Package: bzinb (via r-universe)

September 3, 2024

Type Package

Title Bivariate Zero-Inflated Negative Binomial Model Estimator

Version 1.1.0

Author Hunyong Cho, Chuwen Liu, Jinyoung Park, Di Wu

Maintainer Hunyong Cho <hunycho@live.unc.edu>

Description Provides a maximum likelihood estimation of Bivariate Zero-Inflated Negative Binomial (BZINB) model or the nested model parameters. Also estimates the underlying correlation of the a pair of count data. See Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation) for details.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Imports Rcpp (>= 0.11.0), RcppArmadillo

LinkingTo Rcpp, BH, RcppArmadillo

Repository https://hunyong.r-universe.dev

RemoteUrl https://github.com/hunyong/bzinb

RemoteRef HEAD

RemoteSha 8c4043f451a21f18656aa51025470b8e235bc125

Contents

---

bnb *The bivariate negative binomial distribution*

Description

random generation (rbnb), maximum likelihood estimation (bnb), and log-likelihood. (lik.bnb) for the bivariate negative binomial distribution with parameters equal to (a0, a1, a2, b1, b2).

Usage

```
lik.bnb(xvec, yvec, a0, a1, a2, b1, b2, param = NULL)

rbnb(n, a0, a1, a2, b1, b2, param = NULL)

bnb(
  xvec,
  yvec,
  em = TRUE,
  tol = 1e-08,
  maxiter = 50000,
  vcov = TRUE,
  initial = NULL,
  showFlag = FALSE
)
```

Arguments

xvec, yvec	a pair of bnb random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
a0, a1, a2	shape parameters of the latent gamma variables. must be positive.
b1, b2	scale parameters for the latent gamma variables. must be positive.
param	a vector of parameters ((a0, a1, a2, b1, b2)). Either param or individual parameters (a0, a1, a2, b1, b2) need to be provided.
n	number of observations.
em	if TRUE in bnb, EM algorithm is applied. Otherwise, direct opitmation is used.
tol, maxiter, vcov, initial, showFlag	optional arguments applied only when em is TRUE in bnb.

Value

- rbnb gives a pair of random vectors following BNB distribution.
- bnb gives the maximum likelihood estimates of a BNB pair. Standard error and covariance matrix are provided when em is TRUE.
- lik.bnb gives the log-likelihood of a set of parameters for a BNB pair.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Examples

```
# generating a pair of random vectors
set.seed(1)
data1 <- rbnb(n = 100, a0 = 2, a1 = 1, a2 = 1,
                b1 = 1, b2 = 1)

lik.bnb(xvec = data1[, 1], yvec = data1[ ,2],
          a0 = 1, a1 = 1, a2 = 1, b1 = 1, b2 = 1)

bnb(xvec = data1[,1], yvec = data1[,2], showFlag = FALSE)
```

---

bp *The bivariate poisson distribution*

Description

random generation (rbp), maximum likelihood estimation (bp), and log-likelihood. (lik.bp) for the bivariate Poisson distribution with parameters equal to (m0, m1, m2).

Usage

```
lik.bp(xvec, yvec, m0, m1, m2, param = NULL)

rbp(n, m0, m1, m2, param = NULL)

bp(xvec, yvec, tol = 1e-06)
```

Arguments

xvec, yvec	a pair of bp random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
m0, m1, m2	mean parameters of the Poisson variables. They must be positive.
param	a vector of parameters ((m0, m1, m2)). Either param or individual parameters (m0, m1, m2) need to be provided.
n	number of observations.
tol	tolerance for judging convergence. tol = 1e-8 by default.

Value

- rbp gives a pair of random vectors following BP distribution.
- bp gives the maximum likelihood estimates of a BP pair.
- lik.bp gives the log-likelihood of a set of parameters for a BP pair.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Kocherlakota, S. & Kocherlakota, K. (1992). Bivariate Discrete Distributions. New York: Marcel Dekker.

Examples

```
# generating a pair of random vectors
set.seed(1)
data1 <- rbp(n = 20, m0 = 1, m1 = 1, m2 = 1)

lik.bp(xvec = data1[, 1], yvec = data1[ ,2],
          m0 = 1, m1 = 1, m2 = 1)

bp(xvec = data1[,1], yvec = data1[,2])
```

---

bzinb *The bivariate zero-inflated negative binomial distribution*

Description

random generation (rbzinb), maximum likelihood estimation (bzinb), and log-likelihood. (lik.bzinb) for the bivariate zero-inflated negative binomial distribution with parameters equal to (a0, a1, a2, b1, b2, p1, p2, p3, p4).

Usage

```
lik.bzinb(xvec, yvec, a0, a1, a2, b1, b2, p1, p2, p3, p4, param = NULL)

rbzinb(n, a0, a1, a2, b1, b2, p1, p2, p3, p4, param = NULL)

bzinb(
  xvec,
  yvec,
  initial = NULL,
  tol = 1e-08,
  maxiter = 50000,
  showFlag = FALSE,
  vcov = FALSE
)
```

Arguments

xvec, yvec	a pair of bzinb random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
a0, a1, a2	shape parameters of the latent gamma variables. They must be positive.
b1, b2	scale parameters for the latent gamma variables. They must be positive.
p1, p2, p3, p4	proportions summing up to 1 (p1 + p2 + p3 + p4 = 1). p1 is the probability of both latent Poisson variables being observed. p2 is the probability of only the first Poisson variables being observed. p3 is the probability of only the second Poisson variables being observed, and p4 is the probability of both Poisson variables being dropped out.
param	a vector of parameters ((a0, a1, a2, b1, b2, p1, p2, p3, p4)). Either param or individual parameters (a0, a1, a2, b1, b2, p1, p2, p3, p4) need to be provided.
n	number of observations.
initial	starting value of param for EM algorithm, a vector of nine values.
tol	tolerance for judging convergence. tol = 1e-8 by default.
maxiter	maximum number of iterations allowed. tol = 50000 by default.
showFlag	if TRUE, the updated parameter estimates for each iteration are printed out. If a positive integer, the updated parameter estimates for iterations greater than showFlag are printed out.
vcov	if TRUE, the variance-covariance matrix and information matrix are returned.

Details

EM theoretically guarantees higher likelihood at each iteration than that of previous iterations. See Dempster, Laird, and Rubin (1977). This guarantee comes with an assumption that there is no numerical error in conditional likelihood maximization at each iteration. Small errors can cause decreasing likelihood especially when the iterations reach the point of convergence. Due to this technical error, the EM continues after it reaches the maximum likelihood point (up to 100 iterations). However, the final estimate being returned is the parameter values at the maximum likelihood.

Value

- rbzinb gives a pair of random vectors following BZINB distribution.
- bzinb gives the maximum likelihood estimates of a BZINB pair.
  - rho estimate and standard error of the underlying correlation (ρ) and ($logit(\rho)$)
  - coefficients estimate and standard error of the BZINB parameters
  - lik log-likelihood of the maximum likelihood estimate
  - iter total number of iterations
  - info information matrix. Provided when vcov is TRUE.
  - vcov variance-covariance matrix. Provided when vcov is TRUE.
- lik.bzinb gives the log-likelihood of a set of parameters for a BZINB pair.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Preisser, J., Liu, C., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.

Examples

```
# generating a pair of random vectors
set.seed(2)
data1 <- rbzinb(n = 100, a0 = 2, a1 = 1, a2 = 1,
                b1 = 1, b2 = 1, p1 = 0.5, p2 = 0.2,
                p3 = 0.2, p4 = 0.1)

lik.bzinb(xvec = data1[, 1], yvec = data1[ ,2],
          a0 = 1, a1 = 1, a2 = 1, b1 = 1, b2 = 1,
          p1 = 0.5, p2 = 0.2, p3 = 0.2, p4 = 0.1)

bzinb(xvec = data1[,1], yvec = data1[,2], showFlag = FALSE)
```

---

bzinb.se *The bivariate zero-inflated negative binomial distribution - Standard error estimation*

Description

Standard errors of the BZINB distribution parameter estimates are calculated based on maximum likelihood estimation. If param is NULL, the parameters are first estimated by bzinb function.

Usage

```
bzinb.se(xvec, yvec, a0, a1, a2, b1, b2, p1, p2, p3, p4, param = NULL, ...)
```

Arguments

xvec, yvec	a pair of bzinb random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
a0, a1, a2	shape parameters of the latent gamma variables. They must be positive.
b1, b2	scale parameters for the latent gamma variables. They must be positive.
p1, p2, p3, p4	proportions summing up to 1 (p1 + p2 + p3 + p4 = 1). p1 is the probability of both latent Poisson variables being observed. p2 is the probability of only the first Poisson variables being observed. p3 is the probability of only the second Poisson variables being observed, and p4 is the probability of both Poisson variables being dropped out.
param	a vector of parameters ((a0, a1, a2, b1, b2, p1, p2, p3, p4)). See bzinb for more detail.
...	Other arguments passed on to bzinb function, when param is NULL.

Value

Standard error of rho, logit.rho, a0, a1, a2, b1, b2, p1, p2, p3, and p4 estimates, variance-covariance matrix (vcov) and information matrix. See bzinb for more detail. iter is NA, if the param is given.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Examples

```
set.seed(1)
data1 <- rbzinb(n = 20, a0 = 1, a1 = 1, a2 = 1,
                b1 = 1, b2 = 1, p1 = 0.5, p2 = 0.2,
                p3 = 0.2, p4 = 0.1)
bzinb.se(xvec = data1[,1], yvec = data1[,2],
         param = c(5.5, 0.017, 0.017, 0.33, 0.36,
                   0.53, 0.30, 0.08, 0.09))
```

bzinbReg *The bivariate zero-inflated negative binomial regression.*

Description

the bivariate zero-inflated negative regression.

Usage

```
bzinbReg(x, ...)

expt.names

## S3 method for class 'formula'
bzinbReg(
  mu.formula,
  nu.formula = ~1,
  data,
  zero.inflation = c("full", "co-ZI", "ZINB-NB", "NB-ZINB", "BNB"),
  tol = 1e-08,
  maxiter = 50000,
  showFlag = FALSE,
  vcov = FALSE,
  initial = NULL
)

## S3 method for class 'bzinbReg'
print(x, digits = max(3L, getOption("digits") - 3L), ...)
```

Arguments

tol	tolerance for judging convergence. tol = 1e-8 by default.
maxiter	maximum number of iterations allowed. tol = 50000 by default.
showFlag	if TRUE, the updated parameter estimates for each iteration are printed out. If a positive integer, the updated parameter estimates for iterations greater than showFlag are printed out.
vcov	if TRUE, the variance-covariance matrix and information matrix are returned.
initial	starting value of param for EM algorithm, a vector of nine values.
y1, y2	a pair of bzinb random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
n	number of observations.

Format

An object of class character of length 12.

Value

- the regression coefficients by the MLE of the BZINB regression model.
  - `coefficients` estimate and standard error of the BZINB parameters
  - `lik` log-likelihood of the maximum likelihood estimate
  - `iter` total number of iterations
  - `info` information matrix. Provided when `vcov` is `TRUE`.
  - `vcov` variance-covariance matrix. Provided when `vcov` is `TRUE`.
- `lik.bzinb` gives the log-likelihood of a set of parameters for a BZINB pair.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Preisser, J., Liu, C., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.

Examples

```
library(devtools)
document()
load_all()
set.seed(1)
dat <- rBzinbData()
bzinbReg(cbind(y1, y2) ~ ., ~ X1, data = dat, maxiter = 10)
print(bzinbReg(cbind(y1, y2) ~ ., ~ X1, data = dat, maxiter = 10))
bzinbReg(cbind(y1, y2) ~ ., ~ X1, data = dat, zero.inflation = "co-ZI", maxiter = 10)
bzinbReg(cbind(y1, y2) ~ ., ~ X1, data = dat, zero.inflation = "ZINB-NB", maxiter = 10)
bzinb(y1 = data1[,1], y2 = data1[,2], showFlag = FALSE)
```

`bzip.a` *The bivariate zero-inflated Poisson distribution (A)*

Description

random generation (`rbzip.a`), maximum likelihood estimation (`bzip.a`), and log-likelihood. (`lik.bzip.a`) for the bivariate zero-inflated Poisson (A) distribution with parameters equal to `(m0, m1, m2, p)`.

Usage

```
lik.bzip.a(xvec, yvec, m0, m1, m2, p, param = NULL)

rbzip.a(n, m0, m1, m2, p, param = NULL)

bzip.a(xvec, yvec, tol = 1e-06, initial = NULL, showFlag = FALSE)
```

Arguments

xvec, yvec	a pair of BZIP (A) random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
m0, m1, m2	mean parameters of the Poisson variables. must be positive.
p	zero-inflation probability
param	a vector of parameters ((m0, m1, m2, p)). Either param or individual parameters (m0, m1, m2, p) need to be provided.
n	number of observations.
tol	tolerance for judging convergence. tol = 1e-8 by default.
initial	starting value of param for EM algorithm, a vector of nine values.
showFlag	if TRUE, the updated parameter estimates for each iteration are printed out. If a positive integer, the updated parameter estimates for iterations greater than showFlag are printed out.

Value

- rbzip.a gives a pair of random vectors following BZIP (A) distribution.
- bzip.a gives the maximum likelihood estimates of a BZIP (A) pair.
- lik.bzip.a gives the log-likelihood of a set of parameters for a BZIP (A) pair.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Li, C. S., Lu, J. C., Park, J., Kim, K., Brinkley, P. A., & Peterson, J. P. (1999). Multivariate zero-inflated Poisson models and their applications. Technometrics, 41, 29-38.

Examples

```
# generating a pair of random vectors
set.seed(1)
data1 <- rbzip.a(n = 20, m0 = 1, m1 = 1, m2 = 1, p = 0.5)

lik.bzip.a(xvec = data1[, 1], yvec = data1[ ,2],
```

```
          m0 = 1, m1 = 1, m2 = 1, p = 0.5)

bzip.a(xvec = data1[,1], yvec = data1[,2], showFlag = FALSE)
```

---

bzip.b *The bivariate zero-inflated Poisson distribution (B)*

Description

random generation (rbzip.b), maximum likelihood estimation (bzip.b), and log-likelihood. (lik.bzip.b) for the bivariate zero-inflated Poisson (B) distribution with parameters equal to (m0, m1, m2, p1, p2, p3, p4).

Usage

```
lik.bzip.b(xvec, yvec, m0, m1, m2, p1, p2, p3, p4, param = NULL)

rbzip.b(n, m0, m1, m2, p1, p2, p3, p4, param = NULL)

bzip.b(
  xvec,
  yvec,
  tol = 1e-06,
  initial = NULL,
  showFlag = FALSE,
  maxiter = 200
)
```

Arguments

xvec, yvec	a pair of BZIP (B) random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
m0, m1, m2	mean parameters of the Poisson variables. must be positive.
p1, p2, p3, p4	proportions summing up to 1 (p1 + p2 + p3 + p4 = 1). p1 is the probability of both latent Poisson variables being observed. p2 is the probability of only the first Poisson variables being observed. p3 is the probability of only the second Poisson variables being observed, and p4 is the probability of both Poisson variables being dropped out.
param	a vector of parameters ((m0, m1, m2, p1, p2, p3, p4)). Either param or individual parameters (m0, m1, m2, p1, p2, p3, p4) need to be provided.
n	number of observations.
tol	tolerance for judging convergence. tol = 1e-8 by default.
initial	starting value of param for EM algorithm, a vector of nine values.

showFlag	if TRUE, the updated parameter estimates for each iteration are printed out. If a positive integer, the updated parameter estimates for iterations greater than showFlag are printed out.
maxiter	maximum number of iterations allowed. tol = 50000 by default.

Value

- rbzip.b gives a pair of random vectors following BZIP (B) distribution.
- bzip.b gives the maximum likelihood estimates of a BZIP (B) pair.
- lik.bzip.b gives the log-likelihood of a set of parameters for a BZIP (B) pair.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Examples

```
# generating a pair of random vectors
set.seed(1)
data1 <- rbzip.b(n = 20, m0 = 1, m1 = 1, m2 = 1,
                 p1 = 0.5, p2 = 0.2, p3 = 0.2, p4 = 0.1)

lik.bzip.b(xvec = data1[, 1], yvec = data1[ ,2],
           m0 = 1, m1 = 1, m2 = 1,
           p1 = 0.5, p2 = 0.2, p3 = 0.2, p4 = 0.1)

bzip.b(xvec = data1[,1], yvec = data1[,2], showFlag = FALSE)
```

---

idigamma *Inverse digamma function*

Description

inverse of digamma. digamma function is the first derivative of gamma function divided by gamma function.

Usage

```
idigamma(y)
```

Arguments

y	a numeric vector.

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

Examples

```
idigamma(2)
plot(digamma, 0.1, 3)
plot(idigamma, -10.4, 0.9)
```

`pairwise.bzinb` *Pairwise underlying correlation based on bivariate zero-inflated negative binomial (BZINB) model*

Description

For each pair of rows in the data, underlying corelation (ρ) is calculated based on bivariate zero-inflated negative binomial (BZINB) model.

Usage

```
pairwise.bzinb(
  data,
  nonzero.prop = TRUE,
  fullParam = FALSE,
  showFlag = FALSE,
  nsample = NULL,
  ...
)
```

Arguments

`data`	a matrix with nonnegative integers. rows represent the feature (or gene), and columns represent the sample. If not integers, rounded to the nearest integers.
`nonzero.prop`	logical. If `TRUE`, proportion of nonzero for each of the pair is returned.
`fullParam`	logical. If `TRUE`, estimates of all parameters are returned.
`showFlag`	logical. If `TRUE`, for each pair, the estimates are printed out.
`nsample`	positive integer. If provided, `nsample` random pairs will only be considered for correlation. A non-integer will be rounded to the nearest integer.
`...`	Other arguments passed on to `bzinb` function.

Value

a table of pairwise underlying correlation (ρ) and related statistics.

- `1` row number of the first vector of the pair
- `2` row number of the second vector of the pair
- `pair` row numbers of the pair
- `rho` underlying correlation estimate
- `se.rho` standard error of the underlying correlation estimate
- `nonzero.1`, `nonzero.2` non-zero proportion of the first and the second vector. Returned if `nonzero.prop` is `TRUE`.
- `nonzero.min` pairwise minimum of non-zero proportions Returned if `nonzero.prop` is `TRUE`.
- `a0`, `a1`, `...`, `p4` parameter estimates
- `se.a0`, `se.a1`, `...`, `se.p4` standard error of the parameter estimates
- `logLik` log-likelihood of the maximum likelihood estimates

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Liu, C., Preisser, J., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Examples

```
# generating four random vectors
set.seed(7)
data1 <- rbzinb(n = 20, a0 = 0.5, a1 = 1, a2 = 1,
                b1 = 1, b2 = 1, p1 = 0.5, p2 = 0.2,
                p3 = 0.2, p4 = 0.1)
set.seed(14)
data2 <- rbzinb(n = 20, a0 = 0.5, a1 = 1, a2 = 1,
                b1 = 2, b2 = 2, p1 = 0.5, p2 = 0.2,
                p3 = 0.2, p4 = 0.1)
data3 <- t(cbind(data1, data2))

# calculating all pairwise underlying correlations
## Not run: pairwise.bzinb(data3, showFlag = TRUE)
```

`rBzinbData` *rBzinbData BZINB data generation with random normal covariates where the covariates corresponding to eta and gamma coincide up to the smaller dimension.*

Description

install.packages("MASS")

Usage

```
rBzinbData(
  n = 200,
  alpha = c(1, 1, 1),
  eta1 = c(-3, 0.5, 0.5, 0.5, 0.5),
  eta2 = c(-2, 0, 1, 0, 1),
  gamma1 = c(1, 1),
  gamma2 = c(-1, 1),
  gamma3 = c(-1, 1),
  rho = 0.3
)
```

Examples

```
rBzinbData()
```

`weighted.pc` *Weighted Pearson Correlation (WPC) based on bivariate zero-inflated negative binomial (BZINB) model*

Description

weighted.pc calculates Pearson's correlation with less weights for pairs containing zero(s). The weights are determined by BZINB model.

Usage

```
weighted.pc(xvec, yvec, param = NULL, ...)
```

Arguments

`xvec`, `yvec`	a pair of bzinb random vectors. nonnegative integer vectors. If not integers, they will be rounded to the nearest integers.
`param`	a vector of parameters (`(a0, a1, a2, b1, b2, p1, p2, p3, p4)`). See `bzinb` for details. If `param` is `null`, it will be estimated by `bzinb()`.
`...`	optional arguments used passed to `bzinb`, when `param` is `null`.

Value

weighted Pearson correlation (WPC) estimate

Author(s)

Hunyong Cho, Chuwen Liu, Jinyoung Park, and Di Wu

References

Cho, H., Preisser, J., Liu, C., and Wu, D. (In preparation), "A bivariate zero-inflated negative binomial model for identifying underlying dependence"

Examples

```
# generating a pair of random vectors
set.seed(2)
data1 <- rbzinb(n = 20, a0 = 1, a1 = 1, a2 = 1,
                b1 = 1, b2 = 1, p1 = 0.5, p2 = 0.2,
                p3 = 0.2, p4 = 0.1)

weighted.pc(xvec = data1[,1], yvec = data1[,2],
            param = c(0.769, 0.041, 0.075, 3.225, 1.902, 0.5, 0.084, 1e-20, 0.416))
weighted.pc(xvec = data1[,1], yvec = data1[,2])
```

Index