

Spring 2019

Test: Probability and Statistics - Midterm Topics

Evan Agovino
CUNY City College

NYC Tech-in-Residence Corps

Follow this and additional works at: https://academicworks.cuny.edu/cc_oers



Part of the [Computer Sciences Commons](#)

[How does access to this work benefit you? Let us know!](#)

Recommended Citation

Agovino, Evan and NYC Tech-in-Residence Corps, "Test: Probability and Statistics - Midterm Topics" (2019). *CUNY Academic Works*.

https://academicworks.cuny.edu/cc_oers/184

This Other is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Open Educational Resources by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

Midterm Topics

Descriptive Statistics

Categorical Data - What types of data are there?

Measures of Central Tendency - What are some common measures of central tendency? When are each of them useful? When are each of them less useful?

Measures of Spread - What are some common measures of spread? How are they found?

Data Visualization:

Histograms - What does a histogram show? What are the benefits of a histogram? What are the drawbacks?

Swarm Plots - What does a swarm plot show? What are the benefits of a swarm plot? What are the drawbacks?

ECDF - What does an ECDF show? What are the benefits of an ECDF? What are the drawbacks?

Boxplots - What does a boxplot show? What are the benefits of a boxplot? What are the drawbacks?

Probability:

What is the difference between statistics and probability?

Definitions - What is a sample space? What is an event? What is a probability function? What is the intersection of two sets? What is the union of two sets? What is the complement of a set?

What is the Inclusion-Exclusion principle?

What are disjoint sets?

What is the rule of product?

Permutations and Combinations - What is a permutation? What is a combination? What is the difference between the two? Do you know when to use one or the other?

Sampling - What is sampling with replacement? What is sampling without replacement?

Conditional Probability - What is conditional probability? What is the multiplication rule?

What is the law of total probability?

When are two events independent? Can you identify whether two events are independent or not?

Bayes' Theorem - What is Bayes' Theorem? When can it be particularly useful?

What is a True Positive? What is a False Positive? What is a True Negative? What is a False Negative?

Classification Metrics - What is Accuracy? What is Precision? What is Recall?

If I give you a classification scenario, can you identify which metric we want to maximize? (i.e. we want to create software that detects cheating and ensure that we minimize false positives at the cost of letting some people get away with cheating)

Discrete Distributions

What is a Bernoulli distribution?

What is a Binomial distribution?

What is a Geometric distribution? What is a Poisson distribution? Which parameters do each of these distributions require?

Which distribution can the Poisson be used in place of? Why might people do this? Why may they not need to do it anymore? Can you identify which distribution is used when?

Can you give me an example of each of these distributions? What is the probability mass function?

What is the cumulative distribution function?

Continuous Distributions

What is the difference between a discrete distribution and a continuous distribution?

What is the probability density function?

What is the fundamental difference between the probability mass function and the probability distribution function, and thus a central tenet of the difference between discrete and continuous variables, and the method in which we measure them?

What are the parameters for the normal distribution? What is the standard normal distribution?

What is a Z-score? Why is it important? What does a negative Z-score mean? What does a positive Z-score mean?

What is the law of large numbers?

What is the Central Limit Theorem? Why is the Central Limit Theorem so important?

Which distribution(s) can the normal distribution be an approximate for?

Why might this be useful? Why might this be not so useful anymore?