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

Name:

You will have three hours to complete the exam. It consists of 24 written questions and one separate computational problem. 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 computational problem, you will hand in your answers to the written portion and then get out your laptop to submit a solution to the last question electronically.

For the written part of the exam, no calculators or other materials are allowed, except the Julia-Python-R reference sheet and a list of your current scores on the standards. 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.*

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

Problem 1 **[PMF]**

Suppose that X and Y are independent random variables whose probability mass functions m_X and m_Y are defined as follows:

$$m_X(1/3) = 1/2 \quad m_X(1) = 1/4 \quad m_X(2) = 1/4$$

$$m_Y(1/3) = 1/3 \quad m_Y(1) = 2/5 \quad m_Y(2) = 1/5 \quad m_Y(3) = 1/15$$

(a) How many points $(x, y) \in \mathbb{R}^2$ have the property that $m_{X,Y}(x, y) \neq 0$, where $m_{X,Y}$ is the joint PMF of X and Y?

(b) How many points $z \in \mathbb{R}$ have the property that $m_Z(z) \neq 0$, where m_Z is the PMF of $Z = X + Y$?

Put your answer in the box as an ordered pair (answer to (a), answer to (b)).

Solution

Final answer:

Problem 2 **[PDF]**

Suppose that X is a random variable whose density is given by

$$f_X(x) = 2x\mathbf{1}_{\{0 \le x \le 1\}}.$$

Which random variable has larger expected value, $\sqrt{X}$ or X^2?

Solution

Final answer:

Problem 3 **[CONDPROB]**

Three cards are drawn without replacement from a well-shuffled standard deck. Find the conditional probability that the cards are all diamonds given that they are all red cards. (Note: 13 of the cards are diamonds, 26 of the cards are red, and all of the diamonds are red).

Solution

Final answer:

Problem 4 **[BAYES]**

(a) Suppose that the conditional probability of an email (chosen uniformly at random from a large collection of emails) containing the phrase "additional income", given that the email is spam, is 14%. Suppose that the conditional probability of an email being spam, given that it contains the phrase "additional income", is 88%. Find the ratio of the probability that an email is spam to the probability that an email contains the phrase "additional income".

(b) We flip a weighted coin that has probability $\frac{3}{4}$ of turning up heads. If we get heads, we roll a six-sided die, and otherwise we roll an eight-sided die. Given that the die turns up 4, what is the conditional probability that the coin turned up heads?

Solution

Final answer:

Problem 5 **[IND]**

(a) Suppose that $X_1, \ldots, X_{10}$ are independent Bernoulli(p) random variables defined on a probability space Ω. What is the smallest possible cardinality of Ω?

(b) Suppose that U and V are independent random variables, each selected uniformly at random from $[0, 1]$. Find the probability of the event $\{\frac{1}{2}U + V \leq 1\}$.

Solution

Final answer:

Problem 6 **[EXP]**

(a) Find the expected value of the sum of the sum and product of two independent die rolls.

(b) You roll a die, and if the result is prime you roll two more dice, and if it isn't prime you roll *three* more dice. Find the expected number of pips showing on the top faces of all of the dice rolled (so, either three dice or four dice).

Solution

Final answer:

Problem 7 **[COV]**

Suppose that X_1 and X_2 are independent and identically distributed.

(a) Find the covariance of $X_1 + X_2$ and $X_1 - X_2$.

(b) Show that if X_1 and X_2 are normal random variables, then $X_1 + X_2$ and $X_1 - X_2$ are independent. Hint: use your knowledge of the multivariate normal distribution density.

Solution

Final answer:

Problem 8 **[CONDEXP]**

(a) Suppose that, for all $x \in \mathbb{R}$, the conditional distribution of Y given $X = x$ is exponential with parameter $\lambda = 2|x| + 1$. Find $\mathbb{E}[Y \mid X]$.

(b) What is the strongest conclusion that can be drawn about the distribution of X, based on the information in (a)?

Solution

Final answer:

Problem 9 **[COMDISTD]**

Suppose that $S = X_1 + \ldots + X_n$, where the X_i's are independent $\mathrm{Ber}(p)$ random variables.

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

(b) Find the probability mass function for the conditional distribution of S given $\{X_1 = 1\}$.

(c) You collect some data over a few years, and you find that the number of near-doorings you experience per month on your bicycle commute is approximately Poisson distributed. Give an explanation for why the Poisson distribution might be expected to emerge in this context.

Solution

Final answer:

Problem 10 **[COMDISTC]**

(a) Find the probability density function of the distribution of $\sqrt{X}$, where X is an exponential random variable with parameter λ.

(b) Find $\mathbb{P}(Z = 0.5)$, where Z is a standard normal random variable.

Solution

Final answer:

Problem 11 **[RVINEQ]**

(a) Suppose $k > 0$. Explain why $\mathbb{P}(|X - \mu| > k\sigma) < 1/k^2$, if μ and σ are the mean and standard deviation of X, respectively.

(b) Use Chebyshev's inequality to find an interval centered at 3.5 which contains X with probability 99%, where X is the average of 10,000 independent fair die rolls. (Note: the variance of a fair die roll is $\frac{35}{12}$.) Feel free to leave your answer in unsimplified form.

Solution

Problem 12 **[CLT]**

The **chi-squared distribution** with parameter n is the distribution of the sum of the squares of n independent standard normal random variables.

Let S_k be the sum of k independent chi-squared random variables with parameter 8. Find the limit as $k \to \infty$ of

$$\mathbb{P}(8k \leq S_k \leq 8.01k).$$

Solution

Final answer: