Simulating null distributions

Sayani Gupta

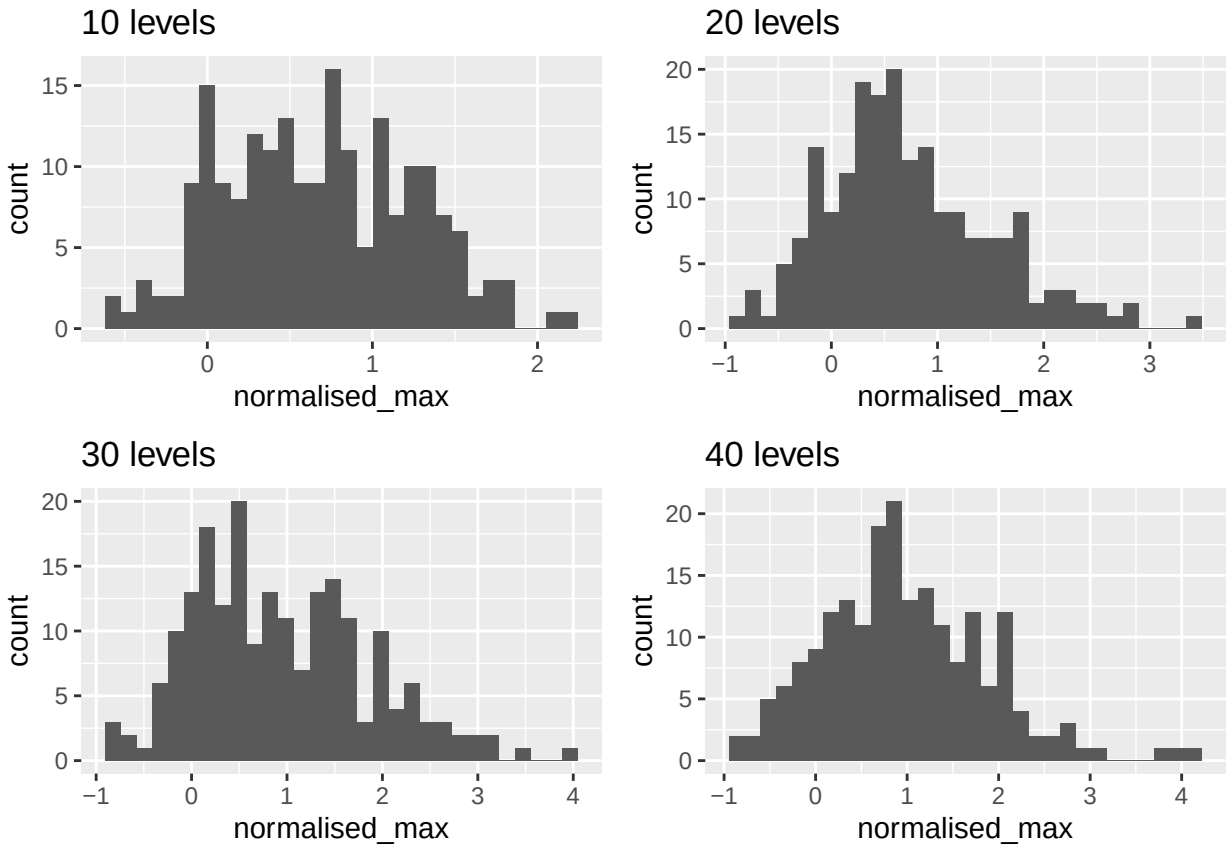
19/08/2020

If we suppose that MMPD is useful to normalise for the number of levels, then distribution of MMPD should be same for different levels, provided the response variable for all the levels comes from the same distribution. Following scenarios are considered to test that by breaking down the problem and checking - A) Does normalising works for different x-axis levels? B) if yes, then does further normalisation using median works for different facet categories?

A) To test if normalisation works for different x-axis levels:

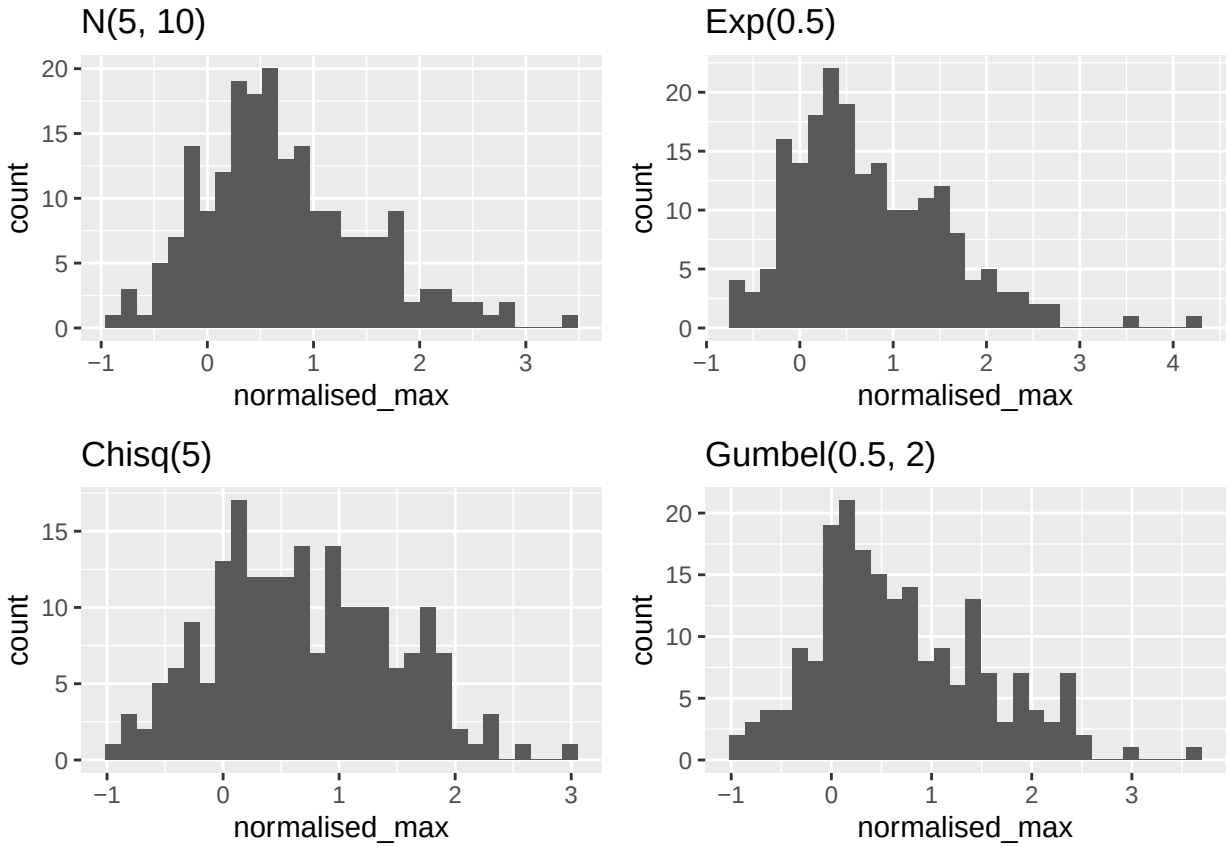
For x-axis, first maximum of pairwise JS distances are computed and it is normalised using Fisher–Tippett–Gnedenko theorem. Different x-axis levels viz, 10, 20, 30, 40 are considered and data is simulated 200 times for each of these cases to learn about the distribution of these normalised distances by plotting their histograms.

All levels drawn from N(5, 10) distribution



From the above graphs, we could see that irrespective of the levels, the distribution of the normalised distances look like a normal distribution when the underlying observations are drawn from a N(5, 10) distribution. Let us check, if the same holds true if the underlying distributions are from Exp(0.5), Chi-squared(5), Gumbel(0.5, 2). Here, we fix the number of levels to 20.

Different distributions with 20 levels



B) To test if further normalisation using median works for different facet categories?

If the normalisation along x-axis is working, we can vary the levels across facets keeping levels of x-axis constant to see if normalisation along facets work or not.

For facets, first all normalised maximum distances are obtained and then their median is taken and futher divided by log(number of facet levels). Different facet levels viz, 10, 20, 30, 40 are considered for a fixed x-axis level and data is simulated 200 times for each of these cases to learn about the distribution of these normalised distances through a histogram.

All levels drawn from N(5, 10) distribution