Confidence intervals

The differential privacy library supports confidence intervals to assess the scale of the noise that has been added to a metric during the anonymization process and narrow down its original, true value. Given a noised metric M and a confidence level of 1 - alpha, the differential privacy library computes a confidence interval [L, R] that contains the raw metric m (where m is the metric after contribution bounding, but before the application of noise) with a probability of at least 1 - alpha, i.e. Pr[L ≤ m ≤ R] ≥ 1 - alpha.
The computation is performed purely based on the noisy metric M and on privacy parameters such as epsilon, delta, sensitivities and the contribution bounds. As a result, no privacy budget is consumed for the computation of confidence intervals.
Note that the confidence intervals provided by the library do not take the effects of contribution bounding into account. For instance, consider a bounded sum over the raw entries 1, 1, 2, 5, 14, 42 and 132 with a lower bound of 10 and an upper bound of 20. The confidence intervals provided by the library will be based on the bounded sum, i.e. m = 94, rather than the actual sum, i.e. m ≠ 197.

Why is alpha used instead of confidence level

1) Using alpha for confidence intervals provides more accuracy for confidence levels which are closer to 1, which is the parameterization we expect.

2) For float values, very high confidence levels can only be represented in terms of alpha. For example: $1 - 10^{-10},\ 1 - 20^{-15}$

3) The alpha error is the standard terminology for confidence intervals, therefore it should be at least equally accessible for the users.