**ZTP-Bin implementation**

*Date: Aug. 16, 2017*

**Ruofei Du, Zhide Fang**

Contents

[Functions 1](#_Toc490641168)

[ZTPBin 1](#_Toc490641169)

[mle.lambda 1](#_Toc490641170)

[Example code 2](#_Toc490641171)

# Functions

There are two R functions being provided: **ZTPBin** and **mle.lambda**.

## ZTPBin

**Description**

This function carries out the zero-truncated Poisson and Binomial modelling for length and conservation bias correction in functional profiling of a metagenomics sample generated by short NGS reads.

**Usage**

ZTPBin (estParam, crtCount, nonCDS='Non-CDS')

**Arguments**

|  |  |
| --- | --- |
| estParam | a dataframe with three columns (COG, lambda, p) containing the estimated lambda and p for a COG |
| crtCount | a dataframe with two columns (COG, count) containing the corrected counts to a COG |
| nonCDS | the name for the non-coding sequence (Non-CDS) appearing in the COG column of crtCount |

**Value**

a dataframe that contains the predicted number of CDS associated to a COG (column NCDS) and the relative abundance of COG in the query dataset (column relAbun).

## mle.lambda

**Description**

This function applies Newton-Raphson method to obtain a maximum likelihood estimator (MLE) of the parameter lambda in a zero-truncated Poisson distribution.

**Usage**

mle.lambda (u, lambda0)

**Arguments**

|  |  |
| --- | --- |
| u | a numeric vector containing the number of reads fragmented from a CDS that is associated to the COG of study |
| lambda0 | an initial value of lambda set for solving the ML equation |

**Value**

the MLE of lambda obtained for the COG of study

# Example code

To facilitate a statistician to fit the ZTP-Bin, using the given R functions, on a metagenomic sample generated by a NGS technology other than 454 platform, an example code is provided to walk one through the process. The datasets used in the example code is saved under ./data folder.

One should estimate the two parameters (lambda, p) from the simulated datasets (Simu\_Ui, Simu\_blastBH) by using the mle.lambda function; and then apply ZTPBin function on a real dataset with the input of these estimated parameters. Detailed comments are provided in the code.