Due to the concepts of underrepresented and overrepresented DFPs is not clearly explained, I understand them as follows:

1.Underrepresented DFPs. In sample log, the actual frequency of DFPs is lower than its expected frequency, i.e., this DFP is not fully represented in sample log.

2.Overrepresented DFPs. In sample log, the actual frequency of DFPs is higher than its expected frequency, i.e., this DFP is over represented in sample log.

Under the above premise, in the fourth row and third column of the table obtained in step 4 of Fig. 2, the value of trace ⟨a,c,e,f,g,h⟩ should be 5 instead of 4. Because this trace contains five DFPs, where underrepresented DFPs = 5, overrepresented DFPs = 0. Therefore, the difference between them is 5.

| Variant | Decimal place | Difference between the number of underrepresented and overrepresented DFPs in preliminary Sample 1 | Normalized difference value |
|---|---|---|---|
| a,b,e,g,h | 0 | 4 | 1 |
| a,b,e,g,i | 6 | 4 | 1 |
| a,b,e,f,e,g,i | 4 | 6 | 1 |
| a,c,e,f,g,h | 4 | 4 | 1 |
| a,c,e,g,i | 4 | 4 | 1 |
| a,c,e,f,e,g,i | 2 | 6 | 1 |

Similarly, the values in the third and fourth columns of Fig. 3 using the above method are shown in blue font in the following figure.

For example, combining Sample2, the underrepresented DFPs of trace ⟨a,b,e,g,i⟩ is ⟨g,i⟩, and its overrepresented DFPs is NULL, so the difference value is 1. Its normalized difference value is 1/4=0.25, where the denominator is the number of DFPs contained in trace. The underrepresented DFPs of trace⟨a,b,e,f,e,g,i⟩ are ⟨e,f⟩⟨f,e⟩⟨g,i⟩, and the overrepresented DFPs is NULL. Therefore, the difference value is 3. The value of the fourth column is 0.5.