

MATH40005

Probability and Statistics

Revised Mock Examination 1

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

1(a). A box contains cards labelled $1, 2, \dots, 8$. One card is drawn, replaced, and a second card is drawn. Briefly describe a probability space $(\Omega, \mathcal{F}, P)$ when all cards are equally likely. (3 marks)

1(b)(i). A password has 10 characters: 5 letters chosen from 52 letters with repetition allowed, 3 digits chosen from 10 digits with repetition allowed, and 2 distinct symbols chosen from 5 symbols. Count all possible passwords. (5 marks)

1(b)(ii). Assuming all pairs of symbol positions are equally likely, find the probability that the two symbols are adjacent. (4 marks)

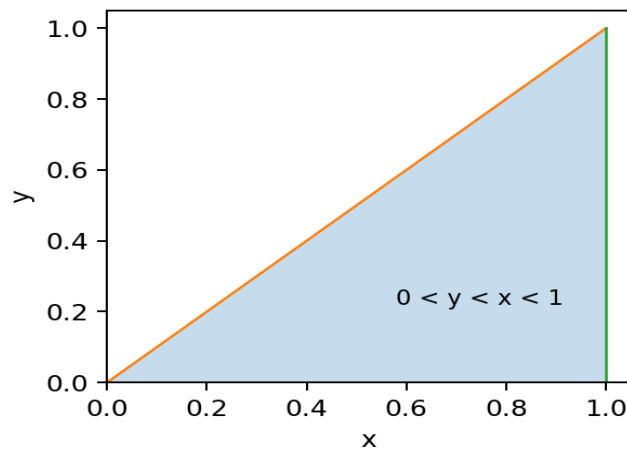
1(c). Let r be an integer and $0 < p < 1$. A random variable has pmf $p_X(x) = p(1-p)^{(x-r)}$ for $x = r, r+1, \dots$ and zero otherwise. Verify that it is a valid pmf and find its CDF for every real x . (5 marks)

1(d). There are n independent components, each failing on a specified day with probability $1/d$. Let Y_i indicate failure of component i and $B = \sum Y_i$. Identify the distributions of Y_i and B , including parameters. (3 marks)

Total: 20 marks

Question 2

2(a). For $0 < y < x < 1$ let $f_{XY}(x,y)=c(x+y)$, and let it be zero otherwise. Find c . (3 marks)



2(b). Find the marginal densities of X and Y . (4 marks)

2(c). Find the conditional density of Y given $X=x$, and state where it is defined. (3 marks)

2(d). Are X and Y independent? Give a decisive reason. (2 marks)

2(e). Compute $\text{Cov}(X,Y)$. (5 marks)

2(f). Let $R=Y/X$. Find the CDF of R . (3 marks)

Total: 20 marks

Question 3

3(a). Let X have finite first and second moments. State the minimizer of $E[(X-a)^2]$, the minimum value, and a minimizer of $E|X-b|$. (3 marks)

3(b). Prove the bias-variance decomposition for an estimator $\hat{\theta}$ of a fixed parameter θ . (3 marks)

3(c). Fifteen independent observations are assumed normal with unknown mean and variance. The sample mean is 74.2 and sample variance is 6.25. Compute a 95% confidence interval for the mean. Use $t_{(14,0.975)}=2.145$. (4 marks)

3(d). Using the same data, test $H_0: \mu=75$ against $H_1: \mu \neq 75$. Compute a two-sided p-value and state the conclusion at level 0.05. (5 marks)

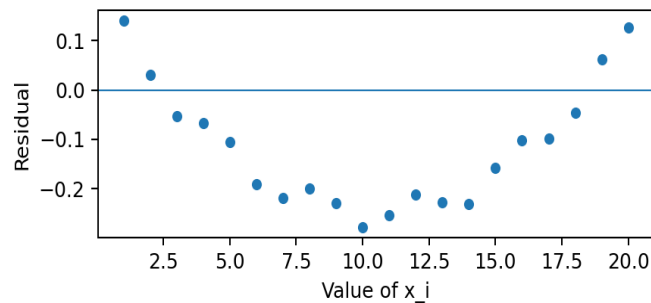
3(e). Define a p-value, explain how it is interpreted, and state when it is $\text{Uniform}(0,1)$. (5 marks)

Total: 20 marks

Question 4

4(a). Suppose $X \sim N(5, 9)$, $Y \sim N(4, 4)$, and $\text{Corr}(X, Y) = -1/3$. Let $Z = X - 2Y$. Compute $\text{Corr}(X, Z)$. (4 marks)

4(b). Let X and Y have equal finite variances. Prove that $X + Y$ and $X - Y$ are uncorrelated. (5 marks)



4(c). A simple linear regression has $R^2 = 0.94$ and the residual plot shown. Does the model fit well? Suggest a specific improvement. (4 marks)

4(d). Independent observations are modeled as $\text{Uniform}[\theta, \theta + 2]$. Assuming $\max(x_i) - \min(x_i) \leq 2$, find the full set of maximum likelihood estimates of θ . (5 marks)

4(e). For original data $\{12, 18, 21, 27, 30\}$, decide whether (i) $\{18, 18, 30, 12, 27\}$ and (ii) $\{12, 18, 21, 27\}$ are bootstrap samples. Justify. (2 marks)

Total: 20 marks

MATH40005

Probability and Statistics

Revised Mock Examination 2

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

- 1(a). Passengers use routes A,B,C with probabilities 0.5,0.3,0.2. Their late probabilities are 0.04,0.10,0.05 respectively. Find the probability that a randomly selected passenger is late. (5 marks)
- 1(b). Given that the passenger is late, find the probability that route A was used. (3 marks)
- 1(c). A signal failure now makes every train on routes B and C late, while route A keeps late probability 0.04. Find $P(A|L)$ under the new conditions. (4 marks)
- 1(d). A 10-character password contains 5 letters (52 choices, repetition allowed), 3 digits (10 choices, repetition allowed), and 2 distinct symbols from 5 symbols. Count all passwords. (5 marks)
- 1(e). State the most suitable plot for each purpose: (i) the relationship between house price and number of bedrooms; (ii) median, spread and outliers of prices in one area; (iii) daily stock prices over three years. (3 marks)

Total: 20 marks

Question 2

2(a). Let $U \sim \text{Uniform}(0,1)$ and $\lambda > 0$. Identify the distribution of $T = -\log(U)/\lambda$ by deriving its CDF. (4 marks)

2(b). For independent $X_1, \dots, X_n \sim \text{Uniform}(0,1)$, find the CDF and density of $M = \max(X_1, \dots, X_n)$. (4 marks)

2(c). Let X be standard normal. Derive its moment generating function. (4 marks)

2(d). If $X_1, \dots, X_n$ are independent standard normal variables, use MGFs to identify $S = \sum X_i$. (4 marks)

2(e). Suppose $E[U^k] = E[V^k]$ for every non-negative integer k . State a sufficient additional condition under which U and V must have the same distribution, and justify. (4 marks)

Total: 20 marks

Question 3

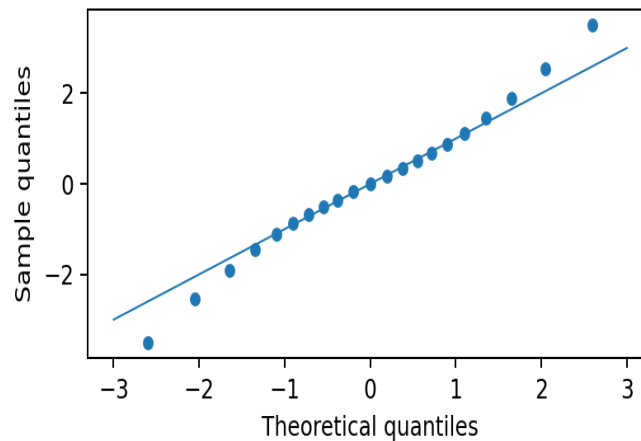
3(a). Independent $X_i \sim N(\mu_i, \sigma^2)$ may have different means. For $\bar{X} = n^{-1} \sum X_i$, prove $\text{Var}(\bar{X}) = \sigma^2/n$. (2 marks)

3(b). With $S^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ and $\bar{\mu} = n^{-1} \sum \mu_i$, derive $E[S^2]$. (6 marks)

3(c). For $n=12$ normal observations with $\bar{x}=181$ and $s^2=3$, compute a 90% confidence interval for μ . Use $t_{(11,0.95)}=1.796$. (4 marks)

3(d). Using the same sample, compute a 95% confidence interval for σ^2 . Use $\chi^2_{(11,0.025)}=3.816$ and $\chi^2_{(11,0.975)}=21.920$. (4 marks)

3(e). Interpret the normal Q-Q plot shown. Does it support a normal model? (4 marks)



Total: 20 marks

Question 4

4(a). Independent observations are modeled as $\text{Uniform}[\theta_1, \theta_2]$, where $0 < \theta_1 < \theta_2$ are unknown and the observed sample is not constant. Find the MLEs. (5 marks)

4(b). In simple linear regression with fixed x_i and independent errors of variance σ^2 , prove $S_{xy} = \sum (x_i - \bar{x})y_i$ and derive $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \bar{x} / S_{xx}$. (4 marks)

4(c). Describe how to construct a percentile bootstrap 95% confidence interval for a population median from an observed sample of size n . (4 marks)

4(d). Rugby and basketball heights are independent normal samples. The population variances are known: $\sigma_1^2 = 10$ and $\sigma_2^2 = 4$. Data give $\bar{x} = 182, n = 15$ and $\bar{y} = 185, m = 12$. Test $H_0: \mu_1 = \mu_2$ against a two-sided alternative at $\alpha = 0.01$, compute the p-value, and conclude. (7 marks)

Total: 20 marks

MATH40005

Probability and Statistics

Revised Mock Examination 3

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

1(a). Assume each birthday week is independent and uniformly distributed over 52 weeks. For a class of 20 pupils, find the probability that at least two share a birthday week. (5 marks)

1(b). Assume birthdays are independent and uniform over 365 days. For $3 \leq n \leq 300$, find the probability that exactly one pair shares a birthday and all other birthdays are distinct. (5 marks)

1(c). A sick child has flu with probability 0.95 and measles with probability 0.05. A rash occurs with probability 0.01 under flu and 0.90 under measles. Given a rash, find the probability of measles. (5 marks)

1(d). Let X and Y be discrete random variables and A an event. Define $Z=X$ on A and $Z=Y$ on A^c . Prove that Z is a discrete random variable. (5 marks)

Total: 20 marks

Question 2

2(a). Let $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$, but do not assume independence. Must $X+Y$ be Poisson? Give a proof or counterexample. (4 marks)

2(b). Let $N \sim \text{Poi}(\lambda)$, independent of i.i.d. $X_i \sim \text{Uniform}(0,1)$. For $0 < x < 1$, find $P(\min\{X_1, \dots, X_{N+1}\} > x)$. (6 marks)

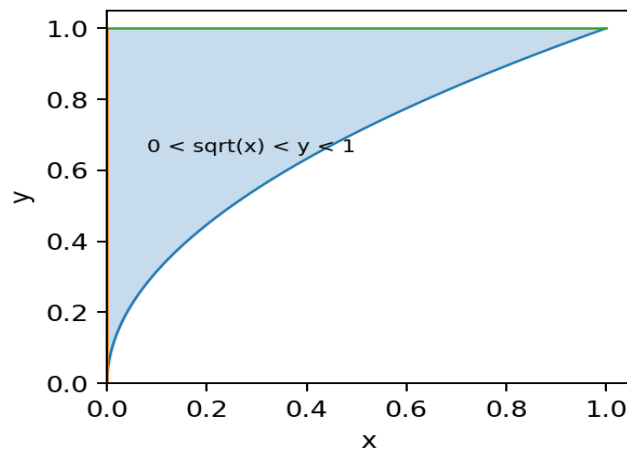
2(c). Let X, Y be independent standard normal variables. Find $P(X > 0, Y > -X)$. (5 marks)

2(d). Calls arrive as $N \sim \text{Poi}(\lambda)$. Independently, each call is labelled urgent with probability p and routine otherwise. Let U and R be the counts. Show that $U \sim \text{Poi}(\lambda p)$, $R \sim \text{Poi}(\lambda(1-p))$, and that U and R are independent. (5 marks)

Total: 20 marks

Question 3

3(a). For $f_{XY}(x,y)=c(x+y)$ on $0 < \sqrt{x} < y < 1$ and zero otherwise, find c , both marginal densities, and decide whether X and Y are independent. (8 marks)



3(b). For $0 < x < 1$, find the conditional CDF $P(Y \leq y | X=x)$, including all cases in y . (5 marks)

3(c). Let $X \sim \text{Uniform}(1,3)$ and $W = X^4$. Find the CDF and density of W , and compute $E[W]$. (7 marks)

Total: 20 marks

Question 4

4(a). Let $N \sim \text{Poi}(\lambda)$, independent of i.i.d. Y_i with mean m and variance v . For $S = \sum_{i=1}^N Y_i$ (zero when $N=0$), derive $E[S]$ and $\text{Var}(S)$. (8 marks)

4(b). Twelve glasses contain six cola and six other drinks. A person selects six as cola and gets five correct. Under random guessing, perform a one-sided exact test of ability at level 0.05. (5 marks)

4(c). A normal population has known standard deviation 6. A random sample of $n=36$ has $\bar{x}=102$. Construct a 95% confidence interval for μ . (4 marks)

4(d). Explain the frequentist meaning of a 95% confidence interval after it has been computed. (3 marks)

Total: 20 marks

MATH40005

Probability and Statistics

Revised Mock Examination 4

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

1(a). Define a sigma-algebra $\mathcal{F}$ on a sample space Ω . (3 marks)

1(b). Define a probability measure P on $(\Omega, \mathcal{F})$. (3 marks)

1(c). How many ordered n -tuples of positive integers sum to k ? State all cases. (4 marks)

1(d). Prove algebraically that $n 4^{n-1} = \sum_{k=0}^n C(n, k) 3^k 1^{n-k}$. (4 marks)

1(e). Give a story proof of the identity in part (d). (6 marks)

Total: 20 marks

Question 2

2(a). A discrete variable has $P(X=1)=1/2$, $P(X=3)=1/3$, $P(X=5)=1/6$. Write its CDF for every real x . (4 marks)

2(b). Let $X \sim \text{Geom}(p_1)$ and $Y \sim \text{Geom}(p_2)$ on $\{1, 2, \dots\}$, independently. Find the distribution of $\min(X, Y)$. (6 marks)

2(c). A fair coin is tossed until the pattern HT first appears. Find the expected number of tosses. (6 marks)

2(d). Prove the law of total expectation for a discrete X and a countable partition $\{B_i\}$ with $P(B_i) > 0$, assuming absolute convergence. (4 marks)

Total: 20 marks

Question 3

3(a). Let X_i be i.i.d. with mean θ and variance σ^2 . As an estimator of θ define $T = (n+1)^{-1} \sum X_i$. Find its bias, variance and MSE. (6 marks)

3(b). Prove Markov's inequality for a non-negative random variable X . (4 marks)

3(c). An unknown Bernoulli proportion p is estimated by the sample proportion. Using Chebyshev, find n sufficient for $|\hat{p} - p| < 0.02$ with confidence at least 0.95 uniformly in p . (5 marks)

3(d). One hundred hypotheses are tested with family-wise level 0.01. Under Bonferroni, which p-values among 0.03, 0.009, 0.0004, 0.00002, 0.000005 are significant? (5 marks)

Total: 20 marks

Question 4

4(a). For data (x_i, y_i) , derive the least-squares estimates in $y_i = \beta_0 + \beta_1 x_i + e_i$. (6 marks)

4(b). For i.i.d. observations from $\text{Uniform}[0, \theta]$, $\theta > 0$, find the MLE of θ . (5 marks)

4(c). Let $X_i | \theta$ be independent $\text{Exponential}(\text{rate } \theta)$ and $\theta \sim \text{Gamma}(a, b)$ in rate parameterization. Derive the posterior. (4 marks)

4(d). Suppose $E[X]=2$, $E[Y]=3$, $E[XY]=4$ and $2 \leq Y \leq 5$. Obtain a non-trivial lower bound on $\text{sd}(X)$. (5 marks)

Total: 20 marks

MATH40005

Probability and Statistics

Revised Mock Examination 5

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

1(a). For $f:U \rightarrow V$ and A subset V , prove $I_A(f(u)) = I_{f^{-1}(A)}(u)$. (2 marks)

1(b). If G is a sigma-algebra on V , prove $f^{-1}(G) = \{f^{-1}(A) : A \in G\}$ is a sigma-algebra on U . (6 marks)

1(c). Construct three Bernoulli random variables that are pairwise independent but not mutually independent. (4 marks)

1(d). Let $\Omega = \{1,2,3,4\}$ and $F = \{\emptyset, \Omega, \{1,2\}, \{3,4\}\}$. Define $X(1)=X(3)=0$ and $X(2)=X(4)=1$. Is X measurable? Give a non-constant measurable random variable on the same space. (4 marks)

1(e). A fair six-sided die is rolled seven times. Give a probability space and the cardinality of its power set. (4 marks)

Total: 20 marks

Question 2

2(a). Let $X \sim N(0,1)$, $Y \sim \text{Gamma}(\alpha, \beta)$ in rate form, independently, and $Z = \sqrt{Y}X$. Find the MGF of Z and its domain. (7 marks)

2(b). Let X be continuous with density f_X and let independent Y take values $1, \dots, 10$. For $Z = XY$, derive f_Z and verify it is a density. (6 marks)

2(c). A joint density is constant c on $0 < x < y < z < 1$. Find c and $E[XYZ]$. (7 marks)

Total: 20 marks

Question 3

3(a). Ten cups contain five tea and five coffee. A person selects five cups as tea and identifies four correctly. Under random guessing, perform a one-sided exact test at level 0.05. (5 marks)

3(b). For the data {14.3, 8.3, 15.9, 12.7, 9.4, 24.0, 7.2, 6.1, 11.6, 10.5}, compute the lower and upper quartiles using the median-of-halves convention, then apply Tukey's rule for outliers. (4 marks)

3(c). Combined data favor a placebo, but within each of two risk groups a drug has a lower disease rate. Name the phenomenon and explain how it can occur. (3 marks)

3(d). Potato weights have unknown distribution and only the bound $\sigma < 10$ g is known. Using the bound 10 g to construct a distribution-free 99% Chebyshev interval, find the minimum integer n that guarantees its total width is strictly less than 4 g. (5 marks)

3(e). Define Type I and Type II errors and explain qualitatively how increasing α affects them. (3 marks)

Total: 20 marks

Question 4

- 4(a). Independent positive observations have lognormal density proportional to $\exp[-(\log x - \theta)^2 / (2\sigma^2)]/x$, where σ^2 is known. Find the MLE of θ . (6 marks)
- 4(b). Show that the fitted least-squares line in simple linear regression passes through $(\bar{x}, \bar{y})$. (4 marks)
- 4(c). Suppose all non-negative integer moments of U and V agree. State a sufficient condition for equality in distribution and justify. (4 marks)
- 4(d). Describe a percentile bootstrap 95% confidence interval for a median. (3 marks)
- 4(e). For i.i.d. observations with mean μ and variance σ^2 , construct an unbiased estimator of μ^2 using $\bar{X}$ and the unbiased sample variance S^2 . (3 marks)

Total: 20 marks

MATH40005

Probability and Statistics

Revised Mock Examination 6

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

1(a). Prove by double counting that $n 5^{n-1} = \sum_{k=0}^n C(n,k) 4^k 1^{n-k}$. Give both an algebraic and a story interpretation. (6 marks)

1(b). A test has sensitivity 0.95 and false-positive rate 0.05. Compute $P(\text{disease}|\text{positive})$ when prevalence is (i) 0.01 and (ii) 0.10, and explain the change. (6 marks)

1(c). Assume independent uniform birthdays over 365 days. Find the probability that among 12 people exactly one pair shares a birthday and all others are distinct. (5 marks)

1(d). A fair four-sector spinner is spun and then a fair coin is tossed. Describe a probability space. (3 marks)

Total: 20 marks

Question 2

2(a). For $f_{XY}(x,y)=c x(y+1)$ on $0<x<y<1$, find c , $f_X(x)$, and $E[Y|X=x]$, stating the domain. (8 marks)

2(b). Let $N \sim \text{Poi}(\lambda)$, independent of i.i.d. Y_i with mean m and variance v . Derive the mean and variance of $S = \sum_{i=1}^N Y_i$. (7 marks)

2(c). Let $M = \max(X_1, \dots, X_n)$ for i.i.d. $\text{Uniform}(0,1)$, and define $T = -\log M$. Identify the distribution of T . (5 marks)

Total: 20 marks

Question 3

3(a). Independent $X_i \sim N(\mu_i, \sigma^2)$ may have unequal means. Derive $E[S^2]$ for the usual sample variance. (6 marks)

3(b). Football and basketball heights are independent normal samples with equal unknown variance. Data: $n=16$, $\bar{x}=176.5$, $s_x^2=4$; $m=9$, $\bar{y}=181$, $s_y^2=4$. Test $H_0: \mu_1 + \mu_2 = 360$ against a two-sided alternative at 0.05, giving a p-value. (7 marks)

3(c). For $n=12$ normal observations with $s^2=3$, compute a 99% confidence interval for σ^2 . Use $\chi^2_{11,0.005}=2.603$ and $\chi^2_{11,0.995}=26.757$. (4 marks)

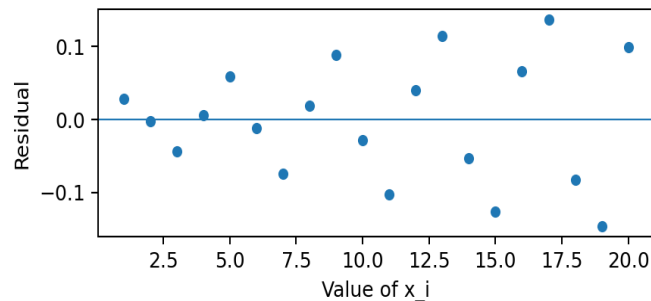
3(d). Define a p-value and prove that an exact upper-tail p-value is $\text{Uniform}(0,1)$ under H_0 when the test statistic has a continuous null CDF F_0 . (3 marks)

Total: 20 marks

Question 4

4(a). For the no-intercept model $y_i = \beta x_i^2 + e_i$, derive the least-squares estimate of β . (5 marks)

4(b). The residual plot shown comes from a fitted straight-line model. Diagnose the problem and give a remedy. (3 marks)



4(c). Independent observations are $\text{Uniform}[\theta, \theta+1]$. Find the full set of MLEs for θ . (5 marks)

4(d). Let $X_i | \theta$ be independent geometric variables with $P(X_i = x) = \theta(1-\theta)^{x-1}$, and $\theta \sim \text{Beta}(a, b)$. Derive the posterior. (4 marks)

4(e). Two unrelated increasing time series have sample correlation 0.97. State what can and cannot be concluded. (3 marks)

Total: 20 marks