

# Package: ZIPFA (via r-universe)

October 18, 2024

**Type** Package

**Title** Zero Inflated Poisson Factor Analysis

**Version** 0.8.1

**Date** 2020-03-30

**Author** Tianchen Xu [aut, cre]

(<<https://orcid.org/0000-0002-0102-7630>>), Ryan T. Demmer  
[aut], Gen Li [aut]

**Maintainer** Tianchen Xu <tx2155@columbia.edu>

**Description** Estimation methods for zero-inflated Poisson factor analysis (ZIPFA) on sparse data. It provides estimates of coefficients in a new type of zero-inflated regression. It provides a cross-validation method to determine the potential rank of the data in the ZIPFA and conducts zero-inflated Poisson factor analysis based on the determined rank.

**URL** <https://zjph602xtc.github.io/ZIPFA/>,  
<https://arxiv.org/abs/1910.11985>

**BugReports** <https://github.com/zjph602xtc/ZIPFA/issues>

**Depends** R (>= 3.2.0)

**Imports** Matrix, doParallel, foreach, optimx, trustOptim

**License** GPL (>= 2)

**Encoding** UTF-8

**LazyData** no

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Date/Publication** 2020-03-31 20:30:02 UTC

**Repository** <https://zjph602xtc.r-universe.dev>

**RemoteUrl** <https://github.com/cran/ZIPFA>

**RemoteRef** HEAD

**RemoteSha** ea6e3724a13d8b8b89b87a56e59190e5ac9d6642

## Contents

cv_ZIPFA . . . . .	2
EMzeropoisson_mat . . . . .	3
X . . . . .	4
ZIPFA . . . . .	5
<b>Index</b>	<b>7</b>

---

cv_ZIPFA	<i>Cross validation for Zero Inflated Poisson factor analysis</i>
----------	---

---

### Description

To conduct a cross validation for Zero Inflated Poisson factor analysis to find the number of factors.

### Usage

```
cv_ZIPFA(X, k, fold, tau = 0.1, cut = 0.8, tolLnlikelihood = 5e-4,
         iter = 20, tol = 1e-4, maxiter = 100, initialtau = 'iteration',
         Madj = TRUE, display = TRUE, parallel = FALSE)
```

### Arguments

X	The matrix to be decomposed.
k	A vector containing the number of factors to try.
fold	The number of folds used in cross validation.
tau	Initial tau value to fit. Will be overwritten by the first value in initial argument.
cut	To delete columns that has more than 100('Cut')% zeros. Cut = 1, if no filtering.
tolLnlikelihood	The max percentage of log likelihood differences in two iterations.
iter	Max iterations.
initialtau	A character specifying the way to choose the initial value of tau at the beginning of EM iteration. <i>stable</i> : estimate tau from fitted beta in last round; <i>initial</i> : always use the initially assigned tau in tau or initial. Use the default tau = 0.1 if 'initial' is empty. <i>iteration</i> : use fitted tau in last round.
tol	Percentage of l2 norm change of [tau beta].
maxiter	Max iteration number in the zero inflated poisson regression.
Madj	If TRUE then adjust for relative library size M.
display	If TRUE display the fitting procedure.
parallel	Use doParallel and foreach package to accelerate.

**Details**

The function conducts cross validation on the zero-inflated Poisson factor analysis to determine the rank.

**Value**

The function returns a matrix. Each row the CV likelihood of one fold. Each column is the result of number of factors in k.

**Author(s)**

Tianchen Xu

**Examples**

```
data(X)
cv_ZIPFA(X, fold = 10, k = c(3,4))
```

---

EMzeropoisson\_mat      *Zero Inflated Poission Regression*

---

**Description**

The zero inflated poission regression model.

**Usage**

```
EMzeropoisson_mat(data, tau = 0.1, initial = NULL, initialtau = 'iteration',
  tol = 1e-4, maxiter = 100, Madj = FALSE, m = NULL,
  display = TRUE, intercept = TRUE)
```

**Arguments**

data	A matrix with the first columns is y and the rest columns are x.
tau	Initial tau value to fit. Will be overwritten by the first value in initial argument.
initial	A list of initial values for the fitting. c(tau beta).
initialtau	A character specifying the way to choose the initial value of tau at the beginning of EM iteration. <i>stable</i> : estimate tau from fitted beta in last round; <i>initial</i> : always use the initially assigned tau in tau or initial. Use the default tau = 0.1 if 'initial' is empty. <i>iteration</i> : use fitted tau in last round.
tol	Percentage of l2 norm change of [tau beta].
maxiter	Max iteration number.

<code>Madj</code>	If TRUE then adjust for relative library size M.
<code>m</code>	A vector containing relative library size M.
<code>display</code>	If TRUE display the fitting procedure.
<code>intercept</code>	If TRUE then the model contains an intercept.

### Details

The function estimates the coefficients in a new type of zero-inflated Poisson regression where the underlying Poisson rate is negatively associated with true zero probability.

### Value

The function turns a matrix. Each row is fitted value in each iteration. The last row the final result. The first column is fitted tau. If `intercept` is true, then the second column is the intercept, and the rest columns are other coefficients. If `intercept` is false, the rest columns are other coefficients.

### Author(s)

Tianchen Xu

### Examples

```
n = 5000;
x1 = rnorm(n);
x2 = rnorm(n);
lam = exp(x1 - 2*x2 + 1.5);
y = rpois(n, lam)
tau = .75
p = 1/(1+lam^tau);
Z = rbinom(n, 1, p);
y[as.logical(Z)] = 0;

res = EMzeropoisson_mat(matrix(c(y,x1,x2),ncol=3), Madj = FALSE, intercept = TRUE)
```

---

X

*A simulated data X.*

---

### Description

For example run.

### Usage

```
data("X")
```

**Format**

The format is: int [1:200, 1:100] 1 1 1 0 0 0 0 0 2 ... - attr(\*, "dimnames")=List of 2 ..\$ : NULL  
 ..\$ : chr [1:100] "V1" "V2" "V3" "V4" ...

**Examples**

```
data(X)
```

---

 ZIPFA

*Zero Inflated Poisson factor analysis*


---

**Description**

To conduct a Zero Inflated Poisson factor analysis.

**Usage**

```
ZIPFA(X, k, tau = 0.1, cut = 0.8, tolLnlkelihood = 5e-4,  
      iter = 20, tol = 1e-4, maxiter = 100, initialtau = 'iteration',  
      Madj = TRUE, display = TRUE, missing = NULL)
```

**Arguments**

X	The matrix to be decomposed.
k	The number of factors.
tau	Initial tau value to fit. Will be overwritten by the first value in initial argument.
cut	To delete columns that has more than 100('Cut')% zeros. Cut = 1, if no filtering.
tolLnlkelihood	The max percentage of log likelihood differences in two iterations.
iter	Max iterations.
initialtau	A character specifying the way to choose the initial value of tau at the beginning of EM iteration. <i>stable</i> : estimate tau from fitted beta in last round; <i>initial</i> : always use the initially assigned tau in tau or initial. Use the default tau = 0.1 if 'initial' is empty. <i>iteration</i> : use fitted tau in last round.
tol	Percentage of l2 norm change of [tau beta].
maxiter	Max iteration number in the zero inflated poisson regression.
Madj	If TRUE then adjust for relative library size M.
display	If TRUE display the fitting procedure.
missing	Reserved for cv_ZIPFA.

**Details**

The function conducts a zero-inflated Poisson factor analysis where the underlying Poisson rate is negatively associated with true zero probability.

**Value**

<code>tau</code>	Fitted tau value.
<code>Ufit</code>	A list containing fitted U matrix in each iteration. The last one is the final fit.
<code>Vfit</code>	A list containing fitted V matrix in each iteration. The last one is the final fit.
<code>itr</code>	Number of iterations.
<code>Likelihood</code>	The likelihood for the training data.
<code>CVLikelihood</code>	The likelihood for the testing data (if applicable)

**Author(s)**

Tianchen Xu

**Examples**

```
data(X)  
ZIPFA(X, k = 3)
```

# Index

\* **cross validation**

cv\_ZIPFA, 2

\* **datasets**

X, 4

\* **factor analysis**

EMzeropoisson\_mat, 3

ZIPFA, 5

\* **zero inflated**

EMzeropoisson\_mat, 3

ZIPFA, 5

cv\_ZIPFA, 2

EMzeropoisson\_mat, 3

X, 4

ZIPFA, 5