1. Introduction

Welcome to the series of E-learning modules on Markov's inequality. In this module, we are going to cover Markov's inequality, its proof and applications, relationship with Tchebyscheff's inequality and sample examples to apply the inequality.

By the end of this session, you will be able to:

- Explain the Markov's inequality
- Derive the inequality
- Explain the implications and applications of the inequality
- Explain the relationship with Tchebyscheff's inequality

One of the most important tasks in analyzing randomized algorithms is to understand what random variables arise and how well they are **concentrated**. A variable with good concentration is one that is close to its mean with good probability. A "concentration inequality" is a theorem proving that a random variable has good concentration. Such theorems are also known as "tail bounds".

Markov's inequality
This is the simplest concentration inequality. In probability theory, **Markov's inequality** gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant.

Markov's inequality is named after the Russian mathematician Andrey Markov. Although it appeared earlier in the work of Pafnuty Chebyshev (Markov's teacher), and many sources, especially in analysis, refer to it as Chebyshev's inequality or Bienayme's inequality.

Markov's inequality (and other similar inequalities) relate probabilities to expectations, and provide (frequently) loose but still useful bounds for the cumulative distribution function of a random variable.
An example of an application of Markov's inequality is the fact that (assuming incomes are non-negative) no more than 1 by 5 of the population can have more than 5 times the average income.

If the expectation value of a non-negative random variable is small, then the random variable must itself be small with high probability. Markov's inequality quantifies this observation.
Markov's inequality is a helpful result in probability that gives information about a probability distribution. The remarkable aspect about it is that the inequality holds for any distribution with positive values, no matter what other features that it has.

Markov's inequality gives an upper bound for the percent of the distribution that is above a particular value.
Let us now look at the Markov Inequality. Even though the statement looks very simple, clever application of the inequality is at the heart of more powerful inequalities like Chebyshev or Chernoff.

2. Statement of Markov's Inequality and the Reverse Markov Inequality

Statement of Markov's Inequality
Markov's inequality states that for a positive random variable *X* and any positive real number epsilon, the probability that *X* is greater than or equal to epsilon is less than or equal to the expected value of *X* divided by epsilon. The above description can be stated more succinctly using mathematical notation.

Suppose a random variable X takes only nonnegative values so that, probability of X greater than or equal to zero is equal to 1. How much probability is there in the tail of the distribution of X? More specifically, for a given value epsilon greater than zero, what can we say about the value of Probability of X greater than or equal to epsilon?
Let X be a non- negative random variable with finite expectation E of x. Then for any epsilon greater than 0, **Probability of X greater than or equal to epsilon is less than or equal to expected value of X by epsilon.** Markov's inequality takes mu is equal to Expected value of X into account and provides an upper bound on **Probability of X greater than or equal to** epsilon that depends on the values of epsilon and mu. **-**

Proof:
Case 1: Let X be a continuous random variable with probability density function f of x
Mu is equal to Expected value of X which is equal to integral from 0 to infinity x into f of x dx which is equal to integral from 0 to epsilon x into f of x dx plus integral from epsilon to infinity x into f of x dx
which is greater than or equal to integral from epsilon to infinity x into f of x dx
The above inequality holds because the integral ignored integral from 0 to epsilon x into f of x dx is non-negative.

The below inequality holds because, X greater than or equal to epsilon for X in [epsilon, infinity).
which is greater than or equal to integral from epsilon to infinity epsilon into f of x dx which is equal to epsilon into integral from epsilon to infinity f of x dx
which is equal to epsilon into Probability of epsilon less than or equal to X less than infinity which is equal to epsilon into Probability of X greater than or equal to epsilon
from which we obtain Markov's Inequality:
Probability of X greater than or equal to epsilon less than or equal to mu by epsilon is equal to expected value of X by epsilon

Case 2:
Let X be a Discrete random variable with probability mass function p of x
The proof for a discrete random variable is similar, with summations replacing integrals.
Mu is equal to Expected value of X which is equal to summation from 0 to infinity x into p of x which is equal to summation from 0 to epsilon x into p of x plus summation from epsilon to

infinity x into p of x
which is greater than or equal to summation from epsilon to infinity x into p of x
The above inequality holds because the summation ignored summation from 0 to epsilon x into p of x is non-negative.

The below inequality holds because, X greater than or equal to epsilon for X in [epsilon, infinity).
which is greater than or equal to summation from epsilon to infinity epsilon into p of x which is equal to epsilon into summation from epsilon to infinity p of x
which is equal to epsilon into Probability of epsilon less than or equal to X less than infinity which is equal to epsilon into Probability of X greater than or equal to epsilon
From which we obtain Markov's Inequality:
Probability of X greater than or equal to epsilon less than or equal to mu by epsilon is equal to expected value of X by epsilon

There are some basic things to note here. First, the term Probability of X greater than or equal to epsilon estimates the probability that the random variable will take the value that exceeds epsilon. The term Probability of X greater than or equal to epsilon is related to the cumulative density function as 1 minus Probability of X less than epsilon. Since the variable is non-negative, this estimates the deviation on one side of the error.
Markov's Inequality gives an upper bound on Probability of X greater than or equal to epsilon that applies to any distribution with positive support.

The Reverse Markov inequality
In some scenarios, we would also like to bound the probability that Y is much smaller than its mean. Markov's inequality can be used for this purpose if we know an upper bound on$Y$.

The following result is an immediate corollary to the above theorem.
Corollary:
Let Y be a random variable that is never larger than B. Then, for all epsilon less than B,
Probability of Y less than or equal to epsilon is less than or equal to Expected value of (B minus Y) by (B minus epsilon)
The downside is that it only gives very weak bounds, but the upside is that needs almost no assumptions about the random variable. It is often useful in scenarios where not much concentration is needed, or where the random variable is too complicated to be analyzed by inequalities that are more powerful.

3. Relationship to the Chebyshev Inequality and Illustration of the Inequality

Relationship to the Chebyshev Inequality
Take X is equal to X minus E of X. This is non-negative. Even if **X** is negative, and its expectation exists then, V of X is the variance**.** Then,
According to Markov's Inequality
Probability of X greater than or equal to epsilon is less than or equal to Expected value of X by epsilon
By substituting X is equal to X minus E of X

Probability of X minus E of X greater than or equal to epsilon is less than or equal to Expected value of X minus E of X by epsilon.
By squaring, we get,
Probability of X minus E of X whole square greater than or equal to epsilon square is less than or equal to Expected value of X minus E of X whole square by epsilon square, which is equal to V of X by epsilon square
According to Tchebyscheff's inequality, Probability of modulus of X minus mu greater than or equal to k sigma is less than or equal to 1 by k square

Let E of X is equal to mu and epsilon is equal to k sigma then k is equal to epsilon by sigma and V of X by epsilon square is equal to sigma square by k square sigma square which is equal to 1 by k square
Probability of X minus E of X whole square greater than or equal to epsilon square is less than or equal to V of X by epsilon square implies Probability of X minus mu whole square greater than or equal to k sigma whole square less than or equal to 1 by k square
Which implies Probability of modulus of X minus mu greater than or equal to k sigma is less than or equal to 1 by k square
This is the Chebyshev inequality. Similar bounds can be derived for higher moments, but do not have eponyms.

One advantage of Markov's inequality is that the computation of the expectation value is sufficient, so it is typically easy to apply. However, Markov's inequality does not depend on any property of the probability distribution of the random variable. Therefore, one can often improve upon this estimate if further information on the probability distribution is available.

Inequalities from an Adversarial Perspective
One interesting way of looking at the inequalities is from an adversarial perspective. The adversary has given you some limited information and you are expected to come up with some bound on the probability of an event.
For example, in the case of Markov inequality, all you know is that the random variable is non-negative and its (finite) expected value.

Based on this information, Markov inequality allows you to provide some bound on the tail inequalities. Similarly, in the case of Chebyshev inequality, you know that the random variable has a finite expected value and variance. Armed with this information Chebyshev inequality allows you to provide some bound on the tail inequalities.

The most fascinating thing about these inequalities is that you do not have to know the probabilistic mass function. For any arbitrary pmf satisfying some mild conditions, Markov and Chebyshev inequalities allow you to make intelligent guesses about the tail probability.

Illustration of the Inequality

To illustrate the inequality, let us consider a distribution with non-negative values (such as a chi-square distribution). If this random variable *X* has expected value of 3, we will look at probabilities for a few values of epsilon.

- For epsilon is equal to 10 Markov's inequality says that Probability of X greater than or equal to 10 is less than or equal to 3 by 10 which is equal to thirty percent. Therefore, there is a thirty percent probability that *X* is greater than 10.

- For epsilon is equal to thirty, Markov's inequality says that Probability of X greater than or equal to thirty is less than or equal to 3 by thirty which is equal to ten percent. Therefore, there is a ten percent probability that X is greater than thirty.

- For epsilon is equal to 3, Markov's inequality says that Probability of X greater than or equal to 3 is less than or equal to 3 by 3 which is equal to 1 percent. Events with probability of 1 is equal to hundred percent are certain. Therefore, this says that some value of the random variable is greater than or equal to 3. This should not be too surprising. When the value of X less than 3, then the expected value would also be less than 3.

- As the value of epsilon increases, the quotient E of X by epsilon will become smaller and smaller. This means that the probability is very small that X is very, very large. Again, with an expected value of 3, we would not expect much of the distribution with values that were very large.

4. Uses of the Inequality and Intuitive Explanation of Markov Inequality

Uses of the Inequality

If we know more about the distribution that we are working with, then we can usually improve on Markov's inequality. The value of using it is that it holds for any distribution with non-negative values.

For example, if we know the mean height of the students at an elementary school, then the Markov's inequality tells us that no more than one sixth of the students can have a height greater than six times the mean height.

The other major use of Markov's inequality is to prove Chebyshev's inequality. This fact results in the name "Chebyshev's inequality" being applied to Markov's inequality as well. The confusion of the naming of the inequalities is also due to historical circumstances. Andrey Markov was the student of Pafnuty Chebyshev. Chebyshev's work contains the inequality that is attributed to Markov.

Intuitive Explanation of Markov Inequality

Intuitively, given a non-negative random variable and its expected value E(X)

- The probability that X takes a value that is greater than twice the expected value is at most half. In other words, if you consider the pmf curve, the area under the curve for values that are beyond 2 into E(X) is at most half
- The probability that X takes a value that is greater than thrice the expected value is at most one third and so on

Let us see why that makes sense.

Let X be a random variable corresponding to the scores of hundred students in an exam. The variable is clearly non-negative as the lowest score is 0. Tentatively, let us assume the highest value is hundred (even though we will not need it). Let us see how we can derive the bounds given by Markov inequality in this scenario.

Let us also assume that the average score is twenty (must be a lousy class!). By definition, we know that the combined score of all students is two thousand (that is twenty into hundred).

Let us take the first claim - The probability that X takes a value that is greater than twice the expected value is at most half. In this example, it means the fraction of students who have scored greater than forty (2 into twenty) is at most point five. In other words, at most fifty students could have scored forty or more.

If fifty students got exactly 40 and the remaining students all got 0, then the average of the whole class is 20. Now, if even one additional student got a score greater than 40, then the total score of hundred students become two thousand forty and the average becomes twenty point four, which is a contradiction to our original information.

Note that the scores of other students that we assumed to be 0 is an over simplification and we can do without that. For example, we can argue that if 50 students got 40, then the total score is at least two thousand and hence the mean is at least 20.

5. Practical Consequences

Practical Consequences:

For most distributions of practical interest, the probability in the tail beyond epsilon is noticeably smaller than mu by epsilon for all values of epsilon.

For continuous and discrete distributions, each with mu is equal to 1, we use R to show that the non increasing "reliability function"

R of epsilon is equal to 1 minus F of epsilon, which is equal to Probability of X greater than epsilon is bounded above by mu by epsilon equals to 1 by epsilon.

For continuous distributions there is no difference between Probability of X less than or equal to epsilon and Probability of X less than epsilon and for discrete distributions the discrepancy is sometimes noticeable. The Markov bound 1 by epsilon is not useful for epsilon less than 1 (that is 1 by epsilon greater than 1) because no probability exceeds 1.

Markov and Chebyshev inequalities are two simplest, yet very powerful inequalities. Careful utility of them provide very useful bounds without knowing anything about the distribution of the random variable. Markov inequality bounds the probability that a non-negative random variable exceeds any multiple of its expected value (or any constant).

On the other hand, Chebyshev's inequality bounds the probability that a random variable deviates from its expected value by any multiple of its standard deviation. Chebyshev does not expect the variable to non-negative but needs additional information to provide a tighter bound. Both Markov and Chebyshev inequalities are tight - This means with the information provided, the inequalities provide the most information that can be provided by them.

Here's a summary of our learning in this session, where we understood:

- The concept of Markov's inequality
- The proof of Markov's inequality or theorem
- The applications and implications of Markov's inequality
- The relationship with Tchebyscheff's inequality