vary_distribution

Sayani Gupta

18/12/2020

Behavior of raw wpd (weighted pairwise distances)

Simulation study

We conduct simulations to understand the behavior of raw WPD for different parameters l_x (levels of x-axis), l_f (levels of facets), λ (tuning parameter), ω (increment in each panel design) $data$ (data sets), $sample - size$ (tuning parameter), $nsim$ (number of simulations), $nperm$ (number of permutations (might not need depending on if norm works) and designs
D_{null} (No difference in distribution)
D_{var_f} (Difference in distribution only across facets)
D_{var_x} (Difference in distribution only across x-axis)
$D_{var_{all}}$ (Difference in distribution in both facets and x-axis)

Location and scale of the distribution under null design

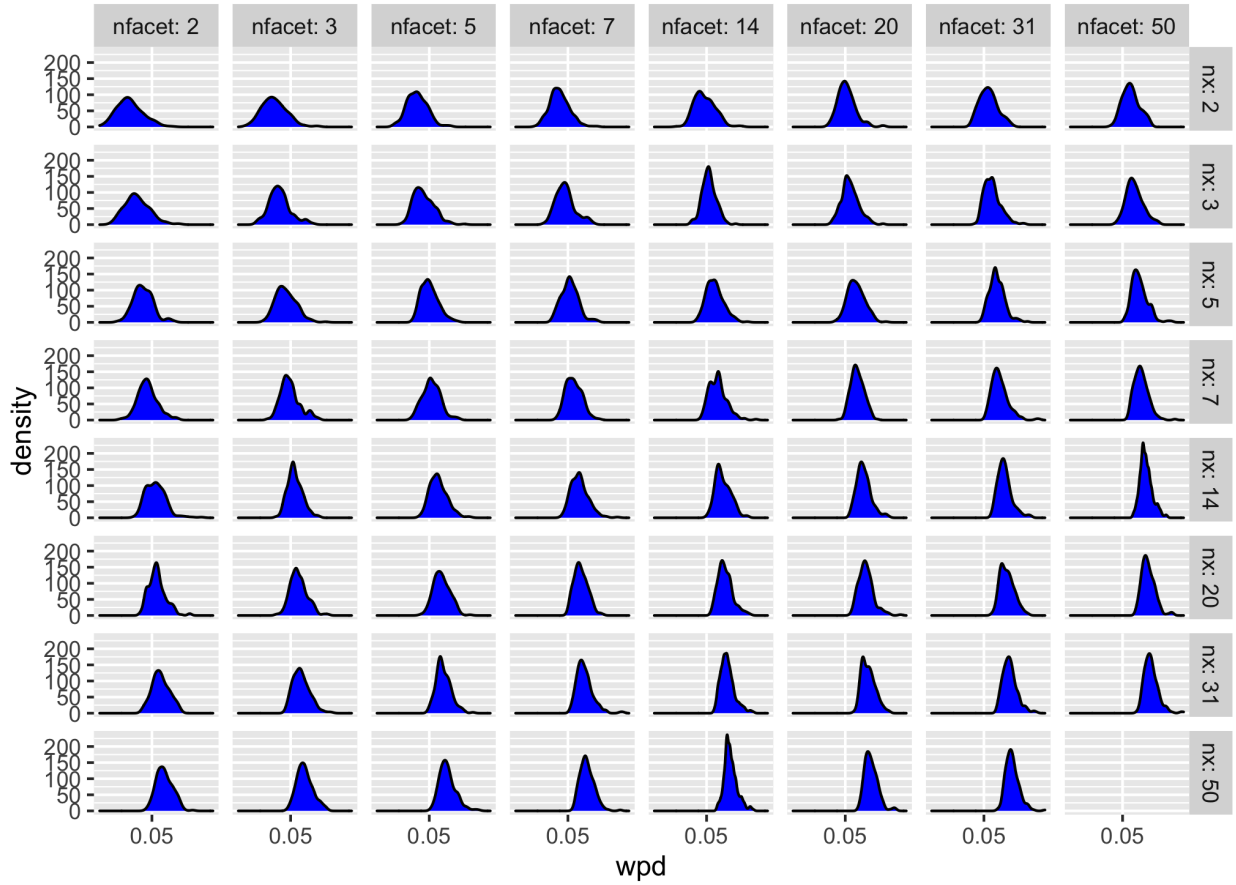
Null design only where all combinations are equal for fixed values of μ and σ Dist 1: $N(0, 1)$ individual graph Dist 2: $N(\mu, 1)$ (faceted with different values of μ: not yet just for one value now) Dist 3: $N(0, \sigma)$ (faceted with different values of σ: not yet just for one value now) Dist 4: $N(\mu, \sigma)$ faceted with different values of mu and sigma, for all levels of x and facet to see how value of wpd changes with that of change in parameters.

Objective: 1. Behavior of wpd under different null designs N(0,1), N(5, 1), Gamma(0.5,1), Gamma(2,1) for different nx and nfacets. Here, we have to fix a value of lambda = 0.67.

Assumptions: There is no difference in distribution between any facet or x-category nsim = 200 lambda = 0.67

Questions: - How raw value of wpd changes with different nx and nfacet for different location and scale of a Normal and non-normal distribution

Standard normal distribution



How mean and sd changes with increasing x-axis and facet levels?

Comparing normal distributions with different means

Gamma Distribution with different locations

Gamma Distribution with different locations after quantile transformation

Normalise by dividing log(number of distances)

How to design the alternate designs when distributions are actually different?

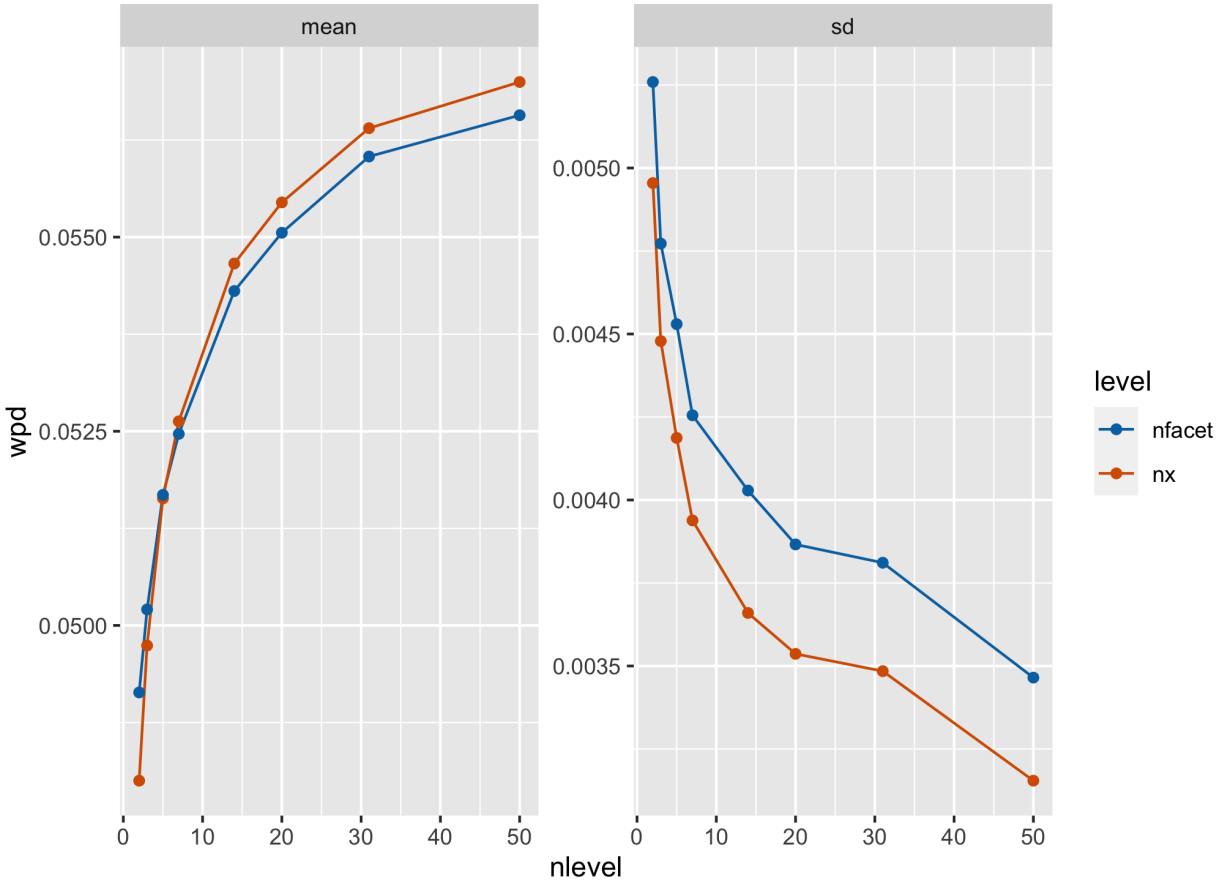


Figure 1: Movement of mean and sd for raw wpd is shown for different number of levels (nlevel) of x-axis and facets through line plots. Mean increases and sd decreases more sharply for increasing x-axis levels as compared to facet levels. It seems like both mean and standard deviation are affected more by change in the x-axis levels than the facet levels.

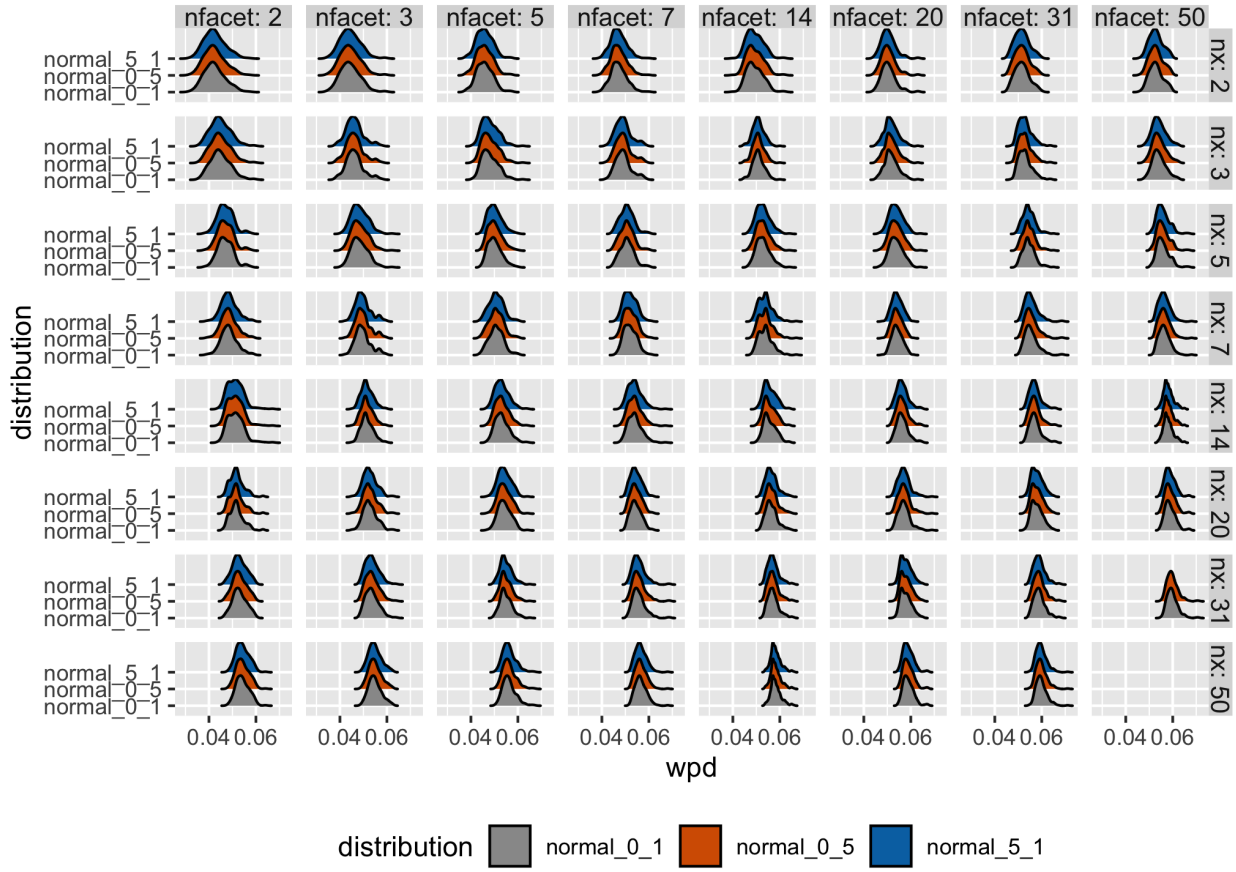


Figure 2: Ridge plots of raw wpd is shown for N(0,1) and N(5,1) distribution. The densities change across different facet and x levels but look same for the two distributions, which implies wpd value is unaffected by the change in mean value of the normal distribution

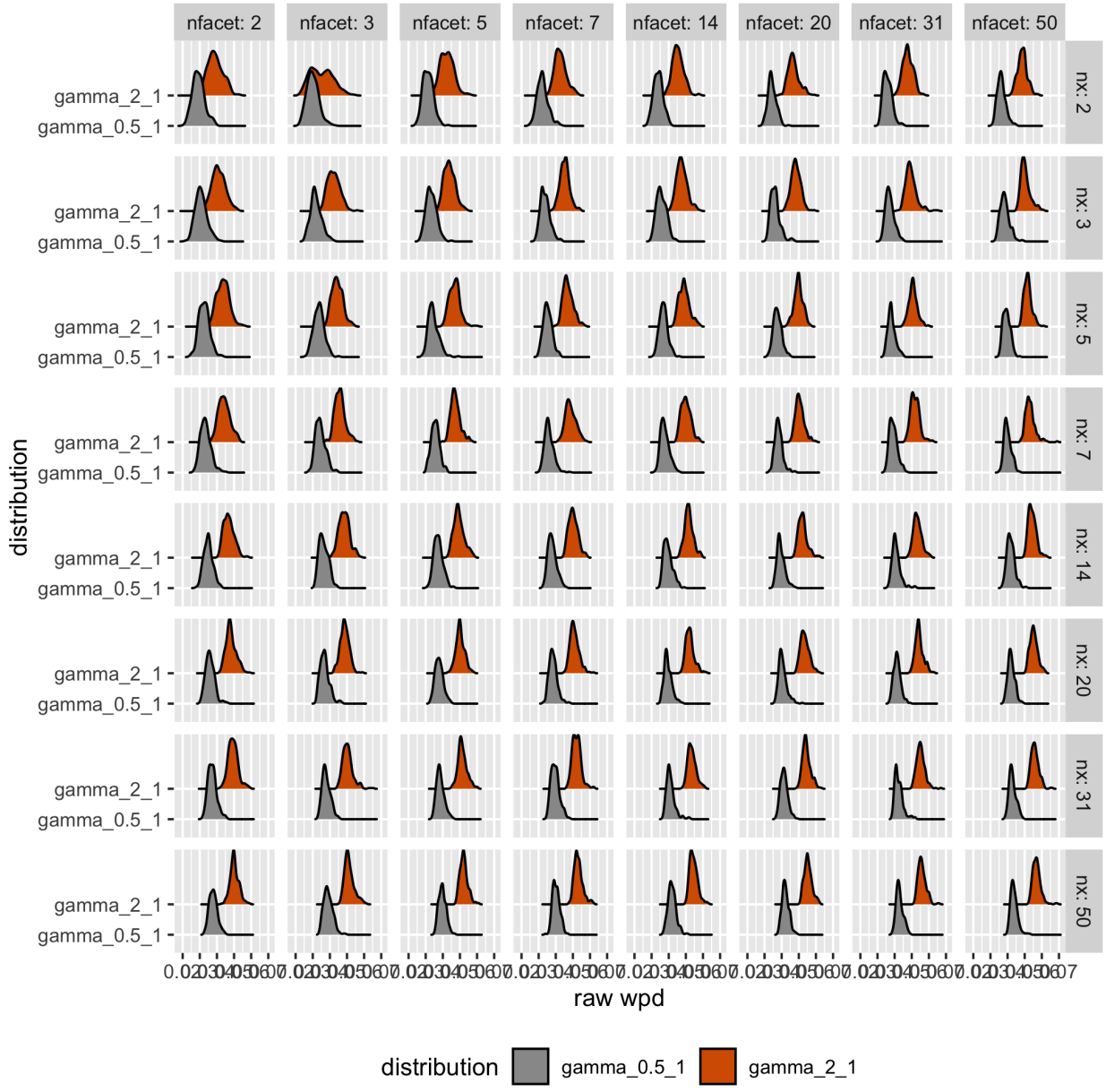


Figure 3: Ridge plots of raw wpd is shown for G(0.5,1) and G(2,1) distribution. The densities change across different facet and x levels and also for the two distributions, which implies wpd value is affected by the change in location value when distributions are not normal.

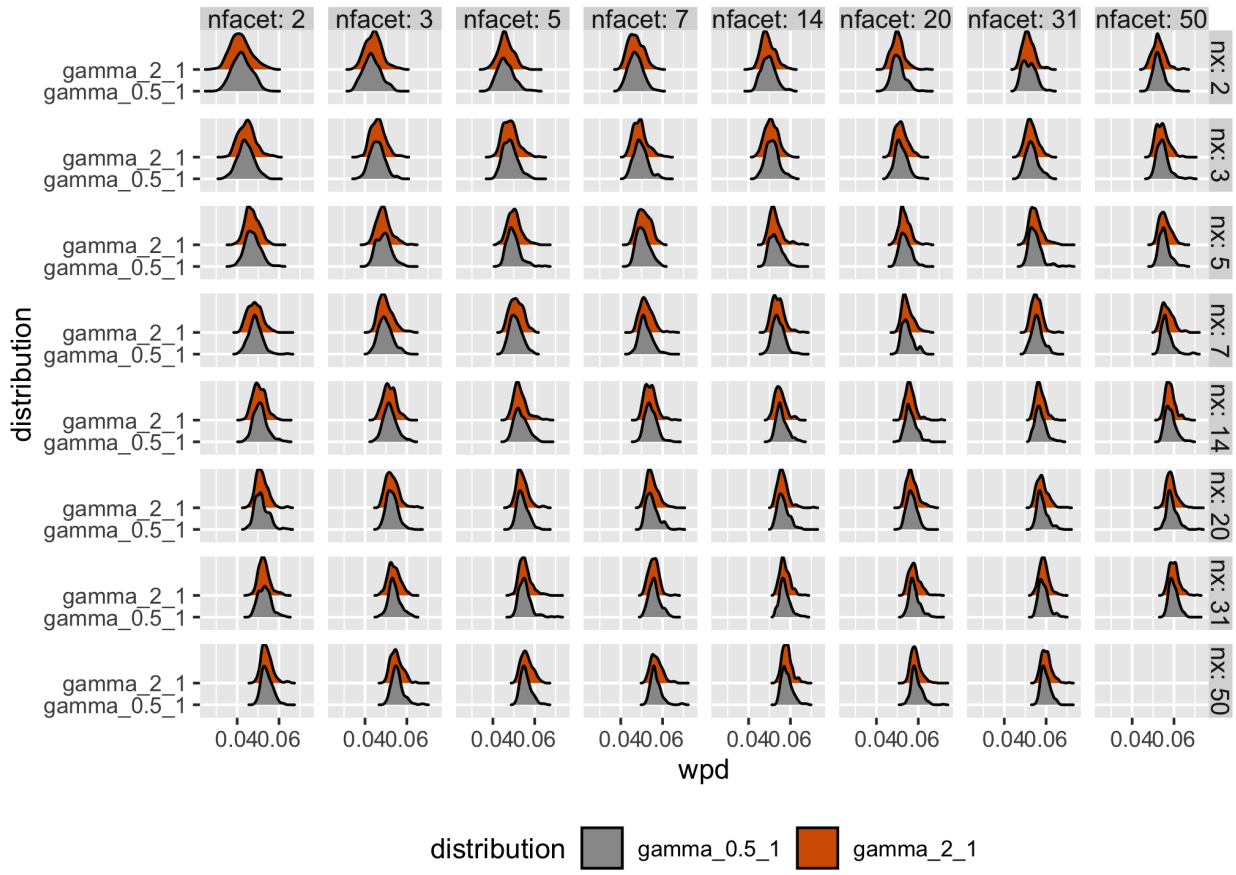


Figure 4: Ridge plots of raw wpd is shown for G(0.5,1) and G(2,1) distribution after quantile transformation looks similar and hence is unaffected by change in location

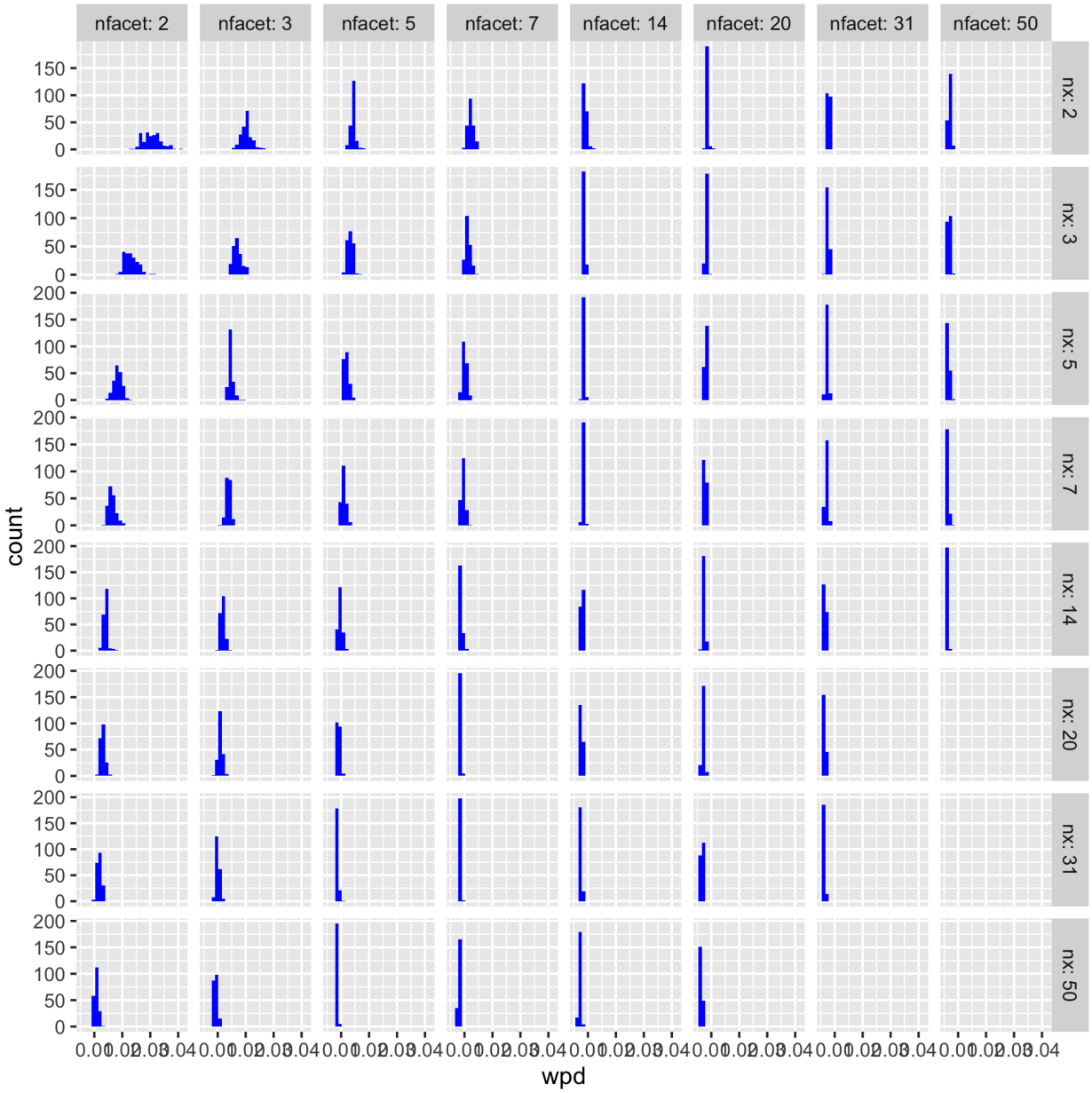


Figure 5: Density plots of raw wpd is shown for N(0,1) distribution after dividing max by log(number of distances). The densities are reduced to zero and varies. So this normalisation approach is not efficient. The raw wpd values are already very low and then dividing them by logarithm of a large number brings it back to zero.