confidence
- $\sigma = \sqrt{9} = 3$ (given population variance is 9)

Plug in the values:

$$0.7 = 1.96 \cdot \frac{3}{\sqrt{n}}$$

Solve for n:

$$\frac{3}{\sqrt{n}} = \frac{0.7}{1.96} \Rightarrow \sqrt{n} = \frac{3 \cdot 1.96}{0.7} = 8.4$$

$$n = 8.4^2 = 70.56 \Rightarrow \boxed{n = 71} \text{ (round up to ensure margin is met)}$$

Step 3: Python Verification

We can use Python to verify this calculation:

```
import math

from scipy.stats import norm



# Known values

sigma = 3  # Standard deviation
```

```
E = 0.7  # Margin of error

confidence = 0.95  # Confidence level

z = norm.ppf(1 - (1 - confidence)/2) # z-value for 95%



# Solve for n

n = (z * sigma / E)**2

print("n (rounded up):", math.ceil(n))
```

Output:

```
n (rounded up): 71
```

Final Answer

The smallest sample size required is:

$$\boxed{n = 71}$$

General Tip for Similar Problems

To solve **sample size for confidence intervals**, remember:

1. Use the margin of error formula:

$$E = Z \cdot \frac{\sigma}{\sqrt{n}}$$

2. Rearrange to solve for n:

$$n = \left(\frac{Z \cdot \sigma}{E} \right)^2$$

3. Use the appropriate Z-value for your desired confidence level:

- 90%: $Z = 1.645$
- 95%: $Z = 1.96$
- 99%: $Z = 2.576$

4. Always **round up** your answer when calculating n to guarantee precision.

This applies anytime you're estimating a **mean** with known variance and need a certain **precision** (margin of error).

Discrete Stochastic Variable Analysis — Full Walkthrough with Explanations

We are given a discrete random variable (X) with the following probability mass function (PMF):

$$p_X(x) = \begin{cases} \frac{1}{3} & \text{if } x \in \{-1, 0, 1\} \\ 0 & \text{otherwise} \end{cases}$$

Let (Y) be a stochastic variable **independent** of (X) and with the same PMF.

Step 1: Compute the Expected Value $\mathbb{E}[X]$

Formula:

$$\mathbb{E}[X] = \sum_x x \cdot P(X = x)$$

Substitution:

$$\begin{aligned} \mathbb{E}[X] &= (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} \\ &= \frac{-1 + 0 + 1}{3} = \frac{0}{3} = 0 \end{aligned}$$

> 📘 **Explanation**: The expected value is a weighted average. Since -1 and 1 cancel out and 0 has zero weight, the average is 0.

WolframAlpha Input:

```
expectation {-1:1/3, 0:1/3, 1:1/3}
```

Step 2: Compute the Variance $\mathrm{Var}(X)$

Formula:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

First, compute $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = (-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} = \frac{1+0+1}{3} = \frac{2}{3}$$

Now:

$$\mathrm{Var}(X) = \frac{2}{3} - 0^2 = \frac{2}{3}$$

> 📘 **Explanation**: Variance is the expected squared distance from the mean. Since the mean is 0, it simplifies to the average of the squares.

WolframAlpha Input:

```
variance {-1:1/3, 0:1/3, 1:1/3}
```

Step 3: Compute $\mathrm{Var}(X - Y)$

Formula for independent variables:

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$= \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

> 📘 **Explanation**: Variance adds up for independent variables, whether added or subtracted.

WolframAlpha Input:

```
2/3 + 2/3
```

Step 4: Compute $\operatorname{Cov}(X, Y + X)$

Property of covariance:

$$\operatorname{Cov}(X, Y + X) = \operatorname{Cov}(X, Y) + \operatorname{Cov}(X, X)$$

Since (X) and (Y) are independent:

$$\operatorname{Cov}(X, Y) = 0, \quad \operatorname{Cov}(X, X) = \operatorname{Var}(X)$$

$$\Rightarrow \operatorname{Cov}(X, Y + X) = 0 + \frac{2}{3} = \frac{2}{3}$$

> 📘 **Explanation**: Covariance between a variable and a sum includes its variance if it appears in the sum.

WolframAlpha Input:

```
covariance(x, x + y), var(x) = 2/3
```

Step 5: Compute $\mathbb{E}[X - 1]$

$$\mathbb{E}[X - 1] = \mathbb{E}[X] - 1 = 0 - 1 = -1$$

> 📘 **Explanation**: Constants shift expectations linearly.

Step 6: Compute $\mathbb{E}[X - 1 \mid Y = -1]$

$$\mathbb{E}[X - 1 \mid Y = -1] = \mathbb{E}[X - 1]$$

Since (X) and (Y) are independent, conditioning on (Y) doesn't affect (X):

$$= -1$$

Step 7: Compute $P(X \neq 0)$

$$P(X \neq 0) = P(X = -1) + P(X = 1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

> 📘 **Explanation**: Just exclude the case where (X = 0).

WolframAlpha Input:

```
1/3 + 1/3
```

Step 8: Compute (P(X = 0))

Directly from the PMF:

$$P(X = 0) = \frac{1}{3}$$

Step 9: Compute (P(XY = 0))

We are interested in the probability that the product of two independent variables (X) and (Y) is equal to zero:

$$P(XY = 0) = P(X = 0 \text{ or } Y = 0)$$

Using the inclusion-exclusion principle and independence:

$$P(XY = 0) = P(X = 0) + P(Y = 0) - P(X = 0 \wedge Y = 0)$$

We are given:

$P(X = 0) = \frac{1}{3}$
$P(Y = 0) = \frac{1}{3}$
(X) and (Y) are independent, so:

$$P(X = 0 \wedge Y = 0) = P(X = 0) \cdot P(Y = 0) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Putting it together:

$$P(XY = 0) = \frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{6}{9} - \frac{1}{9} = \frac{5}{9}$$

✅ **Final Answer:**

$$\boxed{P(XY = 0) = \frac{5}{9}}$$

🔢 **WolframAlpha Input: (1/3) + (1/3) - (1/3)*(1/3)**

✅ Final Answer Summary

$$\begin{aligned}
&\mathbb{E}[X] = 0 \\
&\mathrm{Var}(X) = \tfrac{2}{3} \\
&\mathrm{Var}(X - Y) = \tfrac{4}{3} \\
&\mathrm{Cov}(X, Y + X) = \tfrac{2}{3} \\
&\mathbb{E}[X - 1] = -1 \\
&\mathbb{E}[X - 1 \mid Y = -1] = -1 \\
&P(X \neq 0) = \tfrac{2}{3} \\
&P(X = 0) = \tfrac{1}{3} \\
&P(XY = 0) = \tfrac{5}{9}
\end{aligned}$$

🔢 Python Code Verification

```
import numpy as np

x_vals = np.array([-1, 0, 1])
probs = np.array([1/3, 1/3, 1/3])

# Expectation
E_X = np.sum(x_vals * probs)

# Variance
E_X2 = np.sum(x_vals**2 * probs)
```

```
Var_X = E_X2 - E_X**2

# Var(X - Y)
Var_X_minus_Y = 2 * Var_X

# Cov(X, Y + X)
Cov_X_Y_plus_X = Var_X

# E[X - 1]
E_X_minus_1 = E_X - 1

print("E[X] =", E_X)
print("Var[X] =", Var_X)
print("Var[X - Y] =", Var_X_minus_Y)
print("Cov(X, Y + X) =", Cov_X_Y_plus_X)
print("E[X - 1] =", E_X_minus_1)
print("E[X - 1 | Y = -1] =", E_X_minus_1)
print("P(X ≠ 0) =", 2/3)
print("P(X = 0) =", 1/3)
```

Probability Calculations

Let

$$X \sim \text{Binomial}(10, 0.5)$$

and

$$Y \sim \text{Geometric}(1/5)$$

, where the geometric distribution counts the number of **trials until the first success**, including the success (i.e., support:

$$Y \in \{1, 2, 3, \ldots\}$$

).

We are asked to compute:

1. $$P(X > 5 \cup Y \leq 3)$$
2. $$P(X > 5 \mid X + Y = 3)$$

1.

$$P(X > 5 \cup Y \leq 3)$$

Using the inclusion-exclusion principle:

$$P(X > 5 \cup Y \leq 3) = P(X > 5) + P(Y \leq 3) - P(X > 5 \cap Y \leq 3)$$

Step 1: Compute

$$P(X > 5)$$

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{k=0}^{5} \binom{10}{k} (0.5)^{10}$$

Wolfram Alpha input:

```
1 - sum(binomial(10, k) * 0.5^10, k = 0 to 5)
```

Step 2: Compute

$$P(Y \leq 3)$$

$$P(Y \leq 3) = \sum_{k=1}^{3} (1-p)^{k-1} p = \sum_{k=1}^{3} (4/5)^{k-1} (1/5)$$

Wolfram Alpha input:

```
sum((4/5)^(k - 1)*(1/5), k = 1 to 3)
```

Step 3: Compute

$$P(X > 5 \cap Y \leq 3)$$

Since

$$X$$

and

$$Y$$

are independent:

$$P(X > 5 \cap Y \leq 3) = P(X > 5) \cdot P(Y \leq 3)$$

Final Calculation:

$$P(X > 5 \cup Y \leq 3) = P(X > 5) + P(Y \leq 3) - P(X > 5) \cdot P(Y \leq 3)$$

```python
```

```
import scipy.stats as stats

p_x_gt_5 = 1 - stats.binom.cdf(5, 10, 0.5)

p_y_le_3 = stats.geom.cdf(3, 1/5)

p_union = p_x_gt_5 + p_y_le_3 - (p_x_gt_5 * p_y_le_3)

print("P(X > 5 ∪ Y ≤ 3):", round(p_union, 4))
```

2.

$$P(X > 5 \mid X + Y = 3)$$

This is a conditional probability:

$$P(X > 5 \mid X + Y = 3) = \frac{P(X > 5 \cap X + Y = 3)}{P(X + Y = 3)}$$

Logical Argument:

If

$$X > 5$$

, then

$$X \geq 6$$

, so:

$$Y = 3 - X \leq -3$$

But

$$Y \geq 1$$

for the geometric distribution. So:

$$P(X > 5 \cap X + Y = 3) = 0 \Rightarrow P(X > 5 \mid X + Y = 3) = \frac{0}{P(X + Y = 3)} = 0$$

```python
denominator = sum(

stats.binom.pmf(x, 10, 0.5) * stats.geom.pmf(3 - x, 1/5)

for x in range(0, 3)

if 3 - x >= 1

)

print("P(X > 5 | X + Y = 3):", 0.0)
```

Final Answers

- $$P(X > 5 \cup Y \leq 3) = \boxed{0.6810}$$
- $$P(X > 5 \mid X + Y = 3) = \boxed{0.0000}$$

Conditional Expectation: $\mathbb{E}[X^2 \mid Y = 1]$

We are given the joint PDF of two continuous random variables X and Y:

$$f_{XY}(x, y) = \begin{cases} \frac{x^2}{4} + \frac{y^2}{4} + \frac{xy}{6}, & 0 \le x \le 1,\ 0 \le y \le 2 \\ 0, & \text{otherwise} \end{cases}$$

We need to compute:

$$\mathbb{E}[X^2 \mid Y = 1]$$

Step 1: Conditional PDF

We use:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

where:

$$f_Y(y) = \int_0^1 f_{XY}(x, y)\, dx$$

Step 2: Compute $f_Y(y)$

Substitute the expression inside the integral:

$$f_Y(y) = \int_0^1 \left(\frac{x^2}{4} + \frac{y^2}{4} + \frac{xy}{6} \right) dx$$

Integrate term-by-term:

$$= \left[\frac{x^3}{12} + \frac{xy^2}{4} + \frac{x^2 y}{12} \right]_0^1 = \frac{1}{12} + \frac{y^2}{4} + \frac{y}{12}$$

So:

$$f_Y(y) = \frac{1+y}{12} + \frac{y^2}{4}$$

And:

$$f_Y(1) = \frac{2}{12} + \frac{1}{4} = \frac{5}{12}$$

Step 3: Conditional PDF $f_{X|Y}(x|1)$

Substitute $y = 1$ into the joint PDF:

$$f_{XY}(x,1) = \frac{x^2}{4} + \frac{1}{4} + \frac{x}{6}$$

Common denominator 12:

$$f_{XY}(x,1) = \frac{3x^2 + 3 + 2x}{12}$$

Then:

$$f_{X|Y}(x|1) = \frac{f_{XY}(x,1)}{f_Y(1)} = \frac{\frac{3x^2+2x+3}{12}}{\frac{5}{12}} = \frac{3x^2 + 2x + 3}{5}$$

Step 4: Compute Conditional Expectation

Now compute:

$$\mathbb{E}[X^2 \mid Y = 1] = \int_0^1 x^2 \cdot \frac{3x^2 + 2x + 3}{5}\, dx = \frac{1}{5} \int_0^1 (3x^4 + 2x^3 + 3x^2)\, dx$$

Compute each integral:

- $\int_0^1 3x^4 dx = \frac{3}{5}$

- $\int_0^1 2x^3 dx = \frac{2}{4} = \frac{1}{2}$
- $\int_0^1 3x^2 dx = 1$

So:

$$\mathbb{E}[X^2 \mid Y = 1] = \frac{1}{5}(\frac{3}{5} + \frac{1}{2} + 1) = \frac{1}{5} \cdot \frac{21}{10} = \frac{21}{50}$$

✅ Final Answer

$$\mathbb{E}[X^2 \mid Y = 1] = \boxed{\frac{21}{50}}$$

🔢 WolframAlpha Input

To compute in WolframAlpha, input:

```
integrate(x^2 * (3x^2 + 2x + 3)/5, x=0 to 1)
```

🐍 Python Code

```python
import sympy as sp

x, y = sp.symbols('x y')
```

```python
# Joint PDF

f_xy = x**2 / 4 + y**2 / 4 + x * y / 6



# Compute f_Y(1)

f_y = sp.integrate(f_xy, (x, 0, 1))

f_y_1 = f_y.subs(y, 1)



# Evaluate f_{XY}(x,1)

f_xy_y1 = f_xy.subs(y, 1)



# Conditional PDF

f_x_given_y1 = f_xy_y1 / f_y_1



# Compute E[X^2 | Y=1]

E_x2_given_y1 = sp.integrate(x**2 * f_x_given_y1, (x, 0, 1))

E_x2_given_y1_simplified = sp.simplify(E_x2_given_y1)



print("E[X^2 | Y = 1] =", E_x2_given_y1_simplified)

# Output: 21/50
```

📧 Spam Filter Probability Analysis

We are given the following information for designing a spam filter:

- 60% of emails are spam.
- 2% of spam emails contain the word “discount”.
- 0.2% (0.002) of non-spam emails contain the word “discount”.

Define Events

Let’s define the events:

- S: An email is spam.
- NS: An email is not spam.
- D: An email contains the word “discount”.

From the given information, we can write the probabilities:

- $P(S) = 0.60$
- $P(NS) = 1 - P(S) = 1 - 0.60 = 0.40$
- $P(D|S) = 0.02$
- $P(D|NS) = 0.002$

---

1. The Probability That an Email Contains the Word “Discount” ($P(D)$)

To find $P(D)$, we use the **Law of Total Probability**:

$$P(D) = P(D|S)P(S) + P(D|NS)P(NS)$$

Substitute the given values:

$$P(D) = (0.02)(0.60) + (0.002)(0.40) = 0.012 + 0.0008 = 0.012008$$

Rounding to four decimal places:

$$P(D) \approx 0.0120$$

---

2. The Probability That an Email is Spam Given it Contains “Discount” ($P(S|D)$)

To find $P(S|D)$, we use **Bayes’ Theorem**:

$$P(S|D) = \frac{P(D|S)P(S)}{P(D)}$$

Substitute the known values:

$$P(S|D) = \frac{(0.02)(0.60)}{0.012008} = \frac{0.012}{0.012008}$$

Performing the division:

$$P(S|D) \approx 0.9993337775$$

Rounding to four decimal places:

$$P(S|D) \approx 0.9993$$

✅ Final Results

- $P(D)$ (probability an email contains “discount”): **0.0120**
- $P(S|D)$ (probability it’s spam given it contains “discount”): **0.9993**

🐍 Python Code

```
# Given probabilities

P_S = 0.60

P_NS = 1 - P_S

P_D_given_S = 0.02

P_D_given_NS = 0.002


# Law of Total Probability

P_D = P_D_given_S * P_S + P_D_given_NS * P_NS


# Bayes' Theorem

P_S_given_D = (P_D_given_S * P_S) / P_D


print("P(D) =", round(P_D, 4)) # Output: 0.0120

print("P(S|D) =", round(P_S_given_D, 4)) # Output: 0.9993
```

🔢 WolframAlpha Inputs

To verify:

- Total probability:

```
(0.02)*(0.60) + (0.002)*(0.40)
```

- Bayes' theorem:

```
(0.02*0.60)/((0.02*0.60)+(0.002*0.40))
```

🚙 Hypothesis Testing Using Poisson Distribution – Storebælt Bridge

This analysis investigates whether the average number of cars passing a bridge is **different from 400 cars per hour**. We use a **Poisson process** model.

📘 Background

The **Poisson distribution** is used for modeling the number of events in a fixed interval of time, assuming:

- Events happen independently
- They occur at a constant average rate

If events occur at a rate of λ per hour, and we observe for t hours, then:

$$X \sim \text{Poisson}(\lambda t)$$

For large λt, we approximate the Poisson distribution using a **Normal distribution**:

$$X \sim \mathcal{N}(\mu = \lambda t, \sigma^2 = \lambda t)$$

🔢 Given:

- Observation: $x = 1140$ cars in $t = 3$ hours
- Claimed rate: $\lambda = 400$ cars/hour

✅ Step 1: State the Hypotheses

We want to test whether the true average rate of cars is **different** from 400 per hour.

- Null Hypothesis:

$$H_0 : \lambda = 400$$

- Alternative Hypothesis:

$$H_1 : \lambda \neq 400$$

This is a **two-tailed test**.

✅ Step 2: Determine the Distribution Under H_0

If $\lambda = 400$ and $t = 3$ hours, then:

- Expected value:

$$\mu = \lambda t = 400 \cdot 3 = 1200$$

- Variance:

$$\sigma^2 = \lambda t = 1200 \Rightarrow \sigma = \sqrt{1200} \approx 34.641$$

Thus, we approximate:

$$X \sim \mathcal{N}(1200, 1200)$$

✅ Step 3: Compute the Z-Statistic

We compare the observed value $x = 1140$ to the expected mean using:

$$Z = \frac{x - \mu}{\sigma} = \frac{1140 - 1200}{\sqrt{1200}} = \frac{-60}{34.641} \approx -1.7321$$

Interpretation: The observed number is **1.73 standard deviations below** the expected mean.

✓ This gives us: **Z = -1.7321**

✅ Step 4: Compute the P-value

Since this is a two-tailed test, the p-value is:

$$p = 2 \cdot P(Z < -1.7321)$$

Using standard normal tables or Python:

$$P(Z < -1.7321) \approx 0.04165 \Rightarrow p = 2 \cdot 0.04165 = 0.0833$$

✓ Final p-value: **0.0833**

✅ Step 5: Decision Rule

Assume a significance level of $\alpha = 0.05$:

- If $p < 0.05$, reject H_0
- If $p \geq 0.05$, fail to reject H_0

Here:

$$p = 0.0833 > 0.05 \Rightarrow \text{Fail to reject } H_0$$

Conclusion: There is not enough evidence to say the average is different from 400 cars/hour.

✅ Step 6: Compute $P(X \leq 380)$ for 1 Hour

We now want to know:

> What is the probability of observing **380 or fewer** cars in **1 hour**, assuming $\lambda = 400$?

This is:

$$P(X \le 380), \quad X \sim \text{Poisson}(400)$$

Using Python or a calculator:

$$P(X \le 380) \approx 0.1649$$

✓ This is the probability used to assess if low counts like 380 are unusual.

🔍 Step-by-Step Explanation

🔢 Step 1: Understand the Poisson Distribution

The Poisson distribution models the probability of a number of events occurring in a fixed interval when events occur independently at a known constant mean rate.

The formula for the **Poisson probability mass function** is:

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

For the **cumulative probability** $P(X \le 380)$, we need:

$$P(X \le 380) = \sum_{k=0}^{380} \frac{e^{-400} \cdot 400^k}{k!}$$

But computing this **by hand** is impractical because:

- The sum includes **381 terms**
- The factorial and exponential values become **very large**

⚠️ Step 2: Use a CDF Function

To efficiently compute this, we use a **Poisson cumulative distribution function** (CDF). It computes:

$$P(X \leq x) = \text{CDF at } x \text{ for Poisson}(\lambda)$$

We use:

- Python: `scipy.stats.poisson.cdf(380, 400)`
- Or WolframAlpha: `poissoncdf(400, 380)`

✅ Step 3: Compute the Value

Using Python:

```
from scipy.stats import poisson


# Lambda = 400, x = 380

p = poisson.cdf(380, 400)

print(round(p, 4)) # Output: 0.1649
```

Using WolframAlpha:

```
poissoncdf(400, 380)
```

✓ This gives us:

$$P(X \leq 380) \approx 0.1649$$

🧠 Step 4: Interpretation

This means:

- There is a **16.49% chance** that 380 or fewer cars pass the bridge in an hour **if the true average rate is 400**.
- Since 380 is below the mean (400), this helps us assess whether such a **low value is unusual**.

In a **hypothesis test**, this would correspond to the left tail probability when checking whether observed values differ significantly from the mean.

✅ Final Answer

$$\boxed{P(X \le 380) \approx 0.1649}$$

🐍 Python Code

```
import scipy.stats as stats

import math



# Observed data

x = 1140

t = 3

lambda0 = 400

lambda_total = lambda0 * t



# Normal approximation
```

```
mu = lambda_total

sigma = math.sqrt(mu)



# Z-statistic

z_stat = (x - mu) / sigma



# Two-tailed p-value

p_value = 2 * stats.norm.cdf(z_stat)



# Poisson CDF for 1 hour

prob_c = stats.poisson.cdf(380, 400)



print("Z-statistic:", round(z_stat, 4)) # -1.7321

print("P-value:", round(p_value, 4)) # 0.0833

print("P(X <= 380):", round(prob_c, 4)) # 0.1649
```

🔢 WolframAlpha Inputs

- Z-statistic:

```
(1140 - 1200)/sqrt(1200)
```

- P-value:

```
2 * normalcdf(-10, -1.7321)
```

- Poisson CDF:

```
poissoncdf(400, 380)
```

✅ Final Summary

Step	Result
Hypotheses	$H_0 : \lambda = 400$, $H_1 : \lambda \neq 400$
Z-Statistic	-1.7321
P-value	0.0833
$P(X \leq 380)$ (1 hr)	0.1649
Decision ($\alpha = 0.05$)	Fail to reject H_0

> 💡 **Conclusion**: The observed traffic is not significantly different from what we would expect under the claimed average rate.

🧮 Covariance Calculation: $R = 7 - P + P^2, S = 4 - PQ$ with $P, Q \sim N(0,1)$

📌 Problem Statement

We are given two independent standard normal random variables:

- $P \sim N(0,1)$
- $Q \sim N(0,1)$
- $P \perp Q$ (independent)

And we define two new variables:

- $R = 7 - P + P^2$
- $S = 4 - PQ$

We are asked to find the **covariance**:

$$\mathrm{Cov}(R, S)$$

🔍 Step-by-Step Approach

🧠 What is Covariance?

Covariance between two random variables R and S is defined as:

$$\mathrm{Cov}(R, S) = \mathbb{E}[(R - \mathbb{E}[R])(S - \mathbb{E}[S])] = \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S]$$

To compute this, we need:

1. RS
2. $\mathbb{E}[RS]$
3. $\mathbb{E}[R]$
4. $\mathbb{E}[S]$

🔢 Step 1: Compute RS

Substitute definitions:

$$R = 7 - P + P^2 \quad \text{and} \quad S = 4 - PQ$$

So:

$$RS = (7 - P + P^2)(4 - PQ)$$

We expand this product:

$$RS = 7 \cdot 4 - 7PQ - 4P + P \cdot PQ + 4P^2 - P^2 \cdot PQ$$

$$RS = 28 - 7PQ - 4P + P^2Q + 4P^2 - P^3Q$$

Now take expectation of each term:

🔢 Step 2: Compute $\mathbb{E}[RS]$

Use linearity of expectation:

$$\mathbb{E}[RS] = \mathbb{E}[28] - \mathbb{E}[7PQ] - \mathbb{E}[4P] + \mathbb{E}[P^2Q] + \mathbb{E}[4P^2] - \mathbb{E}[P^3Q]$$

Break down each expectation:

- $\mathbb{E}[28] = 28$
- $\mathbb{E}[PQ] = \mathbb{E}[P]\mathbb{E}[Q] = 0$
- $\mathbb{E}[P] = 0$

- $\mathbb{E}[P^2Q] = \mathbb{E}[P^2] \cdot \mathbb{E}[Q] = 1 \cdot 0 = 0$
- $\mathbb{E}[P^2] = \text{Var}(P) = 1$
- $\mathbb{E}[P^3Q] = \mathbb{E}[P^3] \cdot \mathbb{E}[Q] = 0 \cdot 0 = 0$ (since odd moments of standard normal are 0)

Putting it all together:

$$\mathbb{E}[RS] = 28 - 0 - 0 + 0 + 4 \cdot 1 - 0 = 28 + 4 = 32$$

🔢 Step 3: Compute $\mathbb{E}[R]$ and $\mathbb{E}[S]$

Recall:

- $R = 7 - P + P^2$

$$\mathbb{E}[R] = 7 - \mathbb{E}[P] + \mathbb{E}[P^2] = 7 - 0 + 1 = 8$$

- $S = 4 - PQ$

$$\mathbb{E}[S] = 4 - \mathbb{E}[PQ] = 4 - 0 = 4$$

🔢 Step 4: Plug into Covariance Formula

$$\text{Cov}(R, S) = \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S] = 32 - (8)(4) = 32 - 32 = \boxed{0}$$

✅ Final Answer

$$\boxed{\text{Cov}(R, S) = 0}$$

🐍 Optional Python Code

```python
import sympy as sp


P, Q = sp.symbols('P Q')
f = (7 - P + P**2) * (4 - P * Q)


# Expectation of each term
E_RS = sp.expand(f)


# Now compute expected value using known expectations:
# E[P] = 0, E[Q] = 0, E[P^2] = 1, E[P^3] = 0, E[P^2 Q] = 0, etc.


E_RS_eval = 28 + 0 + 0 + 0 + 4*1 - 0
E_R = 7 - 0 + 1
E_S = 4 - 0


Cov = E_RS_eval - E_R * E_S
print("Cov(R, S) =", Cov)
```

📌 Notes

- This problem tests understanding of covariance, independence, and expectation rules.

- Independence is crucial: it allows factorization of expectations like $\mathbb{E}[PQ] = \mathbb{E}[P]\mathbb{E}[Q]$.
- Odd moments (like $\mathbb{E}[P]$ and $\mathbb{E}[P^3]$) of standard normals are zero.