Chebychev’s Theorem gives bounds on how spread out a probability distribution can be from the mean, in terms of the standard deviation. More precisely, if \(X\) is a random variable with mean \(\mu\) and standard deviation \(\sigma\), then \[  
P(|X – \mu| \ge k \sigma) \le \frac {1}{k^2}.  
\]

When I was an undergraduate at Texas, I noticed that this has the following implication. Suppose \(x\_1,\ldots,x\_n\) is any finite set of numbers. Create a random variable \(X\) so that \(P(X = x\_i) = 1/n\) for all \(1\le i \le n\). Then, the mean of \(X\) is \(\mu = \overline{x}\), and the standard deviation of \(X\) is \(s\_X = s \sqrt{\frac{n-1}{n}}\), where \(s\) is the sample distribution. By Chebychev’s Theorem with \(k = n\), \[P(|X – \overline{x}| \ge \sqrt{n} s\_X) \le \frac{1}{n}\] Now, since the probability of *each \(x\_i\)* is \(\frac 1n\), this tells us that *all* of the points \(x\_i\) are within \(\sqrt{n} s\_X\) if the mean of the \(x\_i\).

I totally didn’t believe that was true when I was 20 years old, so I spent (way too much) time computing examples. What I noticed was that not only did it seem to be true, but I couldn’t even find examples where \(\sqrt{n} s\_X\) was even necessary.

Fast forward 30 years later, and I decided to revisit this problem. Let’s consider the case when \(n = 4\) and do some simulations. I will take a bunch of random samples of size 4, compute \(\mu\) and \(s\_X\) and the maximum distance from \(\mu\) to any of the \(x\_i\). Then, we’ll compare it to \(s\_X\) to see how far the farthest point away from the mean is.

set.seed(9132019)

vec <- 1:4

probs <- rep(1/4, 4)

mu <- mean(vec)

sd <- sqrt(sum(probs \* (vec - mu)^2))

best <- max(abs(vec - mu))/sd

best\_vec <- vec

for(i in 1:100000) {

vec <- rnorm(4)

mu <- mean(vec)

sd <- sd(vec) \* sqrt(3/4)

if(max(abs(vec - mu))/sd > best) {

best <- max(abs(vec - mu))/sd

best\_vec <- vec

}

}

best\_vec

## [1] -0.1465015 -0.1494259 1.5503518 -0.1461181

best

## [1] 1.732048

We see a couple of things. First, we see that the extreme case is when all of the values are equal except for one which is different, and second, that all of the data is within 1.732 standard deviations of the mean, when Chebychev would’ve predicted that all of the data is within 2 standard deviations of the mean. What gives? Laguerre-Samuelson, that’s what!

**Theorem** Let \(x\_1,\ldots,x\_n\) be any \(n\) real numbers with mean \(\mu\) and standard deviation \(s\) (with \(n\) in the downstairs). Then, all \(n\) numbers lie within the interval \([\mu – \sqrt{n-1}s, \mu + \sqrt{n-1}s]\).

As hinted at above, this theorem can be considered a sharp version of Chebychev in the case that the rv \(X\) consists of finitely many equally likely values.