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

To show that (a) $f(A \cup B) = f(A) \cup f(B)$, we note that if $y \in f(A \cup B)$, then $y = f(x)$ for some $x \in A \cup B$. This element x is in either A or B, which means that $y = f(x)$ is in either $f(A)$ or $f(B)$. Thus $y \in f(A) \cup f(B)$. So $f(A \cup B) \subset f(A) \cup f(B)$. Similar reasoning shows that $f(A) \cup f(B) \subset f(A \cup B)$ as well.

To see that (b) may fail, consider the function from $\{-1, 0, 1\}$ to $\{0, 1\}$ which squares its input. Let $A = \{-1, 0\}$ and $B = \{0, 1\}$. Then $f(A \cap B) = \{0\}$, while $f(A) \cap f(B) = \{0, 1\}$.

Final answer:

(a)

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

(a) Since S does not have rank n, the rank-nullity theorem tells us that its null space has dimension at least one. And if S maps $\mathbf{v}$ to zero, the $T \circ S$ maps $\mathbf{v}$ to zero as well (since every linear transformation maps the zero vector to itself).

(b) Yes, it's possible. In fact, this happens whenever the three vectors point in three different directions in the same plane.

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

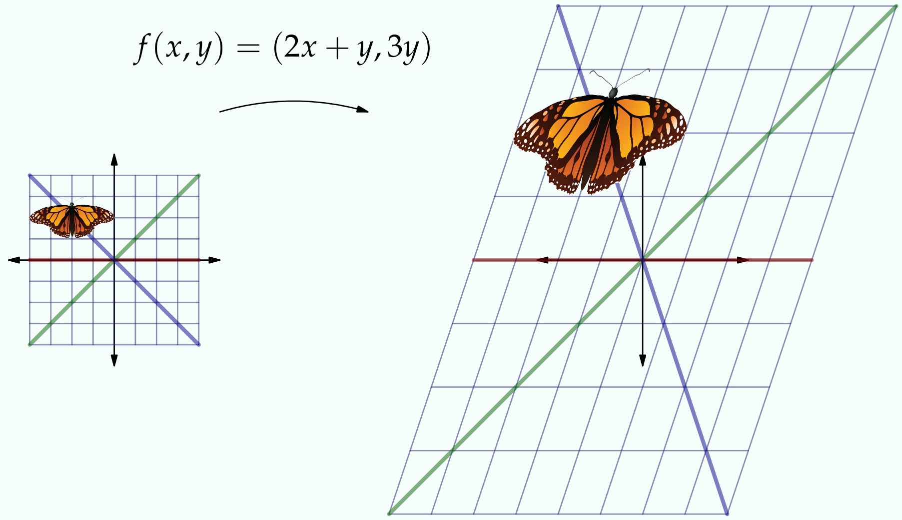
(a) One solution is $[5, 0, 0]$, since the first column times 5 is equal to the vector on the right-hand side.

(b) The matrix does not have an inverse, since its first and last columns are collinear.

(c) The middle component is necessarily zero. This is because the first and last columns are collinear. Geometrically, we can see that a nonzero weight for the second column would cause the corresponding linear combination of the columns of the matrix to have a component which is not in the $[1, 2, 3]$ direction. Since the other two vectors do point in the $[1, 2, 3]$ direction, no linera combination of them would be able to cancel this wrong-direction component.

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

(a) Let L_2 be the line segment which extends due east from the origin to the first slant grid line in the codomain picture, and similarly for L_1 in the domain picture. Then the ratio of the length of L_2 to the length of L_1 is equal to one of the eigenvalues of the transformation.

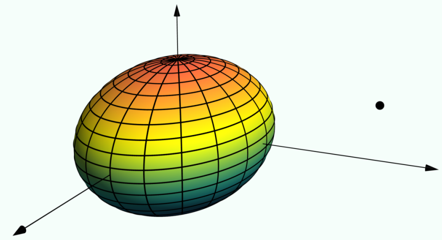
(b) Similarly, the ratio of the lengths of the green line segments from the origin to the first grid intersection is equal to the second eigenvalue of the transformation.

(c) The blue line does not contain an eigenvector, since the direction of its image under the transformation is different from its direction.

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

(a) We define the objective function

$$f(x, y) = (2-x)^2 + (2-y)^2 + \left(1 - \sqrt{(1 - x^2 - 2y^2)/3}\right)^2,$$

which gives the squared distance from $(2, 2, 10)$ to the point on the upper half of the ellipsoid whose first two coordinates are x and y.

We begin at any point, say $(1, 0)$, and we iteratively calculate ∇f at the current location and increment that location by $-\epsilon \nabla f$ for some small value ϵ.

To calculate ∇f, we can apply autodifferentiation by substituting `[x 1; 0 x]` for x and reading off the top right entry of the result.

(b) Yes, gradient descent will work well for this problem, because there are no non-global local minima, or even places where the derivative is close to zero. The direction of fastest decrease is roughly the correct direction, so we expect to get to the minimum promptly.

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

We minimize $|A\mathbf{x} - \mathbf{b}|^2$ by differentiating and setting the result equal to 0:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}}\left[(A\mathbf{x} - \mathbf{b})'(A\mathbf{x} - \mathbf{b})\right] = 0 &\implies \\
(A\mathbf{x} - \mathbf{b})'(A) = 0 &\implies \\
\mathbf{x}'A'A - \mathbf{b}'A = 0 &\implies \\
(A'A)\mathbf{x} = A'\mathbf{b} &\implies \\
\mathbf{x} = (A'A)^{-1}(A'\mathbf{b}).&
\end{aligned}$$

Since the function $\mathbf{x} \mapsto |A\mathbf{x} - \mathbf{b}|^2$ is continuous and goes to ∞ as $|\mathbf{x}| \to \infty$, it has a global minimum on $\mathbb{R}^n$. Since it only has a single critical point, the global minimum must occur there. Therefore, the minimizing value of $\mathbf{x}$ is $(A'A)^{-1}(A'\mathbf{b})$, and the vector closest to $\mathbf{b}$ is $A(A'A)^{-1}(A'\mathbf{b})$.

Final answer:

$$A(A'A)^{-1}(A'\mathbf{b})$$

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

The function counts up to 1.0 in 16 steps, and none of the subsequent additions change the value of `x`, since the step $(1/2)2^{-52}$ only takes you halfway to the next tick, and then the rounding rule takes you back down to 1.0.

Final answer:

1.0

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

A PRNG produces a *deterministic* sequence, which means that it can be computed in a reproducible manner if the seed is known. Random numbers generated from physical processes cannot be reproduced except by recording the values and reading them back. Also, generating randomness in this way is very slow compared to producing a pseudorandom sequence.

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

(a) The condition number measures the magnification of *relative* error, not the magnification of absolute error. Relative error can be large even by virtue of error being large or by virtue of the quantity itself being small. For values of x near 5, the relative error is large because $x - 5$ is much smaller than x.

(b) A separate function exists because the function $x \mapsto e^x - 1$ is well-conditioned, but the algorithm "exponentiate then subtract 1" is unstable for values of x near 0, because the step of subtracting 1 is ill-conditioned. The separate function uses some other algorithm which is presumably (and, indeed, is) stable.

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

There are $3 \times 6 \times 6 = 108$ valid tuples which begin with a prime value and $3 \times 6 \times 6 \times 6 = 648$ tuples which begin with a composite value. Therefore, there are $108 + 648 = 756$ total ways for the experiment to come out.

Final answer:

756

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

(a) We put C inside the region where A and B overlap.

(b) By the principle of inclusion-exclusion,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Substituting, we find that $\mathbb{P}(A \cap B) = 1/3 + 1/3 - 1/2 = 1/6$. Therefore, $\mathbb{P}(C)$ could be as large as $\frac{1}{6}$ but no larger.

Final answer:

$$\frac{1}{6}$$

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

We loop through each array and return `true` if we encounter a sorted array:

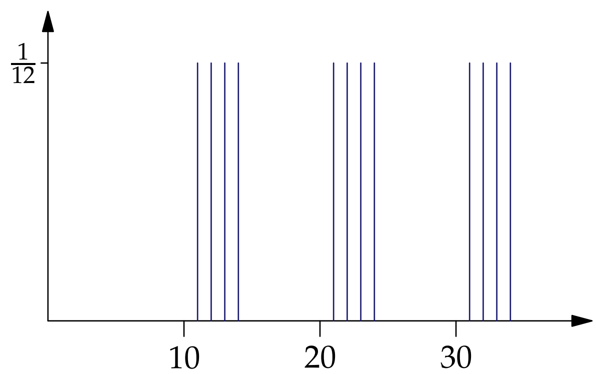
```
function is_sorted(a)
   for i=1:length(a)-1
       if a[i+1] < a[i]
           return false
       end
   end
   true
end
function somesorted(a)
   for vector in a
       if is_sorted(vector)
           return true
       end
   end
   false
end
```

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 $X + Y$.

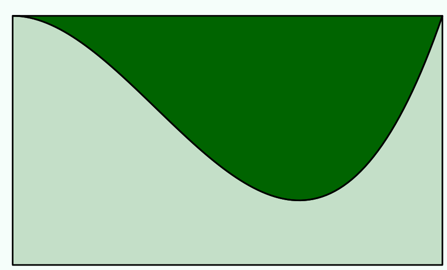
Solution

The probability of each outcome is $\frac{1}{12}$, and the possible outcomes are $\{11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34\}$:



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).

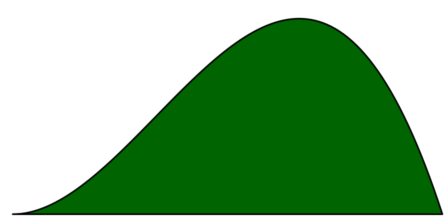


Sketch the graph of the probability density function of X. (Don't worry about normalizing—just get the shape approximately right.)

Solution

The first point landing in the dark region is very unlikely to fall at the far left end of the rectangle, and it is more likely to fall about three quarters of the way across the region.

In fact, the probability that the first point in the dark region lies between the vertical lines at x and $x + \mathrm{d}x$ is proportional to the area of the dark region between those lines, which is $f(x)\,\mathrm{d}x$, where $f(x)$ is the distance from the boundary between the two regions and the top of the box. In other words, we can flip the boundary curve upside down to get the graph of the density function.



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

We generate the random permutation by choosing the entry in the fifth position uniformly at random from the five digits, then we choose the digit in the third position from the other four, and finally we place the last three digits in the other three positions, uniformly at random from the available options.

With this algorithm, we can see that each of the digits $\{1, 2, 3, 5\}$ is equally likely to fall in the third position given that 4 went in the fifth position. Therefore, the desired conditional probability is $\frac{1}{4}$.

Final answer:

$$\frac{1}{4}$$

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

(a) Looks like Jean. Her percentage is higher on both kinds of days.

(b) Jean brings her umbrella on $(70\%)(10\%) + (10\%)(90\%) = 16\%$ of days, and Marvin brings his umbrella on $(60\%)(25\%) + (5\%)(75\%) = 18.75\%$ of days. So actually Marvin brings his umbrella more.

(c) By Bayes' theorem, the conditional probability that it rained today given that Marvin brought his umbrella is

$$\frac{(60\%)(25\%)}{(60\%)(25\%) + (5\%)(75\%)} = 80\%.$$

Final answer:

80%

Problem 17 **[IND]**

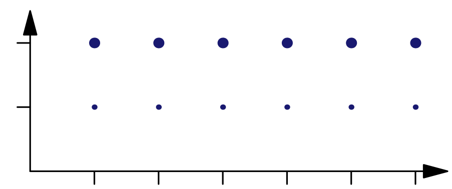
Suppose that X is a fair die roll (that is, its distribution is uniform on $\{1, 2, 3, 4, 5, 6\}$). Suppose Y is a random variable whose distribution has a probability mass of $1/3$ at 1 and a probability mass of $2/3$ at 2.

(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).

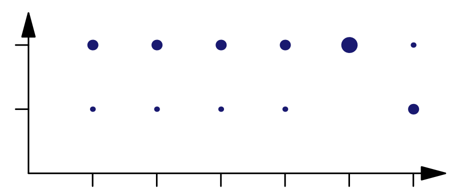
(b) Plot a joint PMF for which X and Y have the marginal distributions specified above but are *not* independent. Use text labels if you want to more clearly identify the amount of mass at each point, and explain the ideas you used to come up with your plot.

Solution

(a) The joint PMF has positive mass at every point in the set $\{1, 2, 3, 4, 5, 6\} \times \{1, 2\}$. Each point in the top row has mass $(2/3)(1/6) = 2/18$, and each point in the bottom row has mass $1/18$.



(b) We obtain the desired result if we move the mass around in such a way that there is still the same amount in each row and column. One simple way to do that is to move a mass in one of the columns to the other row in that same column, and move the same amount of mass in the other direction in a different column. One example of this is shown in the figure.



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

(a) Yes, $\mathbb{E}[X] = \mathbb{E}[Y]$ whenever X and Y have the same distribution. And no, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ does not hold in general; it only holds if $\operatorname{Cov}(X, Y) \neq 0$.

(b) We have $\mathbb{E}[X] = 26.4 + 26.2 = 52.6$ by linearity of expectation. The correlation is irrelevant.

Final answer:

52.6

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 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

(a) There are many possibilities. A simple one is to place four equal masses at the vertices of a square whose diagonals run in the directions of the coordinate axes.

(b) We have

$$\begin{aligned} \mathbb{E}[(X_1 + X_2 + X_3 + X_4)^2] &= \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + \mathbb{E}[X_3^2] + \mathbb{E}[X_4^2] + 2(\mathbb{E}[X_1X_2] + \mathbb{E}[X_1X_2] + \cdots + \mathbb{E}[X_3X_4]) \\ &= (3)(4) + (2)(6)(2) \\ &= 36. \end{aligned}$$

Final answer:

36

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

(a) The conditional distribution of Y given $X = x$ is obtained by restricting the density function to the vertical line at position x. Therefore, the conditional distribution of Y is the uniform distribution on the interval $[0, \sqrt{1-x^2}]$.

(b) The mean of the uniform distribution on $[0, \sqrt{1-x^2}]$ is $\frac{1}{2}\sqrt{1-x^2}$, so $\mathbb{E}[Y \mid X] = \frac{1}{2}\sqrt{1-X^2}$.

Final answer:

$$\frac{1}{2}\sqrt{1-X^2}$$

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

(a) If we write out each X as a sum of 10 Bernoulli's, we see that S is a sum of 400 Bernoulli's (all with $p = 0.02$. Therefore, S is binomial with parameters 400 and 0.02.

(b) Since $\text{Bin}(n, \lambda/n) \approx \text{Poiss}(\lambda)$ when n is large and λ is of modest size, the distribution of S is approximately Poisson(8).

Final answer:

8

Problem 22 **[COMDISTC]**

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

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

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

(a) We recognize the density as a multivariate normal density, so the distribution of X is a normal distribution with mean -3 and variance 2.

(b) Correlation is normalized covariance:

$$\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}} = \frac{1}{\sqrt{2}\sqrt{2}} = \frac{1}{2}.$$

Final answer:

$\frac{1}{2}$

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

If we apply the given equality to the suggested random variable and set $k = 1$, then we get

$$\mathbb{P}\left(\left|X - \frac{1}{2}\right| \leq \frac{1}{2}\right) < 1,$$

which is false since the event $\left\{\left|X - \frac{1}{2}\right| \geq \frac{1}{2}\right\}$ occurs with probability 1.

Problem 24 **[CLT]**

Suppose that S_n 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.

$$
\begin{aligned}
A &= \mathbb{P}(S_n \le n/3 + \sqrt{n}) \\
B &= \mathbb{P}(S_n \ge 0.33n) \\
C &= \mathbb{P}(S_n < 0) \\
D &= \mathbb{P}(S_n < 10^4)
\end{aligned}
$$

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

Solution

By the central limit theorem, A converges to a number strictly between $\frac{1}{2}$ and 1 as $n \to \infty$. Since $0.33n = n/3 - 0.00\overline{3}n$ and since $0.00\overline{3}n$ is much larger than $\sqrt{n}$ when n is large, the central limit theorem implies that B is very close to 1. Since S's distribution is not supported on the negative numbers, we have $C = 0$. And D is very small but nonzero. So the correct order is C, D, A, B.

Final answer:

CDAB

Additional space

Additional space

Additional space

Additional space

Additional space