Final Exam Prep

Probability & Statistics

Vrije Universiteit Amsterdam

This document is not part of the official course material

2025/2026

This handbook contains fully worked solutions to all problems from three past exams:

- AI Final Exam - March 2025 (20 questions)
- CS Final Exam - October 2024 (24 questions)
- CS Resit Exam - January 2025 (24 questions)

Last updated: March 2025

Contents

AI Exam – March 2025

Q1 – Conditional probability with coin tosses

AI Exam Q1

You toss a fair coin three times. Given that you observed at least one heads, what is the probability that you observed at least two heads?

A $4/7$
B $3/7$
C $1/5$
D $3/7$

Solution

We use the **definition of conditional probability**:

$$P(A_2 \mid A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)}$$

where $A_1 = \{\text{at least one heads}\}$ and $A_2 = \{\text{at least two heads}\}$.
The sample space has $2^3 = 8$ equally likely outcomes. Since every outcome with ≥ 2 heads automatically has ≥ 1 heads, we have $A_2 \subseteq A_1$, so $A_2 \cap A_1 = A_2$.
Compute $P(A_1)$: The only outcome with zero heads is TTT, so

$$P(A_1) = 1 - P(\text{TTT}) = 1 - \frac{1}{8} = \frac{7}{8}$$

Compute $P(A_2)$: The outcomes with ≥ 2 heads are HHH, HHT, HTH, THH, giving

$$P(A_2) = \frac{4}{8} = \frac{1}{2}$$

Apply the formula:

$$P(A_2 \mid A_1) = \frac{P(A_2)}{P(A_1)} = \frac{4/8}{7/8} = \boxed{\frac{4}{7}}$$

Answer: A

Q2 – Conditional probability with two dice

AI Exam Q2

Suppose you roll two fair dice. What is the probability that their sum is at least 7, given that exactly one of the dice shows a 3?

A $2/5$
B $1/3$
C $3/5$
D $1/2$

Solution

We use the formula $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$, but since all outcomes in the conditioned event are equally likely, we can simply count.

Step 1 – List the conditioned event. "Exactly one die shows 3" gives pairs $(3, x)$ and $(x, 3)$ with $x \neq 3$:

$$(3,1),(3,2),(3,4),(3,5),(3,6),(1,3),(2,3),(4,3),(5,3),(6,3) \;\Rightarrow\; 10 \text{ outcomes}$$

Step 2 – Count favourable outcomes (sum ≥ 7):

$$(3,4),(3,5),(3,6),(4,3),(5,3),(6,3) \;\Rightarrow\; 6 \text{ outcomes}$$

Step 3 – Compute the probability:

$$P(\text{sum} \geq 7 \mid \text{exactly one 3}) = \frac{6}{10} = \boxed{\frac{3}{5}}$$

Answer: C

Q4 – Computing a 95% CI for a population proportion

AI Exam Q4

A study was conducted to estimate the proportion of people in a city who prefer public transportation. A random sample shows that 290 out of 400 people prefer public transportation. What is an approximate 95% confidence interval for the proportion of people who prefer public transportation?

A $[0.68, 0.77]$
B $[0.66, 0.79]$
C $[0.70, 0.75]$
D $[0.64, 0.81]$

Solution

The approximate confidence interval for a population proportion is given by:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

where $\hat{p}$ is the sample proportion, n is the sample size, and $z_{\alpha/2}$ is the critical value from the standard normal distribution.

Step 1 – Identify the values:

$$\hat{p} = \frac{290}{400} = 0.725, \quad n = 400, \quad z_{0.025} = 1.96 \text{ (for 95\% confidence)}$$

Step 2 – Compute the standard error:

$$\text{SE} = \sqrt{\frac{0.725 \times 0.275}{400}} = \sqrt{\frac{0.199375}{400}} = \sqrt{0.000498} \approx 0.02233$$

Step 3 – Compute the margin of error:

$$E = 1.96 \times 0.02233 \approx 0.0438$$

Step 4 – Build the interval:

$$0.725 \pm 0.0438 \;\Rightarrow\; [0.681,\ 0.769] \approx \boxed{[0.68,\ 0.77]}$$

Answer: A

Q5 – CI for a mean with known variance

AI Exam Q5

A random sample $X_1, \ldots, X_{100}$ is given from a distribution with known variance $\sigma^2 = 16$. For the observed sample, the sample mean is 23.5. An approximate 95% confidence interval for $\theta = E(X_i)$ is

A $[22.3, 24.6]$
B $[22.2, 24.7]$
C $[22.1, 24.9]$
D $[22.7, 24.3]$

Solution

When the population variance σ^2 is known, the confidence interval for the mean is:

$$\bar{X} \;\pm\; z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}$$

Step 1 – Identify the values:

$$\bar{X} = 23.5, \quad \sigma = \sqrt{16} = 4, \quad n = 100, \quad z_{0.025} = 1.96$$

Step 2 – Compute the margin of error:

$$E = 1.96 \times \frac{4}{\sqrt{100}} = 1.96 \times \frac{4}{10} = 1.96 \times 0.4 = 0.784$$

Step 3 – Build the interval:

$$23.5 \pm 0.784 \;\Rightarrow\; [22.716,\ 24.284] \approx \boxed{[22.7,\ 24.3]}$$

Answer: D

Q6 – Properties of the CDF

AI Exam Q6

Which of the following is a property of the cumulative distribution function (CDF)?

A Is it differentiable.
B It is non-decreasing.
C It is continuous.
D It can decrease in a countable number of points.

Solution

The three fundamental properties of *any* CDF F are:

1. **Non-decreasing:** if $a < b$, then $F(a) \leq F(b)$.
2. **Right-continuous:** $\lim_{x \downarrow a} F(x) = F(a)$ for all a.
3. **Boundary limits:** $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

Let us check each option:

A: No. A CDF can have jump discontinuities (e.g. for discrete random variables), so it is not necessarily differentiable.

B: Yes. This is one of the defining properties.

C: No. Discrete CDFs have jumps at the probability mass points.

D: No. A CDF can *never* decrease; it is non-decreasing everywhere.

Answer: B

Q7 – Expected value from a power-law PDF

AI Exam Q7

Let X be a random variable with probability density function $f(x) = 3x^{-4}$ for $x > 1$ and $f(x) = 0$ otherwise. The expected value of X equals (round to 1 decimal):

Solution

The expected value is computed using:

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Since $f(x) = 3x^{-4}$ only for $x > 1$:

$$E(X) = \int_{1}^{\infty} x \cdot 3x^{-4}\, dx = 3 \int_{1}^{\infty} x^{-3}\, dx$$

Evaluate the integral:

$$\int_{1}^{\infty} x^{-3}\, dx = \left[\frac{x^{-2}}{-2}\right]_{1}^{\infty} = \left(0 - \left(-\frac{1}{2}\right)\right) = \frac{1}{2}$$

Therefore:

$$E(X) = 3 \times \frac{1}{2} = \frac{3}{2} = \boxed{1.5}$$

Q8 – LOTUS: computing $E(X^2)$ from a PDF

AI Exam Q8

Let X be a random variable with probability density function $f(x) = 4x(1-x)$ for $x \in [0,1]$ and $f(x) = 0$ otherwise. $E(X^2)$ equals:

A $1/4$
B $3/4$
C $1/5$
D $4/5$

Solution

We apply the **Law of the Unconscious Statistician (LOTUS)**:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

With $g(x) = x^2$ and $f(x) = 4x(1-x)$:

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

Evaluate each term:

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

Therefore:

$$E(X^2) = 4\left(\frac{1}{4} - \frac{1}{5}\right) = 4 \times \frac{5-4}{20} = 4 \times \frac{1}{20} = \boxed{\frac{1}{5}}$$

Answer: C

Q9 – Standardizing a normal distribution

AI Exam Q9

If $X \sim N(3,4)$ and if Φ denotes the cumulative distribution function of the standard normal distribution, then $P(X > 4)$ equals

A $1 - \Phi(0.5)$
B $\Phi(0.5)$
C $1 - \Phi(0.25)$
D $2\Phi(0.25)$

Solution

When $X \sim N(\mu, \sigma^2)$, we standardize using $Z = \frac{X-\mu}{\sigma}$, where $Z \sim N(0,1)$.
Here $\mu = 3$ and $\sigma^2 = 4$, so $\sigma = 2$. Therefore:

$$P(X > 4) = P\left(\frac{X-3}{2} > \frac{4-3}{2}\right) = P(Z > 0.5) = 1 - \Phi(0.5)$$

Tip

Remember: $N(\mu, \sigma^2)$ means the second parameter is the **variance**, not the standard deviation. So $\sigma = \sqrt{4} = 2$, not 4.

Answer: A

Q10 – Finding the normalizing constant of a PDF

AI Exam Q10

Let X be a random variable with probability density function $f(x) = ce^{-x}$ for $1 < x < 2$ and $f(x) = 0$ otherwise. Determine the value of c.

A $e + 1$
B $e/(e + 1)$
C $(e + 1)/(e - 1)$
D $e^2/(e - 1)$

Solution

A valid PDF must satisfy $\int_{-\infty}^{\infty} f(x)\,dx = 1$. So we require:

$$\int_1^2 c\,e^{-x}\,dx = 1$$

Evaluate the integral:

$$\int_1^2 e^{-x}\,dx = \left[-e^{-x}\right]_1^2 = -e^{-2} - (-e^{-1}) = e^{-1} - e^{-2}$$

Factor out e^{-1}:

$$e^{-1} - e^{-2} = e^{-1}(1 - e^{-1}) = \frac{1}{e} \cdot \frac{e-1}{e} = \frac{e-1}{e^2}$$

Set equal to 1 and solve:

$$c \cdot \frac{e-1}{e^2} = 1 \quad \Rightarrow \quad c = \boxed{\frac{e^2}{e-1}}$$

Answer: D

Q11 – Poisson distribution: probability over an interval

AI Exam Q11

The number of customers arriving at a grocery store is a Poisson random variable. On average 10 customers arrive per hour. Let X be the number of customers arriving between 10 a.m. and 11:30 a.m. What is $P(14 \leq X \leq 16)$? Round your answer to 2 decimals.

Solution

Step 1 – Determine the parameter. The time window is 1.5 hours. Since the rate is 10 customers/hour:

$$\lambda = 10 \times 1.5 = 15 \quad \Rightarrow \quad X \sim \text{Poisson}(15)$$

Step 2 – Apply the Poisson PMF. The Poisson probability mass function is:

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

Step 3 – Compute each term:

$$P(X = 14) = \frac{e^{-15} \cdot 15^{14}}{14!} \approx 0.1024$$

$$P(X = 15) = \frac{e^{-15} \cdot 15^{15}}{15!} \approx 0.1024$$

$$P(X = 16) = \frac{e^{-15} \cdot 15^{16}}{16!} \approx 0.0960$$

Step 4 – Sum:

$$P(14 \le X \le 16) = 0.1024 + 0.1024 + 0.0960 = \boxed{0.30}$$

Q12 – Computing the t-test statistic

AI Exam Q12

The average adult male height in a certain country is 170 cm. We suspect that adult males in a certain city in that country have a different average height due to some environmental factors. We pick a random sample of size 9 from the adult males in the city and find that its mean is 165.8 cm and its standard deviation is $S = 8.25$. We want to test the null distribution that the average height of adult males in this city is equal to the national average and assume that the heights are normally distributed. What is the observed value of the test-statistic? (round to two decimals)

Solution

When the population variance is unknown and the data are normal, we use the **t-test statistic**:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

where $\bar{X}$ is the sample mean, μ_0 is the hypothesized mean, S is the sample standard deviation, and n is the sample size.

Substitute the values:

$$T = \frac{165.8 - 170}{8.25/\sqrt{9}} = \frac{-4.2}{8.25/3} = \frac{-4.2}{2.75} \approx \boxed{-1.53}$$

Tip

The exact value is $-1.52727\ldots$ The official answer key lists -1.52, but -1.53 is also correct depending on rounding. Both were accepted.

Q13 – Definition of a Type II error

AI Exam Q13

A type II error occurs if

A A false null hypothesis is rejected.
B A true null hypothesis is rejected.
C A false null hypothesis is not rejected.
D A true null hypothesis is not rejected.

Solution

The two types of errors in hypothesis testing are defined as:

	H_0 is **true**	H_0 is **false**
Reject H_0	Type I error (α)	Correct decision (power)
Fail to reject H_0	Correct decision	Type II error (β)

A **Type II error** occurs when we *fail to reject* a null hypothesis that is actually *false* — i.e., we miss a real effect.
Answer: C

Q14 – Decision rule using p-values

AI Exam Q14

In a hypothesis test, the observed p-value is 0.04. If the significance level (α) is set to 0.05, what is the correct conclusion?

A Reject H_0.
B Accept H_0.
C Increase the sample size.
D None of the above.

Solution

The decision rule is:

$$\text{If } p\text{-value} < \alpha, \text{ reject } H_0. \qquad \text{If } p\text{-value} \geq \alpha, \text{ fail to reject } H_0.$$

Here p-value $= 0.04 < 0.05 = \alpha$, so we **reject** H_0.

Tip

We never "accept H_0" — we either reject it or fail to reject it. Failing to reject is not the same as proving H_0 true.

Answer: A

Q15 – Computing the probability of a Type I error

AI Exam Q15

Suppose $X_1, \ldots, X_9$ is a random sample from the $N(\mu, 1)$-distribution. We test the hypothesis $H_0\colon \mu = 1$ versus $H_1\colon \mu > 1$. We decide to reject H_0 if the sample mean is larger than 1.5. Denote the cumulative distribution function of the standard normal distribution by Φ. The probability of a type I error is

A $1 - \Phi(4.5)$
B $\Phi(3/2)$
C $\Phi(4/5)$
D $1 - \Phi(3/2)$

Solution

A **Type I error** is rejecting H_0 when H_0 is true. So we compute $P(\bar{X} > 1.5 \mid \mu = 1)$.
Step 1 – Distribution of $\bar{X}$ under H_0. With $\mu = 1$, $\sigma^2 = 1$, and $n = 9$:

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) = N\left(1,\ \frac{1}{9}\right) \quad \Rightarrow \quad \text{std. dev. of } \bar{X} = \frac{1}{3}$$

Step 2 – Standardize:

$$P(\bar{X} > 1.5) = P\left(\frac{\bar{X} - 1}{1/3} > \frac{1.5 - 1}{1/3}\right) = P\left(Z > \frac{0.5}{1/3}\right) = P(Z > 1.5)$$

Step 3 – Express using Φ:

$$P(Z > 1.5) = 1 - \Phi(1.5) = 1 - \Phi(3/2)$$

Answer: D

Q16 – Bayes' rule for defective items

AI Exam Q16

A company has two machines A and B. Machine A produces 60% of the items and machine B produces 40% of the items. Machine A has a 2% defect rate. The probability of a randomly selected item being defect equals 0.1. We randomly select an item and it turns out to be defect. The probability that the selected item was produced by machine A equals (round to 2 decimals):

Solution

This is a direct application of **Bayes' rule**:

$$P(A \mid D) = \frac{P(D \mid A)\, P(A)}{P(D)}$$

Identify the given values:

$$P(A) = 0.6, \quad P(D \mid A) = 0.02, \quad P(D) = 0.1$$

Substitute:

$$P(A \mid D) = \frac{0.02 \times 0.6}{0.1} = \frac{0.012}{0.1} = \boxed{0.12}$$

Q17 – Identifying true probability identities

AI Exam Q17

Which of the statements below are generally true?

A $P(A \cap B) \leq P(A \cup B)$
B $P(A \mid B) = P(A \cap B)/P(A)$
C $P(A \cup B \cup C) = 1 - P(A^c \cap B^c \cap C^c)$
D $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$

Solution

A: TRUE. Since $A \cap B \subseteq A \cup B$, by the **monotonicity** of probability: $P(A \cap B) \leq P(A \cup B)$.

B: FALSE. The correct definition of conditional probability is $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. The denominator should be $P(B)$, not $P(A)$.

C: TRUE. By **De Morgan's law**, $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$. Therefore $P(A \cup B \cup C) = 1 - P((A \cup B \cup C)^c) = 1 - P(A^c \cap B^c \cap C^c)$.

D: FALSE. The **inclusion–exclusion principle for conditional probability** gives $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$. The term $-P(A \cap B \mid C)$ is missing, so the formula is only valid when A and B are mutually exclusive given C.

Answer: A and C

Q18 – Inclusion–exclusion (sum rule)

AI Exam Q18

Consider two events A and B such that $P(A \cap B) = 0.1$, $P(A) = 0.3$, and $P(B) = 0.2$. What is $P(A \cup B)$?

A 0.2
B 0.4
C 0.1
D 0.5

Solution

Apply the **inclusion–exclusion principle** (also called the sum rule):

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

$$= 0.3 + 0.2 - 0.1 = \boxed{0.4}$$

Answer: B

Q19 – Countability of number sets

AI Exam Q19

Indicate whether the following statements are true or false.

	True	Not true
The set of rational numbers is uncountable.		
The Cartesian product of two countable sets is countable.		

Solution

- "The set of rational numbers is uncountable." – **Not true.**

  The rationals $\mathbb{Q}$ are *countable*. One can enumerate them using the Cantor pairing function on $\mathbb{Z} \times \mathbb{N}$. (In contrast, the *reals* $\mathbb{R}$ are uncountable by Cantor's diagonal argument.)

- "The Cartesian product of two countable sets is countable." – **True.**

  If A and B are countable, then $A \times B$ can be put in bijection with $\mathbb{N}$ (e.g. using a diagonal enumeration), so $A \times B$ is countable.

Q20 – CLT and unbiasedness of the sample mean

AI Exam Q20

Indicate whether the following statements are true or false:

	True	Not true
The Central Limit Theorem (CLT) says that, for a random sample $X_1, \ldots, X_n$ from a distribution with mean μ and variance σ^2, the sample mean $(X_1 + \cdots + X_n)/n$ converges to $N(\mu, \sigma^2/n)$.		
The sample mean is an unbiased estimator for the expected value μ of a distribution.		

Solution

Statement 1: True. The CLT states that for large n:

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

which is equivalent to saying $\bar{X}_n$ is approximately distributed as $N(\mu, \sigma^2/n)$.

Statement 2: True. By linearity of expectation:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu = \mu$$

Since $E(\bar{X}) = \mu$, the sample mean is an **unbiased** estimator for μ.

CS Final Exam – October 2024

Q1 – Bayes' rule for defective items (3-decimal version)

CS Final Q1

A company has two machines, let's call these A and B. Machine A produces 60% of the items and machine B produces 40% of the items. Machine A has a 2% defect rate. The probability of an item being defect equals 0.1. We are given a defect item. The probability that this item was produced by machine A equals (round to 3 decimals)

Solution

By **Bayes' rule**: $P(A \mid D) = \frac{P(D|A)\,P(A)}{P(D)}$.

$$P(A \mid D) = \frac{0.02 \times 0.6}{0.1} = \frac{0.012}{0.1} = \boxed{0.120}$$

Q2 – Birthday problem for 3 people

CS Final Q2

In a group of 3 people, what is the probability that at least two people share the same birthday (assuming 365 possible birthdays and ignoring leap years)? Round to three decimals.

Solution

Use the **complement**: $P(\text{at least one match}) = 1 - P(\text{all different})$.
For 3 people, the probability that all birthdays are different is:

$$P(\text{all different}) = 1 \times \frac{364}{365} \times \frac{363}{365}$$

The first person can have any birthday (probability 1), the second must differ from the first (364/365), and the third must differ from both (363/365).

$$P(\text{at least one match}) = 1 - \frac{364 \times 363}{365^2} = 1 - \frac{132132}{133225} \approx 1 - 0.9918 = \boxed{0.008}$$

Q3 – Identifying CDF properties (5-option version)

CS Final Q3

Which of the following is a property of any cumulative distribution function (CDF)?

A It is always non-decreasing.
B It can decrease at certain points.
C It is constant over the entire real line.
D Its value is always less than or equal to 0.
E It is always left-continuous.

Solution

Recall the three key CDF properties: non-decreasing, right-continuous, limits to 0 (at $-\infty$) and 1 (at $+\infty$).

A: **Correct.** A CDF is always non-decreasing.

B: False – CDFs never decrease.

C: False – it goes from 0 to 1.

D: False – $F(x) \in [0, 1]$.

E: False – CDFs are *right*-continuous, not left-continuous (they may have left-hand jumps at discrete mass points).

Answer: A

Q4 – 99% CI for a population proportion

CS Final Q4

A study was conducted to estimate the proportion of people in a city who prefer public transportation. A random sample of 400 people shows that 240 prefer public transportation. What is an approximate 99% confidence interval for the proportion of people who prefer public transportation?

A $(0.57, 0.63)$
B $(0.54, 0.66)$
C $(0.51, 0.69)$
D $(0.59, 0.61)$

Solution

Use the CI formula for a proportion: $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$.
Values: $\hat{p} = 240/400 = 0.6$, $n = 400$, $z_{0.005} = 2.58$ (for 99%).

$$\text{SE} = \sqrt{\frac{0.6 \times 0.4}{400}} = \sqrt{0.0006} \approx 0.02449$$

$$E = 2.58 \times 0.02449 \approx 0.0632$$

$$0.6 \pm 0.0632 \Rightarrow \boxed{(0.54,\ 0.66)}$$

Answer: B

Q5 – Identifying true probability identities (5-option version)

CS Final Q5

Which statements on rules for computing probabilities are generally true?

A $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$
B $P(A \cup B \mid C) = P(A \mid B \cup C)\, P(C)$
C $P(A \cup B \cup C) = 1 - P(A^c \cap B^c \cap C^c)$
D $P(A \mid B) = P(B \mid A)\, P(B)/P(A)$
E $P(A \cap B) = P(A \mid B)\, P(B)$

Solution

A: FALSE. The correct formula is $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$.

B: FALSE. No such identity exists in general.

C: TRUE. This follows from De Morgan's law: $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$.

D: FALSE. Bayes' rule states $P(A \mid B) = \frac{P(B|A)\, P(A)}{P(B)}$. Here $P(A)$ and $P(B)$ are swapped.

E: TRUE. This is the **multiplication rule**, which is the definition of conditional probability rearranged: $P(A \cap B) = P(A \mid B)\, P(B)$.

Answer: C and E

Q6 – Conditional probability with dice (sum = 7)

CS Final Q6

Suppose you roll two fair six-sided dice. What is the probability that the sum of the dice is 7, given that exactly one of the dice shows a 3?

A 1
B 1/6
C 1/5
D 1/3

Solution

Same setup as AI Q2: there are 10 outcomes where exactly one die shows 3.
Of these, sum = 7 requires the other die to be 4: outcomes $(3, 4)$ and $(4, 3) \Rightarrow 2$ favourable.

$$P(\text{sum} = 7 \mid \text{exactly one 3}) = \frac{2}{10} = \boxed{\frac{1}{5}}$$

Answer: C

Q7 – Expected value of a binomial distribution

CS Final Q7

A random variable X follows the Binomial distribution with parameters $n = 10$ and $p = 0.2$. The expected value of X equals (round to 1 decimal)

Solution

For $X \sim \text{Binomial}(n, p)$, the expected value formula is:

$$E(X) = np = 10 \times 0.2 = \boxed{2.0}$$

Q8 – Expected value from a power-law PDF (CS version)

CS Final Q8

Let X be a random variable with probability density function $f_X(x) = 3x^{-4}$ if $x > 1$. The expected value of X equals (round to 1 decimal)

Solution

$$E(X) = \int_1^\infty x \cdot 3x^{-4}\,dx = 3\int_1^\infty x^{-3}\,dx = 3\left[-\frac{x^{-2}}{2}\right]_1^\infty = 3\left(0 + \frac{1}{2}\right) = \boxed{1.5}$$

Q9 – Identifying a left-tailed test

CS Final Q9

Consider a testing problem with $H_0\colon \mu = 21$ versus $H_1\colon \mu < 21$. This test is

A right tailed (or right-sided).
B left tailed (or left-sided).
C cross tailed (or cross-sided).
D center tailed (or center-sided).

Solution

The alternative hypothesis $H_1\colon \mu < 21$ points to the **left** of the hypothesized value. The rejection region is in the left tail of the distribution.
Answer: B

Q10 – Definition of a Type I error

CS Final Q10

A type I error occurs if

A a false null hypothesis is rejected
B a true null hypothesis is rejected
C a false null hypothesis is not rejected
D a true null hypothesis is not rejected

Solution

A **Type I error** means rejecting the null hypothesis when it is actually true (a "false positive"). Its probability is the significance level α.
Answer: B

Q11 – Law of total probability for defect rates

CS Final Q11

A company has two machines, let's call these A and B. Machine A produces 60% of the items and machine B produces 40% of the items. Machine A has a 2% defect rate and Machine B has a 5% defect rate. The probability that a randomly selected item is defect equals (round to 3 decimals)

Solution

Apply the **law of total probability**:

$$P(D) = P(D \mid A)\,P(A) + P(D \mid B)\,P(B)$$

$$= 0.02 \times 0.6 + 0.05 \times 0.4 = 0.012 + 0.020 = \boxed{0.032}$$

Q12 – Choosing the correct LOTUS integral

CS Final Q12

Suppose X is a continuous random variable with probability density function $f_X(x) = 4x(1-x)$ if $x \in [0,1]$. What is $E[X^2]$?

A $\left(\int x\, 4x(1-x)\, dx\right)^2$
B $\int x\,(4x(1-x))^2\, dx$
C $\int x^2\, 4x(1-x)\, dx$
D $\int (x\, 4x(1-x))^2\, dx$

Solution

By **LOTUS**, for any function g:

$$E[g(X)] = \int g(x)\, f_X(x)\, dx$$

With $g(x) = x^2$ and $f_X(x) = 4x(1-x)$:

$$E[X^2] = \int_0^1 x^2 \cdot 4x(1-x)\, dx$$

This matches option C. **Answer: C**

Q13 – Standardizing $N(3,4)$ to compute a tail probability

CS Final Q13

If $X \sim N(3,4)$ and if Φ denotes the cumulative distribution function of the standard normal distribution, then $P(X > 4)$ equals

A $1 - \Phi(0.5)$
B $\Phi(0.5)$
C $1 - \Phi(0.25)$
D $\Phi(0.25)$
E $2\,\Phi(0.5)$
F $2\,\Phi(0.25)$

Solution

$X \sim N(3, 4)$ means $\mu = 3$, $\sigma^2 = 4$, $\sigma = 2$. Standardize with $Z = (X - \mu)/\sigma$:

$$P(X > 4) = P\left(Z > \frac{4-3}{2}\right) = P(Z > 0.5) = 1 - \Phi(0.5)$$

Answer: A

Q14 – Union probability for three independent events

CS Final Q14

Consider three events A, B and C in a sample space with $P(A) = 0.4$, $P(B) = 0.5$ and $P(C) = 0.3$. If A, B and C are mutually independent, what is $P(A \cup B \cup C)$?

A 0.64
B 0.76
C 0.82
D 0.91
E 0.79
F none of the given options

Solution

Use the **complement method** combined with **independence**:

$$P(A \cup B \cup C) = 1 - P(A^c \cap B^c \cap C^c) = 1 - P(A^c)\,P(B^c)\,P(C^c)$$

$$= 1 - (1 - 0.4)(1 - 0.5)(1 - 0.3) = 1 - 0.6 \times 0.5 \times 0.7 = 1 - 0.21 = \boxed{0.79}$$

Answer: E

Q15 – Rejecting H_0 when $p < \alpha$

CS Final Q15

In a hypothesis test, the p-value obtained is 0.04. If the significance level (α) is set to 0.05, what is the correct conclusion?

A Fail to reject H_0
B Reject H_0.
C Increase the sample size.
D Decrease the significance level.
E None of the above given answers.

Solution

Decision rule: reject H_0 if p-value $< \alpha$. Here $0.04 < 0.05$, so **reject** H_0. **Answer: B**

Q16 – Independence and disjointness of events (spam filter)

CS Final Q16

Suppose we developed some spam filter to identify email spam. It works as follows: if a message contains any of some flagged words, it is marked as spam. It is known that the probability of any message being spam is 20%, the probability that a flagged word is present in spam email is 0.9, whereas the probability that a flagged word is present in non-spam email is 0.05. An email message arrives. Denote W={message contains a flagged word}, S={message is spam}.
Indicate whether the below claims are true or false.

	True	Not true
The events W and S are independent.		
The events W and S are disjoint.		

Solution

Given: $P(S) = 0.2$, $P(W \mid S) = 0.9$, $P(W \mid S^c) = 0.05$.
Independent? Two events are independent iff $P(W \mid S) = P(W)$.
First compute $P(W)$ using the **law of total probability**:

$$P(W) = P(W \mid S)P(S) + P(W \mid S^c)P(S^c) = 0.9(0.2) + 0.05(0.8) = 0.18 + 0.04 = 0.22$$

Since $P(W \mid S) = 0.9 \neq 0.22 = P(W)$, the events are **not independent**. **Not true.**
Disjoint? Disjoint means $P(W \cap S) = 0$. But:

$$P(W \cap S) = P(W \mid S)\, P(S) = 0.9 \times 0.2 = 0.18 \neq 0$$

So they are **not disjoint**. **Not true.**

Q17 – Reverse application of the sum rule

CS Final Q17

Consider two events A and B such that $P(A \cup B) = 0.8$, $P(A) = 0.5$ and $P(B) = 0.4$. What is $P(A \cap B)$?

A 0.1
B 0.3
C 0.4
D 0.5

Solution

Rearrange the **inclusion–exclusion formula**:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \Rightarrow \quad P(A \cap B) = P(A) + P(B) - P(A \cup B)$$
$$= 0.5 + 0.4 - 0.8 = \boxed{0.1}$$

Answer: A

Q18 – Computing a p-value for a left-tailed Z-test

CS Final Q18

Suppose $X_1, X_2, \ldots X_n$ are independent random variables, each with the $N(\mu, \sigma^2)$-distribution. A researcher is testing the hypothesis $H_0 \colon \mu = 25$ *against* $H_1 \colon \mu < 25$. The sample mean is 24, the sample size (n) is 36 and the population standard deviation is known to be 6. What is the p-value for this test?
In the options below, Φ denotes the cumulative distribution function of the standard Normal distribution and φ denotes the probability density function of the standard Normal distribution.

A $\Phi(-1/6)$
B $\Phi(-1)$
C $\varphi(-1/6)$
D $\Phi(-6)$
E $1 - \Phi(-1)$
F $1 - \Phi(-1/6)$

Solution

Step 1 – Compute the test statistic. Since σ is known, we use the Z-test:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{24 - 25}{6/\sqrt{36}} = \frac{-1}{6/6} = \frac{-1}{1} = -1$$

Step 2 – Compute the p-value. For a **left-tailed** test ($H_1 \colon \mu < 25$), the p-value is the probability of observing a value as extreme or more extreme *to the left*:

$$\text{p-value} = P(Z \leq z_{\text{obs}}) = P(Z \leq -1) = \Phi(-1)$$

Answer: B

Q19 – Variance of the sample mean

CS Final Q19

If $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables from a distribution with mean 10 and variance 5. The variance of the sample mean (which is defined as the average of $X_1, X_2, \ldots, X_n$) is given by

A $5/n$
B $10/n$
C $5/\sqrt{n}$
D $25/n$
E $25/\sqrt{n}$
F $10/\sqrt{n}$

Solution

For iid random variables, the **variance of the sample mean** is:

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

With $\sigma^2 = 5$: $\mathrm{Var}(\bar{X}) = \boxed{5/n}$. **Answer: A**

Q20 – Unbiasedness and MSE of the sample mean

CS Final Q20

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the $N(\mu, 4)$-distribution.

	True	Not true
The sample mean is an unbiased estimator for μ.		
The mean squared error for estimating μ equals $2/\sqrt{n}$.		

Solution

Statement 1: $E(\bar{X}) = \mu$, so the sample mean is **unbiased**. **True.**
Statement 2: For an unbiased estimator, $\mathrm{MSE} = \mathrm{Var}(\bar{X}) + \mathrm{Bias}^2 = \mathrm{Var}(\bar{X})$.

$$\mathrm{MSE} = \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{4}{n}$$

Since $4/n \neq 2/\sqrt{n}$, this is **Not true.**

Q21 – MLE properties for the exponential distribution

CS Final Q21

Suppose $x_1, x_2, \ldots, x_n$ are realizations of independent random variables, where each X_i $(1 \leq i \leq n)$ has the Exponential distribution with parameter λ. Verify for the following statements whether they are true or false.

	True	Not true
The likelihood only depends on the sum of $x_1, x_2, \ldots, x_n$		
The maximum likelihood estimate equals the sample mean.		

Solution

Step 1 – Write the likelihood. For $X_i \sim \mathrm{Exp}(\lambda)$, with PDF $f(x) = \lambda e^{-\lambda x}$:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)$$

The data appear only through $\sum x_i$, so **True** for statement 1.
Step 2 – Find the MLE. Maximize the log-likelihood:

$$\ell(\lambda) = n\ln\lambda - \lambda\sum x_i \quad \Rightarrow \quad \frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum x_i = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

The MLE is $1/\bar{x}$ (the *reciprocal* of the sample mean), not $\bar{x}$ itself. **Not true** for statement 2.

Q22 – Computing $P(X > 1/3)$ by integrating a PDF

CS Final Q22

Suppose X is a random variable with probability density function $f(x) = x + 1/2$ for $0 \leq x \leq 1$ and zero for other values of x. The probability of the event $\{X > 1/3\}$ equals (round to 2 decimals)

Solution

$$P(X > 1/3) = \int_{1/3}^{1} f(x)\,dx = \int_{1/3}^{1} \left(x + \frac{1}{2}\right) dx = \left[\frac{x^2}{2} + \frac{x}{2}\right]_{1/3}^{1}$$

Evaluate at the bounds:

$$\text{At } x = 1: \quad \frac{1}{2} + \frac{1}{2} = 1 \qquad \text{At } x = \frac{1}{3}: \quad \frac{1}{18} + \frac{1}{6} = \frac{1}{18} + \frac{3}{18} = \frac{4}{18} = \frac{2}{9}$$

$$P(X > 1/3) = 1 - \frac{2}{9} = \frac{7}{9} \approx \boxed{0.78}$$

(Both 0.77 and 0.78 were accepted.)

Q23 – CDF of the geometric distribution at a non-integer

CS Final Q23

Suppose X has the geometric distribution with parameter 0.2. The value of the cumulative distribution function at 2.5 (i.e. $F_X(2.5)$) is given by (round to 2 decimals)

Solution

The geometric distribution takes values $k = 1, 2, 3, \ldots$ with PMF $P(X = k) = p(1-p)^{k-1}$. Since X is integer-valued and 2.5 is between 2 and 3:

$$F_X(2.5) = P(X \leq 2.5) = P(X = 1) + P(X = 2)$$

$$= 0.2 + 0.2 \times 0.8 = 0.2 + 0.16 = \boxed{0.36}$$

Q24 – Type I error probability from a custom rejection rule

CS Final Q24

Suppose $X_1, X_2, \ldots, X_9$ is a random sample from the $N(\mu, 1)$-distribution. We test the hypothesis $H_0\colon \mu = 1$ versus $H_1\colon \mu > 1$. We decide to reject H_0 if the sample mean is larger than 1.5. Denote the cumulative distribution function of the standard Normal distribution by Φ. The probability of a type I error is given by

A $\Phi(3/2)$
B $\Phi(4/5)$
C $1 - \Phi(4.5)$
D $1 - \Phi(3/2)$
E $\Phi(-3/2)$
F $\Phi(-4.5)$

Solution

Under H_0 ($\mu = 1$, $\sigma^2 = 1$, $n = 9$), the sample mean has distribution:

$$\bar{X} \sim N\left(1,\ \frac{1}{9}\right), \qquad \text{so } \frac{\bar{X} - 1}{1/3} \sim N(0,1)$$

Type I error $= P(\text{reject } H_0 \mid H_0 \text{ true}) = P(\bar{X} > 1.5 \mid \mu = 1)$:

$$= P\left(\frac{\bar{X} - 1}{1/3} > \frac{1.5 - 1}{1/3}\right) = P\left(Z > \frac{0.5}{1/3}\right) = P(Z > 1.5) = 1 - \Phi\left(\frac{3}{2}\right)$$

Answer: D

CS Resit Exam – January 2025

Q1 – Unbiasedness and MSE of the sample mean (correct MSE)

Resit Q1

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the $N(\mu, 4)$-distribution.

	True	Not true
The sample mean is an unbiased estimator for μ.		
The mean squared error for estimating μ equals $4/n$.		

Solution

Unbiased: $E(\bar{X}) = \mu$. **True.**
MSE: $\text{MSE}(\bar{X}) = \text{Var}(\bar{X}) = \sigma^2/n = 4/n$. **True.**

Tip

Compare with CS Final Q20, where MSE was claimed to be $2/\sqrt{n}$ (which was false). Here $4/n$ is the correct MSE.

Q2 – Negative binomial: exactly 5 tosses to get 2 heads

Resit Q2

We continue throwing a coin until we see 2 times head. The coin is unfair; the probability that it lands heads equals 1/3. What is the probability that we need to throw exactly 5 times?

A 0.03
B 0.13
C 0.24
D 0.10
E 0.08
F none of the given other options

Solution

To need exactly 5 throws for 2 heads: the 5th throw must be heads (the 2nd head), and among the first 4 throws there is exactly 1 head.
The number of ways to place 1 head in the first 4 positions: $\binom{4}{1} = 4$.
Each valid sequence has 2 heads and 3 tails, so the probability per sequence is $(1/3)^2 \cdot (2/3)^3$:

$$P = \binom{4}{1}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = 4 \times \frac{1}{9} \times \frac{8}{27} = \frac{32}{243} \approx \boxed{0.13}$$

Answer: B

Q3 – Union of three independent events (partial information)

Resit Q3

Suppose the events A, B and C are mutually independent. We know that $P(A^c \cap B^c) = 1/2$ and $P(C) = 1/3$. The probability of the event $A \cup B \cup C$ is given by (when rounded to 2 decimals)

A 0.12
B 0.25
C 0.37
D 0.66
E 0.67
F 0.83

Solution

Step 1 – Use independence to simplify. Since A, B, C are mutually independent, A^c and B^c are also independent:

$$P(A^c \cap B^c) = P(A^c)\,P(B^c) = \frac{1}{2}$$

Step 2 – Apply the complement method:

$$P(A \cup B \cup C) = 1 - P(A^c \cap B^c \cap C^c) = 1 - P(A^c)\,P(B^c)\,P(C^c)$$

$$= 1 - \frac{1}{2} \times \left(1 - \frac{1}{3}\right) = 1 - \frac{1}{2} \times \frac{2}{3} = 1 - \frac{1}{3} = \frac{2}{3} \approx \boxed{0.66}$$

Answer: D

Q4 – CLT approximation for a binomial probability

Resit Q4

Suppose $X \sim \text{Binomial}(20, 1/5)$. If we use the central limit theorem to approximate $P(X > 8)$, then we get (recall Φ is the cumulative distribution function of the standard normal distribution).

A $1 - \Phi(0.5\sqrt{5})$
B $1 - \Phi(\sqrt{5})$
C $1 - \Phi(2\sqrt{5})$
D $\Phi(0.5\sqrt{5})$
E $\Phi(\sqrt{5})$
F $\Phi(2\sqrt{5})$

Solution

Step 1 – Compute mean and variance:

$$\mu = np = 20 \times \frac{1}{5} = 4, \qquad \sigma^2 = np(1-p) = 20 \times \frac{1}{5} \times \frac{4}{5} = \frac{16}{5} = 3.2$$

$$\sigma = \sqrt{3.2} = \sqrt{\frac{16}{5}} = \frac{4}{\sqrt{5}}$$

Step 2 – Apply CLT:

$$P(X > 8) \approx P\left(Z > \frac{8-4}{4/\sqrt{5}}\right) = P\left(Z > \frac{4\sqrt{5}}{4}\right) = P(Z > \sqrt{5}) = 1 - \Phi(\sqrt{5})$$

Answer: B

Q5 – True/false on conditional probability identities

Resit Q5

Suppose we are given events A, B and C on a sample space. Indicate which statements are true:

	True	Not true
$P(C \mid C) = 1$		
$P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$		

Solution

Statement 1: By definition, $P(C \mid C) = \frac{P(C \cap C)}{P(C)} = \frac{P(C)}{P(C)} = 1$. **True.**
Statement 2: The full formula is $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$. The subtracted term is missing, so this only holds when A and B are mutually exclusive given C, which is not guaranteed in general. **Not true.**

Q6 – Independence vs. disjointness (true/false)

Resit Q6

Indicate true or false:

	True	Not true
If events A and B are independent, then they are disjoint.		
If A and B are disjoint, then they are independent.		

Solution

Both are **Not true.**
Independent $\not\Rightarrow$ disjoint: If A and B are independent with $P(A) > 0$ and $P(B) > 0$, then $P(A \cap B) = P(A)P(B) > 0$, so they are *not* disjoint.
Disjoint $\not\Rightarrow$ independent: If A and B are disjoint with $P(A) > 0$ and $P(B) > 0$, then $P(A \cap B) = 0 \neq P(A)P(B)$, so they are *not* independent.

Q7 – Normalizing constant for a symmetric PDF

Resit Q7

Let X be a random variable with probability density function $f(x) = cx^2$ if $x \in [-1, 1]$ and zero for all other values of x. The value of c should be

A 0.5
B 1
C 1.5
D 2/3
E 2
F −1

Solution

Require $\int_{-1}^{1} f(x)\,dx = 1$:

$$\int_{-1}^{1} cx^2\,dx = c\left[\frac{x^3}{3}\right]_{-1}^{1} = c\left(\frac{1}{3} - \frac{-1}{3}\right) = c \cdot \frac{2}{3} = 1 \quad \Rightarrow \quad c = \frac{3}{2} = \boxed{1.5}$$

Answer: C

Q8 – Conditional probability: dice with maximum condition

Resit Q8

Suppose you roll 2 fair dice. If we know the maximum of the 2 outcomes equals 3, then what is the probability that their sum equals 4? Round to 2 decimals.

Solution

Step 1 – Count outcomes where max = 3**.** Both dice must be ≤ 3, and at least one must equal 3.
Total pairs with both ≤ 3: $3 \times 3 = 9$. Pairs with both ≤ 2: $2 \times 2 = 4$. So $9 - 4 = 5$ outcomes:

$$(3,1),\ (3,2),\ (3,3),\ (1,3),\ (2,3)$$

Step 2 – Count favourable outcomes (sum = 4): $(3,1)$ and $(1,3)$ ⇒ 2 outcomes.

$$P(\text{sum} = 4 \mid \max = 3) = \frac{2}{5} = \boxed{0.40}$$

Q9 – Distribution of a sum of independent normals

Resit Q9

Suppose $X \sim N(3,1)$ and $Y \sim N(-3,2)$. Assume X and Y are independent. Define $Z = X + Y - 2$. Then Z has distribution:

A $N(-2, 5)$
B $N(-2, 3)$
C $N(-2, 3)$
D $N(0, 5)$
E $N(0, 3)$
F the distribution will not be a Normal (Gaussian) distribution

Solution

For independent normal random variables, a linear combination is also normal. The key formulas are:

$$E(aX + bY + c) = aE(X) + bE(Y) + c, \qquad \mathrm{Var}(aX + bY + c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$$

(constants do not affect variance).
Mean: $E(Z) = E(X) + E(Y) - 2 = 3 + (-3) - 2 = -2$
Variance: $\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 1 + 2 = 3$

$$Z \sim \boxed{N(-2,\ 3)}$$

Answer: B (or C, which are identical)

Q10 – Writing the likelihood for a binomial sample

Resit Q10

Suppose X_1, X_2, X_3, X_4, are independent random variables, each with the Binomial$(3, \theta)$ distribution. We observe realisations $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$. The likelihood function $L(\theta)$ is given by

A $L(\theta) = \theta^4(1-\theta)^4$
B $L(\theta) = 13\,\theta^5(1-\theta)^4$
C $L(\theta) = 27\,\theta^8(1-\theta)^4$
D $L(\theta) = 3\,\theta^4(1-\theta)^6$
E $L(\theta) = 12\,\theta\,(1-\theta)^4$

Solution

The PMF of Binomial$(3, \theta)$ is $P(X = k) = \binom{3}{k}\theta^k(1-\theta)^{3-k}$.
The likelihood for independent observations is:

$$L(\theta) = \prod_{i=1}^{4} \binom{3}{x_i} \theta^{x_i}(1-\theta)^{3-x_i}$$

Sum of successes: $1 + 3 + 2 + 2 = 8$. **Sum of failures:** $2 + 0 + 1 + 1 = 4$.
Product of binomial coefficients: $\binom{3}{1}\binom{3}{3}\binom{3}{2}\binom{3}{2} = 3 \times 1 \times 3 \times 3 = 27$.

$$L(\theta) = 27\,\theta^8(1-\theta)^4$$

Answer: C

Q11 – Sample size for a 90% CI with given margin of error

Resit Q11

To estimate the fraction of voters who plan to vote for Candidate A in an election, a random sample of size n from the voters is chosen. The sampling is done with replacement. Let θ be the fraction of voters who plan to vote for Candidate A among all voters. How large does n need to be so that we can obtain a 90% confidence interval with 3% margin of error? That is, how large n needs to be such that $P(U-0.03 < \theta < U+0.03) \geq 0.90$? Here U is the random variable that gives the fraction of voters that plan to vote for candidate A in the sample.
The minimal value of n that ensures the 90% confidence interval with 3% margin of error equals:

A $n = 500$
B $n = 600$
C $n = 700$
D $n = 800$
E $n = 900$

Solution

The margin of error formula for a proportion is $E = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$.
Solving for n and using the worst-case $\hat{p} = 0.5$ (which maximizes the required sample size):

$$n = \frac{z_{\alpha/2}^2 \times \hat{p}(1-\hat{p})}{E^2} = \frac{z_{\alpha/2}^2 \times 0.25}{E^2}$$

For 90% confidence: $z_{0.05} = 1.645$, and $E = 0.03$:

$$n = \frac{1.645^2 \times 0.25}{0.03^2} = \frac{2.706 \times 0.25}{0.0009} = \frac{0.6766}{0.0009} \approx 752$$

Since n must be an integer and we need $n \geq 752$, the smallest option that works is $\boxed{n = 800}$. **Answer: D**

Q12 – Type I error for a geometric hypothesis test

Resit Q12

Suppose $X \sim$ Geometric(θ). We wish to test
$H_0 : \theta = 1/2$ versus
$H_1 : \theta = 0.1$
We reject H_0 if $X \geq 3$. The probability of a type I error equals: (round to 2 decimals).

Solution

Type I error $= P(\text{reject } H_0 \mid H_0 \text{ true}) = P(X \geq 3 \mid \theta = 1/2)$.
For a geometric distribution with parameter $\theta = 1/2$: $P(X = k) = (1/2)(1/2)^{k-1} = (1/2)^k$.

$$P(X \geq 3) = 1 - P(X = 1) - P(X = 2) = 1 - \frac{1}{2} - \frac{1}{4} = \boxed{0.25}$$

Q13 – Poisson tail probability at a non-integer

Resit Q13

Suppose the random variable X has the Poisson distribution with parameter 3. Then $P(X > 1.8)$ equals: (round to 2 decimals)

Solution

Since X takes integer values, $P(X > 1.8) = P(X \geq 2) = 1 - P(X \leq 1)$.

$$P(X = 0) = \frac{e^{-3} \cdot 3^0}{0!} = e^{-3} \approx 0.0498$$

$$P(X = 1) = \frac{e^{-3} \cdot 3^1}{1!} = 3e^{-3} \approx 0.1494$$

$$P(X > 1.8) = 1 - 0.0498 - 0.1494 = \boxed{0.80}$$

Q14 – $E(X^3)$ for a uniform distribution via LOTUS

Resit Q14

Suppose the random variable X is uniformly distributed on the interval $[0, 2]$. Then $E[X^3]$ is given by

A 3
B 1
C 1/2
D 2
E 1/4
F 0.16

Solution

For $X \sim \text{Uniform}[0, 2]$, the PDF is $f(x) = \frac{1}{2-0} = \frac{1}{2}$ for $x \in [0, 2]$.
By LOTUS:

$$E(X^3) = \int_0^2 x^3 \cdot \frac{1}{2}\, dx = \frac{1}{2}\left[\frac{x^4}{4}\right]_0^2 = \frac{1}{2} \times \frac{2^4}{4} = \frac{1}{2} \times \frac{16}{4} = \frac{1}{2} \times 4 = \boxed{2}$$

Answer: D

Q15 – P-value for a left-tailed Z-test (repeated)

Resit Q15

Suppose $X_1, X_2, \ldots X_n$ are independent random variables, each with the $N(\mu, \sigma^2)$-distribution. A researcher is testing the hypothesis $H_0\colon \mu = 25$ *against* $H_1\colon \mu < 25$. The sample mean is 24, the sample size (n) is 36 and the population standard deviation is known to be 6. What is the p-value for this test?
In the options below, Φ denotes the cumulative distribution function of the standard Normal distribution and φ denotes the probability density function of the standard Normal

distribution.

A $\Phi(-1/6)$
B $\Phi(-1)$
C $\varphi(-1/6)$
D $\Phi(-6)$
E $1-\Phi(-1)$
F $1-\Phi(-1/6)$

Solution

Test statistic: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{24-25}{6/\sqrt{36}} = \frac{-1}{1} = -1$.
P-value (left-tailed): $P(Z \leq -1) = \Phi(-1)$. **Answer: B**

Q16 – Chi-square confidence interval for the variance

Resit Q16

A manufacturing company produces bolts with a standard deviation of diameters that is unknown. A random sample of 16 bolts has (sample) variance $s^2 = 0.04$. Assuming the bolt diameters are normally distributed with variance σ^2, construct a 95% confidence interval for σ^2. Choose from below the correct confidence interval. If you believe there is information missing to construct the interval, you can also choose that option.
You can use values from the following table (the rightmost column contains the p-upper quantile of the χ_n-distribution (chi-square distribution with n degrees of freedom).

p	n	$\chi_{p,n}$
0.025	15	27.49
0.05	15	14.34
0.95	15	7.26
0.975	15	6.26
0.025	16	28.85
0.05	16	15.34
0.95	16	7.96
0.975	16	6.91

A $[0.022, 0.096]$
B $[0.021, 0.087]$
C $[0.032, 0.076]$
D $[0.012, 0.136]$
E There is information missing to answer this question.

Solution

The 95% CI for the population variance σ^2 from a normal sample is:

$$\left[\frac{(n-1)\,s^2}{\chi^2_{\alpha/2,\,n-1}}, \quad \frac{(n-1)\,s^2}{\chi^2_{1-\alpha/2,\,n-1}}\right]$$

Identify values: $n = 16$, degrees of freedom $= n - 1 = 15$, $s^2 = 0.04$, $\alpha = 0.05$.
Chi-square quantiles (from the table, $n = 15$): $\chi^2_{0.025,\,15} = 27.49$ and $\chi^2_{0.975,\,15} = 6.26$.

Compute the numerator:

$$(n-1)\,s^2 = 15 \times 0.04 = 0.6$$

Build the interval:

$$\left[\frac{0.6}{27.49},\ \frac{0.6}{6.26}\right] = [0.0218,\ \ 0.0958] \approx \boxed{[0.022,\ 0.096]}$$

Answer: A

Q17 – CDF properties: right-continuity

Resit Q17

Which of the following is a property of any cumulative distribution function (CDF)?

A It is always strictly increasing.
B It can decrease at certain points.
C It is constant over the entire real line.
D If F is a CDF, then the limit of $F(x)$, when x tends to infinity, equals 0.
E It is always right-continuous.

Solution

A: False – a CDF is non-decreasing but can be flat (e.g. between mass points of a discrete distribution).

B: False – CDFs never decrease.

C: False – F goes from 0 to 1.

D: False – $\lim_{x\to\infty} F(x) = 1$, not 0.

E: **True.** Right-continuity ($\lim_{x\downarrow a} F(x) = F(a)$) is a fundamental CDF property.

Answer: E

Q18 – Total probability for three widget types

Resit Q18

A factory produces three types of widgets: W_1, W_2, W_3, with respective probabilities 0.5, 0.3 and 0.2. The probability that a widget is defective is 0.02, 0.05 and 0.1 for widgets 1, 2 and 3 respectively. What is the probability that a randomly selected widget is defective?

A 0.045
B 0.055
C 0.065
D 0.075
E 0.085

Solution

By the **law of total probability**:

$$P(D) = \sum_{i=1}^{3} P(D \mid W_i)\, P(W_i) = 0.02(0.5) + 0.05(0.3) + 0.1(0.2)$$

$$= 0.01 + 0.015 + 0.02 = \boxed{0.045}$$

Answer: A

Q19 – Bayes' rule: breast cancer screening

Resit Q19

Studies show that about one woman in seven who live to be 90 will develop breast cancer. Suppose that of those women who develop breast cancer, a test is negative 2% of the time. Also suppose that in the general population of women, the test for breast cancer is negative about 85% of the time. Let B= {woman develops breast cancer} and let N= {test shows negative result}. Suppose a woman is selected at random.
Given that the woman tests negative, what is the probability that she has breast cancer? Round your answer to 4 decimals.

Solution

Apply **Bayes' rule**: $P(B \mid N) = \frac{P(N|B)\,P(B)}{P(N)}$.
Given: $P(B) = 1/7 \approx 0.1429$, $P(N \mid B) = 0.02$, $P(N) = 0.85$.

$$P(B \mid N) = \frac{0.02 \times \frac{1}{7}}{0.85} = \frac{0.02857}{0.85} \approx 0.00336 \approx \boxed{0.0034}$$

Tip

The result makes intuitive sense: if a woman tests negative and the test is quite reliable (only 2% false negatives), the chance she actually has cancer is very small.

Q20 – Rejection at one level implies rejection at a larger level

Resit Q20

If a hypothesis is rejected at 0.025 level, then

	True	Not true
it must be rejected also at level 0.01		
it must be rejected also at level 0.05		

Solution

"Rejected at level 0.025" means p-value < 0.025.
At level 0.01: We know $p < 0.025$, but we don't know if $p < 0.01$. For example, $p = 0.02$ would mean rejection at 0.025 but not at 0.01. **Not true.**
At level 0.05: Since $p < 0.025 < 0.05$, we can always reject. **True.**

Tip

General rule: if you reject at level α, you also reject at any level $\alpha' > \alpha$, but not necessarily at levels $\alpha' < \alpha$.

Q21 – Conceptual questions about Type II errors

Resit Q21

Which statements are true?

	True	Not true
The probability of a type II error is the probability of not rejecting the null hypothesis when the null hypothesis is true.		
If the rejection region of a test becomes larger, the probability of a type II error gets smaller.		

Solution

Statement 1: Not true. A Type II error is failing to reject H_0 when H_0 is *false*. Not rejecting H_0 when it *is true* is simply a correct decision.
Statement 2: True. A larger rejection region means we reject H_0 more often. When H_0 is false, we are thus more likely to (correctly) reject it, which means the probability of a Type II error (failing to reject a false H_0) decreases. The trade-off is that Type I error probability increases.

Q22 – Type II error and significance levels

Resit Q22

If a hypothesis is rejected at 0.025 level, then

	True	Not true
the probability of a type II error is less than 0.025		
it may or may not be rejected at level 0.01		

Solution

Statement 1: Not true. The Type II error probability β depends on the true parameter value and has no direct relationship to α. One can have $\alpha = 0.025$ and $\beta = 0.5$.
Statement 2: True. We know p-value < 0.025, but the p-value could be anywhere in $(0, 0.025)$. If $p = 0.005$, it would also be rejected at 0.01. If $p = 0.02$, it would not.

Q23 – Computing the MLE from a given likelihood

Resit Q23

Suppose the likelihood function for the parameter θ is given by
$L(\theta) = 12\,\theta^3(1-\theta)^5$.
The maximum likelihood estimate is given by (round to 2 decimals)

Solution

Step 1 – Take the log-likelihood:

$$\ell(\theta) = \ln L(\theta) = \ln 12 + 3\ln\theta + 5\ln(1-\theta)$$

Step 2 – Differentiate and set to zero:

$$\frac{d\ell}{d\theta} = \frac{3}{\theta} - \frac{5}{1-\theta} = 0$$

Step 3 – Solve:

$$\frac{3}{\theta} = \frac{5}{1-\theta} \quad \Rightarrow \quad 3(1-\theta) = 5\theta \quad \Rightarrow \quad 3 = 8\theta \quad \Rightarrow \quad \hat{\theta} = \frac{3}{8} = \boxed{0.38}$$

Tip

For any likelihood of the form $L(\theta) \propto \theta^a(1-\theta)^b$, the MLE is always $\hat{\theta} = \frac{a}{a+b}$.

Q24 – Probability involving a square root transformation

Resit Q24

Suppose X is a random number (so X has the uniform distribution on $(0,1)$). Then $P(\sqrt{X} > 0.5)$ equals (round to 2 decimals)

Solution

Since $\sqrt{X} > 0.5$ is equivalent to $X > 0.25$ (squaring both sides; both are positive):

$$P(\sqrt{X} > 0.5) = P(X > 0.25)$$

For $X \sim \text{Uniform}(0,1)$, $P(X > a) = 1 - a$ for any $a \in [0,1]$:

$$P(X > 0.25) = 1 - 0.25 = \boxed{0.75}$$

Quick Reference: Key Formulas

Probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \text{(Inclusion–exclusion)}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{(Conditional probability)}$$

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \qquad \text{(Bayes' rule)}$$

$$P(B) = \sum_i P(B \mid A_i)\, P(A_i) \qquad \text{(Total probability)}$$

Discrete Distributions

Distribution	PMF	$E(X)$	$\mathrm{Var}(X)$
$\mathrm{Bin}(n,p)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
$\mathrm{Pois}(\lambda)$	$e^{-\lambda}\lambda^k/k!$	λ	λ
$\mathrm{Geom}(p)$	$p(1-p)^{k-1}$	$1/p$	$(1-p)/p^2$

Continuous Distributions

Distribution	PDF	$E(X)$	$\mathrm{Var}(X)$
$\mathrm{Unif}(a,b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
$N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2

Confidence Intervals

$$\text{Mean (known } \sigma\text{):} \quad \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{1}$$

$$\text{Proportion:} \quad \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{2}$$

$$\text{Variance:} \quad \left[\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}},\ \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right] \tag{3}$$

Common z-values: $z_{0.025} = 1.96$ (95%), $z_{0.005} = 2.58$ (99%), $z_{0.05} = 1.645$ (90%).

Hypothesis Testing

$$\text{Known } \sigma\text{: } Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad \text{Unknown } \sigma\text{: } T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Maximum Likelihood

For $L(\theta) \propto \theta^a(1-\theta)^b$: $\hat{\theta}_{\mathrm{MLE}} = \dfrac{a}{a+b}$.

Good luck with your exam!