Chebyshev Mathematical Inequalities +2

What is the relevance of the Chebyshev's inequality in real life application?

This question previously had details. They are now in a comment.

4 Answers

Michael Lamar, PhD in Applied Mathematics
Answered Jun 16, 2015

On the one hand, you can always construct a distribution that satisfies the Chebyshev bound on probabilities in the tail with equality. That means there exists a distribution with finite variance that has the property that exactly $\frac{1}{k^2}$ of the probability is more than k standard deviations away from the mean (for $k > 1$). So it's not like there is some other better bound waiting around to be discovered.

On the other hand, I think it is safe to say that for the vast majority of what you might call "naturally occurring" distributions, the bound given by the Chebyshev inequality is lousy. As Michael Hochster's answer points out, for the normal distribution, the bound is off by a factor of 5 for two standard deviations, and it only gets worse from there. For three standard deviations, the bound is too conservative by a factor of over 40, and for four standard deviations, the factor is nearly 1000. But it's not just the normal distribution with it's rapidly decaying tails that does poorly.

For an exponential distribution with mean μ, the standard deviation is also μ. The cumulative distribution function is $F(x) = 1 - e^{-\frac{x}{\mu}}$ so the probability of being more than k standard deviations away from the mean (which means taking on a value of greater than $(k+1)\mu$) turns out to be exactly $e^{-(k+1)}$. Obviously this exponential function decays far more rapidly than $\frac{1}{k^2}$ so once again, the bound starts out bad and only gets worse. The bound does the best job for $k = 2$ when it's off by a factor of just over 5, but by the time $k = 10$, the factor is nearly 600.

So the point of my answer is to say that it's not the kind of bound that is particularly useful in practice because it's just too conservative for "typical" distributions that people work with. However, when you want to prove things that must be true for ANY distribution, it's a great tool to have because it allows you to control the impact of "rare" events.

Related Questions

What is a real-world application of Chebyshev's inequality?

Mathematics: What are some intuitive proofs of Markov's inequality and Chebyshev's inequality?

What is an intuitive explanation of Chebyshev's inequality?

How tight of a bound is Chebyshev's inequality?

Where are inequalities used in real life?

Do I use Chebyshev's inequality for this question?

Is Chebyshev's inequality relevant to quantum mechanics?

What are main applications of Jensen's Inequality?

Is inequality a fact of life?

What are some cool applications of the Cauchy Schwarz Inequality?

+ Ask New Question

More Related Questions

Question Stats

22 Public Followers

9,812 Views

Last Asked Jun 26, 2015

Edits

Michael Hochster, PhD in Statistics, Stanford
Updated Jan 6, 2014 · Upvoted by Justin Rising, PhD in statistics

Not sure if the following really answers your question. It's mainly just another explanation of what Chebyshev's inequality says.

You often hear "standard deviations away from the mean" as a way of summarizing how rare or unusual an event is (as in six-sigma). A conventional example is that "two or more standard deviations away from the mean" translates to approximately "the 5% most extreme values."

The assumption connecting the "two standard deviations from the mean" to the "5% most extreme" is that the underlying distribution is bell-shaped (Gaussian). Without assuming a bell-shaped curve, we can't make this statement. But can we make any statement at all?

Yes! The Chebyshev inequality says that no more than **25%** of the data can be two or more standard deviations from the mean, whatever the underlying distribution. So the Chebyshev inequality lets us make a weaker connection between "standard deviations from the mean" and "tail probabilities" but at a much lower cost in assumptions.

There are also one-sided Chebyshev-like inequalities that can answer interesting questions. For example, suppose I characterize "high performers" on my exam as those who score more than two standard deviations *above* the mean. Exercise: what is the maximum possible fraction of high performers on my exam?

Chebyshev's inequality can be used to show that when the expected value of a nonnegative random variable is large compared to its variance, it takes on the value zero with probability approaching zero.

For an integer n, let ν(n) be the number of primes that divide n. Then we can use Chebyshev's inequality to prove a result which says that "most" integers have approximately $lnlnn$prime factors.i.e.$|V(x) - lnlnn| = c\sqrt{lnlnn} + 10$

A set of positive integers $\{x_1, \ldots, x_k\}$is said to have distinct sums if all sums $\Sigma_{i \in S} x_i$, $S \subseteq \{1\text{........}, k\}$ are distinct.i.e. All sums of all subsets of the integer set are distinct.

If we define $f(n) = max\{k \in N : \exists\{x_1, \ldots. x_k\}\}$ with distinct sums, Chebyshev's inequality can be used to bound f(n) :
$lg[n] + 1 \leq f(n) \leq lg[n] + \frac{1}{2}lglgn + C$

PS:
Although , proofs to all the above are better DIY , i ll be happy to help with proving any of the above.

[1]e.g. : Suppose that we roll a standard fair die 100 times. Let X be the sum of the numbers that appear over the 100 rolls,you can use Chebyshev's inequality to bound $P[|x - 350| \geq 50]$.

Justin Rising, PhD in statistics
Answered Feb 21, 2013

You can use it to generate distribution-free confidence intervals. They're very wide compared to any CI based on the right distribution, but they always work.