<u>Frequently Asked Questions</u>

1. What do you mean by Markov's inequality?

Answer:

Markov's inequality 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.

It 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.

Hence, Markov's inequality gives an upper bound for the percent of the distribution that is above a particular value.

2. State and prove Markov's inequality for a continuous random variable.

Answer:

Statement:

Let X be a non- negative random variable with finite expectation E(x) .Then for any $\varepsilon > 0$, P $(X \geq \varepsilon) \leq \dfrac{E(X)}{\varepsilon}$

Markov's inequality takes $\mu = E(X)$ into account and provides an upper bound on $P\{X \geq \varepsilon\}$ that depends on the values of ε and μ.

Let X be a continuous random variable with probability density function f(x)

$$\mu = E(X) = \int_0^\infty x\ f(x)\,dx = \int_0^\varepsilon x\,f(x)\,dx + \int_\varepsilon^\infty x\,f(x)\,dx$$

$$\geq \int_\varepsilon^\infty x\,f(x)\,dx$$

The above inequality holds because the integral ignored $\int_0^{\varepsilon} x\, f(x)\, dx$ is nonnegative.

The below inequality holds because, $X \geq \varepsilon$ for X in [ε, ∞).

$$\geq \int_{\varepsilon}^{\infty} \varepsilon\, f(x)\, dx = \varepsilon \int_{\varepsilon}^{\infty} f(x)\, dx = \varepsilon\ P(\varepsilon \leq X < \infty) = \varepsilon\, P\{X \geq \varepsilon\},$$

From which we obtain Markov's Inequality:

$P\{X \geq \varepsilon\} \leq \mu/\varepsilon = E(X)/\varepsilon$

3. Deduce Markov's inequality for a discrete random variable.

Answer:

Let X be a discrete random variable with probability mass function p(x)

The proof for a discrete random variable is similar, with summations replacing integrals.

$$\mu = E(X) = \sum_{0}^{\infty} xp(x) = \sum_{0}^{\varepsilon} xp(x) + \sum_{\varepsilon}^{\infty} xp(x)$$

$$\geq \sum_{\varepsilon}^{\infty} xp(x)$$

The above inequality holds because the summation ignored $\sum_{0}^{\varepsilon} xp(x)$ is nonnegative.

The below inequality holds because, $X \geq \varepsilon$ for X in [ε, ∞).

$$\geq \sum_{\varepsilon}^{\infty} \varepsilon p(x) = \varepsilon \sum_{\varepsilon}^{\infty} p(x) = \varepsilon\ P(\varepsilon \leq X < \infty) = \varepsilon\, P\{X \geq \varepsilon\},$$

from which we obtain Markov's Inequality:

$P\{X \geq \varepsilon\} \leq \mu/\varepsilon = E(X)/\varepsilon$

4. Write a note on the reverse Markov inequality.

Answer:

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 of Markov's theorem

Corollary: Let Y be a random variable that is never larger than B. Then, for all $\varepsilon < B$,

$$P[Y \le \varepsilon] \le \frac{E[B-Y]}{B-\varepsilon}$$

5. How do you relate Markov's inequality to Tchebysheff's inequality?

Answer:

Take X = X-E(X). This is non-negative, whether or not X is, and its expectation, if it exists, is the variance V(X). Therefore

According to Markov's Inequality

P (X$\ge \varepsilon) \le \frac{E(X)}{\varepsilon}$. By substituting X = X-E(X)

$$P[X - E(X) \ \ge \varepsilon] \le \frac{E[(X - E(X))}{\varepsilon}$$. Squaring

$$\Rightarrow P[(X - E(X))^2 \ge \varepsilon^2] \le \frac{E(X - E(X))^2}{\varepsilon^2} = \frac{V(X)}{\varepsilon^2}$$

According to Tchebyscheff's inequality $P\{|x-\mu| \ge k\sigma\} \le \frac{1}{k^2}$

Let E(X) = μ and ε = kσ then k= σ/ε and V(X) / $\varepsilon^2 = \sigma^2 / k^2 \sigma^2 = 1/k^2$

$$P[(X - E(X))^2 \ge \varepsilon^2] \le \frac{V(X)}{\varepsilon^2} \Rightarrow P[(X - \mu)^2 \ge (k\sigma)^2 \le \frac{1}{k^2}]$$

$\Rightarrow P\{|x-\mu| \ge k\sigma\} \le \frac{1}{k^2}$ which is the Chebyshev inequality

6. Illustrate Markov's rule for different values of the constant ε .

Answer:

Suppose we have a distribution with nonnegative 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 ε .

- For ε = 10 Markov's inequality says that $P(X \geq 10) \leq 3/10 = 30\%$. Therefore, there is a 30% probability that X is greater than 10.
-
- For ε = 30 Markov's inequality says that $P(X \geq 30) \leq 3/30 = 10\%$. Therefore, there is a 10% probability that X is greater than 30.
-
- For ε = 3 Markov's inequality says that $P(X \geq 3) \leq 3/3 = 1$. Events with probability of 1 = 100% 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. Were all the value of X less than 3, then the expected value would also be less than 3.

- As the value of ε increases, the quotient $E(X) / \varepsilon$ 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 there to be much of the distribution with values that were very large.

7. How can you consider Markov and Tchebyscheff's inequalities as the Inequalities from an Adversarial Perspective?

Answer:

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.

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

8. What are the uses of Markov's inequality?

Answer:

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

For example, if we know the mean height of students at an elementary school. 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.

One advantage of Markov's inequality is that the computation of the expectation value is sufficient, so it 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.

9. What is the Practical consequences Markov's inequality?

Answer:

For most distributions of practical interest, the probability in the tail beyond ε is noticeably smaller than μ/ε for all values of ε.

For several continuous and discrete distributions, each with $\mu = 1$,

We use R to show that the non increasing "reliability function" $R(\varepsilon) =$

$1 - F(\varepsilon) = P\{X > \varepsilon\}$ is bounded above by $\mu/\varepsilon = 1/\varepsilon$.

(For continuous distributions there is no difference between $P\{X \leq \varepsilon\}$ and $P\{X < \varepsilon\}$,and for discrete distributions the discrepancy is not noticeable in some cases.)The Markov bound $1/\varepsilon$ is not useful for $\varepsilon < 1$ (that is $1/\varepsilon > 1$) because no probability exceeds 1.

10. Give the Intuitive Explanation of Markov Inequality with an example.

Answer:

Intuitively, what this inequality means is that, given a non-negative random variable and its expected value E(X)

(1) 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*E(X) is at most half.

(2) 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 using one example

Let X be a random variable corresponding to the scores of 100 students in an exam. The variable is clearly non-negative as the lowest score is 0. Tentatively, let us assume the highest value is 100 (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 20 (must be a lousy class!). By definition, we know that the combined score of all students is 2000 (20*100).

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 score greater than 40 (2*20) is at most 0.5. In other words, at most 50 students could have scored 40 or more. It is very clear that it must be the case.

If 50 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 100 students become 2040 and the average becomes 20.4, 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** 2000 and hence the mean is **at least** 20.

We can also see how the second claim is true. The probability that X takes a value that is greater than thrice the expected value is at most one third. If 33.3 students got 60 and others got 0, then we get the total score as around 2000 and the average remains the same. Similarly, regardless of the scores of other 66.6 students, we know that the mean is at least 20 now.

This also must have made clear why the variable must be non-negative. If some of the values are negative, then we cannot claim that mean is at least some constant C. The values that do not exceed the threshold may well be negative and hence can pull the mean below the estimated value.

11. Consider a coin that comes up with head with probability 0.2. Let us toss it n times. Use Markov inequality to bound the probability by at least 80% of heads.

Answer:

Let X be the random variable indicating the number of heads we got in n tosses. Clearly, X is non-negative. Using linearity of expectation, we know that E[X] is 0.2n.

We have to bound the probability by at least 80% =0.8

Then we have to find P(X >= 0.8n). Using Markov inequality, we have

$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$.

$P(X \geq 0.8n) \leq \frac{0.2n}{0.8n} = 0.25$

12. Give an example to show that in some situations Markov's inequality provides weak results.

Answer:

Let X be the face that shows up when we toss a die. Let us now find the probability that X >= 5 by Markov inequality.

We know that E[X] = $\sum xp(x) = \frac{1}{6}[1+2+...+6] = \frac{6(7)}{6(2)}$ =7/2 = 3.5.

By Markov's inequality P $(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$

$P(X \geq 5) \leq \frac{3.5}{5} = 0.7$

The actual answer of course is 2/6 and the answer is quite off. This becomes even more bizarre, for example, if we find P(X >= 3) . By Markov inequality,

$P(X \geq 3) \leq \frac{3.5}{3} = \frac{7}{6}$

The upper bound is greater than 1! Of course using axioms of probability, we can set it to 1 while the actual probability is closer to 0.66. You can play around with the coin example or the score example to find cases where Markov inequality provides weak results.

13. X is a random variable with mean two. Find an upper bound to P{X > 8}.

Answer:

Given μ=2

By Markov's inequality P $(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$

Therefore

P (X> 8) $\leq \frac{2}{8} = 0.25$

Hence the upper bound to P{X > 8} is 0.25

14. If X is a random variable with mean =6. Find t such that P{x ≥ t} ≥ 0.96

Solution:

Given μ=6. Therefore,

By Markov's inequality P $(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$

Substituting

P ($X \geq \varepsilon) \leq \frac{6}{\varepsilon}$.----(1)

Put $$\frac{6}{\varepsilon} = 0.96 \Rightarrow \varepsilon = \frac{6}{0.96} = 6.25$$

Substituting in (1)

P (X$\geq 6.25) \leq \frac{6}{\varepsilon}$ =0.96

Therefore, t= 6.25

15. Flip fair coin n times. Obtain a bound to obtain a probability of getting at least 70% of heads.

Answer:

Let us flip fair coin n times. Let Xi be the indicator random variable for the event that the ith coin flip is head.

Then X = ∑ Xi, is the number of heads in the sequence of n coin flips.

Xi = 1 if the ith coin faces head.

= 0 otherwise

Since a fair coin is tossed, P(Xi=1) =1/2 for all i

Therefore,

E[X] = E(∑Xi) = ∑ E(xi)= ∑ P(Xi=1)=n/2

By Markov's inequality P (X$\geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$

P (X$\geq 0.7n) \leq \frac{E(X)}{0.7n} = \frac{n/2}{0.7n} = 0.714$

The probability to obtain 70% or more heads in such a sequence of coin flips is thus bounded by 0.714.