Sensei: Sample size determination for comparing abundance of cells under different conditions

# Notations and assumptions

|  |  |  |
| --- | --- | --- |
| Parameter | Notation | Annotation |
| Number of cells in type A in sample |  | A random variable denoting the number of cells of cells in cell type A. |
|  |  | A random variable denoting the proportion of cells in cell type A. |
| Proportion of cells of type A |  | A random variable denoting the observed proportion of cells in cell type A. |
| Expected proportion of cells of type A |  | Proportion of cells in each sample.  A random variable. |
| Misclassification rate |  | False negative and false positive in clustering |
| Number of samples under condition |  |  |
| Number of cells in sample |  | It should be the number of cells of type A plus reliable background cell types. Other cell types whose cell number vary should not be included to avoid the composition effect |
| Beta parameters |  | Parameters for beta distribution. |
|  |  |  |
| Alpha |  | False positive rate, aka. Critical value |
| Beta |  | False negative rate, aka |

All parameters should be specified except for the ones shaded in green. The green ones are assumptions.

# Hypothesis

# Derivation

## Distribution of

Because

Thus,

We now assume that the beta binomial distribution can be approximated by a normal distribution.



## Distribution of

Then, we can give the distribution of the t-value.

Where

Where

Now for a two-sided test

|  |  |  |
| --- | --- | --- |
| Test | Requires | Critical value |
| One-sided |  |  |
| Two-sided |  |  |

Thus, the false negative rate is

# More about beta distribution

Note that for Beta distributed

Thus,

In fact,

So

Thus