Brown University
DATA 1010
Fall 2018: Midterm II
Samuel S. Watson

Name:

You will have three hours to complete the exam, which consists of 24 questions. Among the first 12 questions, you should only solve problems for standards for which you want to improve your score from the first exam. If you are completing the problem with standard key [JULIA], you will hand in your answers to the written portion and then get out your laptop and implement the solution in Julia. You will be able to submit your answer to that question directly using Gradescope.

For the written part of the exam, no calculators or other materials are allowed, except the Julia-Python-R reference sheet and the provided exam reference sheet. For the computational part of the exam, you may use any internet technologies which do not involve active communication with another person.

You are responsible for explaining your answer to **every** *question. Your explanations do not have to be any longer than necessary to convince the reader that your answer is correct.*

For questions with a final answer box, please write your answer as clearly as possible and strictly in accordance with the format specified in the problem statement. Do not write anything else in the answer box. Your answers will be grouped by Gradescope's AI, so following these instructions will make the grading process much smoother.

I verify that I have read the instructions and will abide by the rules of the exam: ____________________________

Problem 1 [SETFUN]

Which of the following is true for all functions f and subsets A and B of f's domain? Write your answer in the box as (a) or (b).

(a) $f(A \cap B) \stackrel{?}{=} f(A) \cap f(B)$

(b) $f(A \cup B) \stackrel{?}{=} f(A) \cup f(B)$

Solution

Final answer:

Problem 2 [LINALG]

(a) Suppose that S and T are linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ with the property that the range of S is not $\mathbb{R}^n$. Show that the null space of $T \circ S$ contains more than one vector.

(b) Suppose that $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ are vectors in $\mathbb{R}^5$. Is it possible that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent, $\{\mathbf{v}_1, \mathbf{v}_3\}$ is linearly independent, $\{\mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent, and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent?

Solution

Problem 3 **[MATALG]**

(a) Find a solution to the matrix equation

$$\begin{bmatrix} 1 & 4 & -2 \\ 2 & 0 & -4 \\ 3 & 7 & -6 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix}. \tag{3.1}$$

(b) Explain why (3.1) cannot be solved by left-multiplying by the inverse of the coefficient matrix:

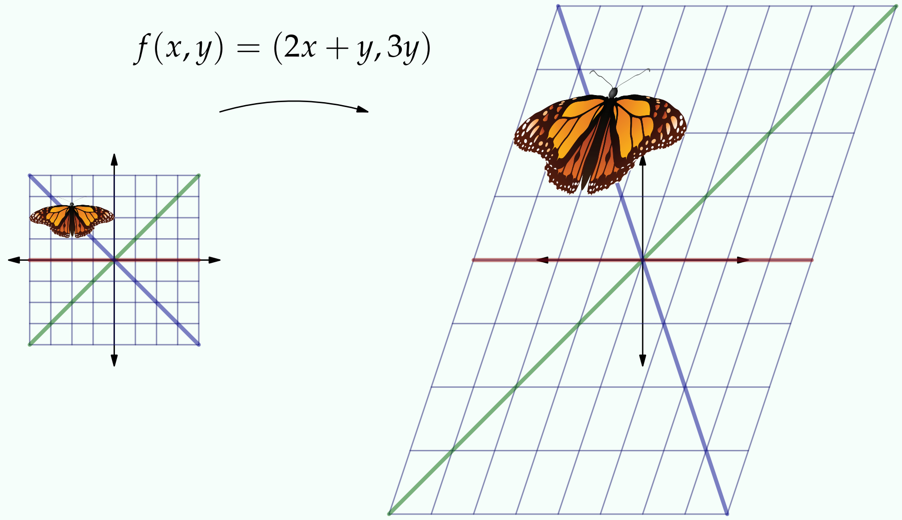
$$\mathbf{x} \stackrel{?}{=} \begin{bmatrix} 1 & 4 & -2 \\ 2 & 0 & -4 \\ 3 & 7 & -6 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix}.$$

(c) Identify which component of $\mathbf{x}$ is necessarily zero for any solution $\mathbf{x}$ of (3.1). Give a geometric explanation for why that component cannot be nonzero.

Solution

Problem 4 **[EIGEN]**

(a) Identify two line segments in the figure below whose ratio of lengths is equal to one of the eigenvalues of the transformation depicted.



(b) Identify two line segments in the below whose ratio of lengths is equal to the *other* eigenvalue of the transformation depicted.

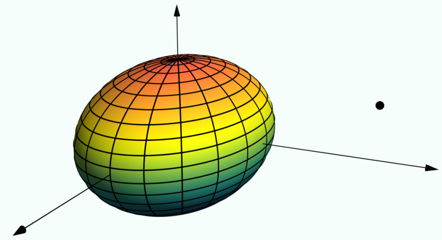
(c) Which colored line in the figure does not contain an eigenvector of the transformation?

Solution

Problem 5 **[OPT]**

(a) Describe an algorithm, using gradient descent, to numerically approximate the point on the ellipsoid $x^2 + 2y^2 + 3z^2 = 1$ which is closest to the point $(2, 2, 1)$. Explain in enough detail that it would be relatively straightforward to translate your description to a working program.

(b) Do you expect gradient descent to work well for this problem? Why or why not?



Solution

Problem 6 **[MATDIFF]**

Find a formula (in terms of A and $\mathbf{b}$) for the vector in the range of an $m \times n$ matrix A which is as close to $\mathbf{b} \in \mathbb{R}^m$ as possible. You may assume that $\operatorname{rank}(A) = n$.

Solution

Final answer:

Problem 7 [MACHARITH]

What value does the function below return? Note that all operations are performed in `Float64` arithmetic.

```
function f()
    x = 1 - 0.5^48
    for i in 1:2^20
        x += 0.5^53
    end
    x
end
```

Solution

Final answer:

Problem 8 [PRNG]

Consider a random number generator which is based on detecting gamma rays entering the earth's atmosphere. Explain why this is **not** a pseudorandom number generator, and discuss some salient differences between this random number generator and a PRNG.

Solution

Problem 9 **[NUMERROR]**

(a) Explain why the function $f(x) = x - 5$ is not well-conditioned everywhere even though its derivative (in other words, its error magnification factor) is equal to 1 everywhere.

(b) Many scientific computing environments include a function `expm1`, which calculates the mathematical function $f(x) = e^x - 1$. Using what you know about numerical error, explain why we would want to have a separate function instead of just using `exp(x)-1.0`. Correctly use the terms 'unstable' and 'well-conditioned' in your explanation.

Solution

Problem 10 **[COUNTING]**

Suppose you roll a die, and if the result is prime, then you roll the die again 2 more times. If the result is not prime, then you roll the die again 3 more times. You record the dice results as a tuple (for example, the result might be $(2, 4, 5)$ or $(1, 6, 6, 4)$). How many elements are in the set of all possible outcomes for this experiment?

Solution

Problem 11 **[PROBSPACE]**

(a) Draw a Venn diagram generically representing the statement "C occurs only if A and B both occur", where A, B, and C are events.

(b) Suppose that A and B each have probability $\frac{1}{3}$ and that the probability that *neither* A nor B occurs is $\frac{1}{2}$. Find the greatest possible value of $\mathbb{P}(C)$, assuming that C occurs only if A and B both occur.

Put your answer to (b), expressed as a reduced fraction, in the box.

Solution

Final answer:

Problem 12 **[JULIA]**

Write a Julia function `somesorted` which accepts a vector of vectors as an argument and returns `true` if at least one of the vectors is sorted (each entry but the last is less than or equal to the next). Write a helper function `is_sorted` which determines whether a single vector is sorted.

```
@assert somesorted([[9,7,1],[4,3,6],[3]]) == true
@assert somesorted([[3,4,1],[1,3,8],[1,10]]) == true
@assert somesorted([[-1,0,0,4],[5,4]]) == true
@assert somesorted([[4,3],[-1,0,-2]]) == false
```

Solution

Problem 13 **[PMF]**

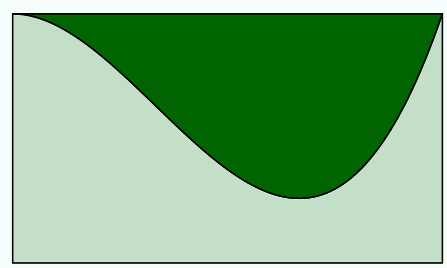
Suppose that X is uniformly distributed on $\{1, 2, 3, 4\}$ and Y is an independent random variable which is uniformly distributed on $\{10, 20, 30\}$. Sketch a graph of the probability mass function of $X + Y$.

Solution

Problem 14 **[PDF]**

A random variable X is obtained in the following way: we sample a point uniformly from the rectangle shown. If the point falls in the dark region, we define X to be the x-coordinate of that point. If it is in the light green region, we reject it and try again (rejecting as many times as necessary until we get a point in the dark green region).

Sketch the graph of the probability density function of X. (Don't worry about normalizing—just get the shape approximately right.)



Solution

Problem 15 **[CONDPROB]**

Suppose that a permutation of $\{1,2,3,4,5\}$ is selected uniformly at random from the set of all 120 permutations of those digits. What is the conditional probability that the permutation has 3 in the third position given that it has 4 in the fifth position? For example, the permutation $(2,1,3,5,4)$ is in the event in question, but the permutation $(3,1,5,2,4)$ is not. Express your answer as a fraction in this box.

Solution

Final answer:

Problem 16 **[BAYES]**

On days when it rains, Marvin brings his umbrella to work with probability 60% and Jean brings her umbrella to work with probability 70%. On days when it does not rain, Marvin brings his umbrella to work with probability 5% and Jean brings her umbrella to work with probability 10%.

(a) Based on the numbers above, who would you guess brings their umbrella to work more often? (no wrong answers here, just speculate)

(b) Jean lives in San Diego, where it rains on 10% of the days of the year. Marvin lives in Seattle, where it rains 25% of the days of the year. What is Jean's probability of bringing her umbrella to work on a given day? What is Marvin's probability of bringing his umbrella to work on a given day?

(c) Conditioned on the event that Marvin brought his umbrella to work today, what is the probability that it rained?

Put your answer to (c), written as an integer followed by a percent sign, in the final answer box.

Solution

Final answer:

Problem 17 **[IND]**

Suppose that X is a fair die roll (that is, its distribution is uniform on $\{1, 2, 3, 4, 5, 6\}$). Suppose Y is a random variable whose distribution has a probability mass of $1/3$ at 1 and a probability mass of $2/3$ at 2.

(a) Plot the probability mass function of the joint distribution X and Y, assuming that they are independent (using dot size to indicate the amount of mass at each point).

(b) Plot a joint PMF for which X and Y have the marginal distributions specified above but are *not* independent. Use text labels if you want to more clearly identify the amount of mass at each point, and explain the ideas you used to come up with your plot.

Solution

Problem 18 **[EXP]**

(a) Suppose that X and Y are random variables which are not independent but which have the same distribution. Is it necessarily true that $\mathbb{E}[X] = \mathbb{E}[Y]$? Is it necessarily true that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$?

(b) Kevin Durant scores 26.4 points per game on average, and Stephen Curry scores 26.2 points per game on average. Their respective point totals in a given game are negatively correlated, since they are teammates. If C is the number of points scored by Curry and D is the number of points scored by Durant, the covariance matrix of $[C, D]$ is

$$\begin{bmatrix} 45.7 & -14.7 \\ -14.7 & 72.2 \end{bmatrix}$$

Let the random variable X be the total number of points scored by Durant and Curry in the next game. Find $\mathbb{E}[X]$ and write your answer (expressed as a decimal) in the box.

Solution

Final answer:

Problem 19 **[COV]**

(a) Draw a probability mass function for the joint distribution of two random variables which have zero covariance but are not independent.

(b) Suppose that X_1, X_2, X_3, X_4 are independent random variables each of which has mean zero and variance 3. Suppose that $\mathbb{E}[X_i X_j] = 2$ for all $i \neq j$ (that is, any pair of distinct random variables from the list has covariance 2). Find

$$\mathbb{E}[(X_1 + X_2 + X_3 + X_4)^2]$$

and write your answer in the box.

Solution

Final answer:

Problem 20 **[CONDEXP]**

Suppose that (X, Y) is uniformly distributed on the upper unit half disk (that is, the set of points above the x-axis whose distance from the origin is less than 1).

(a) Find the conditional distribution of Y given $X = x$.

(b) Find $\mathbb{E}[Y \mid X]$, and write your answer in the box.

Solution

Final answer:

Problem 21 **[COMDISTD]**

Suppose that $S = X_1 + \ldots + X_{40}$, where the X_i's are independent $\text{Bin}(10, 0.02)$ random variables.

(a) The distribution of S is a named distribution. Which one is it, and what are the parameters?

(b) The distribution of S is approximately Poisson with parameter λ for which value of λ? Write your answer in the box.

Solution

Final answer:

Problem 22 **[COMDISTC]**

(a) Suppose that X and Y are random variables whose joint distribution has density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{3}} \mathrm{e}^{-\frac{1}{2}[x+3 \;\; y+4]\left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]^{-1}\left[\begin{smallmatrix} x+3 \\ y+4 \end{smallmatrix}\right]}.$$

Find the distribution of Y.

(b) Find the correlation between X and Y and write your answer (expressed as a fraction) in the box.

Solution

Final answer:

Problem 23 **[RVINEQ]**

Show that the inequality

$$\mathbb{P}(|X - \mu| \geq k\sigma) < 1/k^2$$

is **not** true in general (where X is a finite-variance random variable, μ is its mean, σ is its standard deviation, and k is an arbitrary positive number). Hint: try a Bernoulli random variable with parameter $p = \frac{1}{2}$.

Solution

Problem 24 **[CLT]**

Suppose that S is a binomial random variable with parameters n and $p = \frac{1}{3}$. Assuming that n is extremely large, sort the following quantities from least to greatest.

$$A = \mathbb{P}(S_n \leq n/3 + \sqrt{n})$$
$$B = \mathbb{P}(S_n \geq 0.33n)$$
$$C = \mathbb{P}(S_n < 0)$$
$$D = \mathbb{P}(S_n < 10^4)$$

Express your answer in the box as a sequence of adjacent letters with no separation symbol (like ABCD).

Solution

Final answer:

Additional space

Additional space

Additional space

Additional space

Additional space