Package ‘poilog’

February 20, 2015

Type Package

Title Poisson lognormal and bivariate Poisson lognormal distribution

Version 0.4

Date 2008-04-29

Author Vidar Grøtå and Steinar Engen

Maintainer Vidar Grøtå <vidar.grotan@bio.ntnu.no>

Description Functions for obtaining the density, random deviates and maximum likelihood estimates of the Poisson lognormal distribution and the bivariate Poisson lognormal distribution.

License GPL-3

Encoding latin1

NeedsCompilation yes

Repository CRAN

Date/Publication 2008-04-29 20:16:14

R topics documented:

poilog-package .................................................. 2
Bivariate Poisson lognormal ................................ 2
MLE for bivariate Poisson lognormal distribution .......... 5
MLE for Poisson lognormal distribution ..................... 7
Poisson lognormal ............................................. 9

Index 11
poilog-package

Poisson lognormal and bivariate Poisson lognormal distribution

Description

Functions for obtaining the density, random deviates and maximum likelihood estimates of the Poisson lognormal distribution and the bivariate Poisson lognormal distribution.

Details

Package: poilog
Type: Package
Version: 0.3
Date: 2007-12-03
License: GPL-3

dpoilog returns the density, rpoilog returns random deviates and poilogMLE performs maximum likelihood estimation of parameters for the Poisson lognormal distribution.
dbipoilog returns the density, rbipoilog returns random deviates and bipoilogMLE performs maximum likelihood estimation of parameters for the bivariate Poisson lognormal distribution.

Author(s)

Vidar Grøtan and Steinar Engen
Maintainer: Vidar Grøtan <vidar.grotan@bio.ntnu.no>

References


Bivariate Poisson lognormal

Bivariate Poisson Lognormal Distribution

Description

Density and random generation for the bivariate Poisson lognormal distribution with parameters mu1, mu2, sig1, sig2 and rho.
Bivariate Poisson lognormal

Usage

\[ \text{dbipoilog}(n1, n2, \mu1, \mu2, \sigma1, \sigma2, \rho) \]
\[ \text{rbipoilog}(S, \mu1, \mu2, \sigma1, \sigma2, \rho, \nu1=1, \nu2=1, \]
\[ \text{condS=FALSE, keep0=FALSE}) \]

Arguments

- \( n1 \) vector of observed individuals for each species in sample 1
- \( n2 \) vector of observed individuals for each species in sample 2
  (in the same order as in sample 1)
- \( \mu1 \) mean of lognormal distribution in sample 1
- \( \mu2 \) mean of lognormal distribution in sample 1
- \( \sigma1 \) standard deviation of lognormal distribution in sample 1
- \( \sigma2 \) standard deviation of lognormal distribution in sample 2
- \( \rho \) correlation among samples
- \( S \) Total number of species in both communities
- \( \nu1 \) sampling intensity sample 1
- \( \nu2 \) sampling intensity sample 2
- \( \text{condS} \) logical; if TRUE random deviates are conditional on \( S \)
- \( \text{keep0} \) logical; if TRUE species with count 0 in both communities are included in the random deviates

Details

The following is written from the perspective of using the Poisson lognormal distribution to describe community structure (the distribution of species when sampling individuals from a community of several species).

The following assumes knowledge of the Details section of \texttt{dpoilog}.

Consider two communities jointly and assume that the log abundances among species have the binormal distribution with parameters \((\mu1, \sigma1, \mu2, \sigma2, \rho)\). If sampling intensities are \(\nu1 = \nu2 = 1\), samples from the communities will have the bivariate Poisson lognormal distribution

\[
P(N_1 = n1, N_2 = n2; \mu1, \sigma1, \mu2, \sigma2, \rho) =
q(n1,n2; \mu1,\sigma1,\mu2,\sigma2,\rho) =
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{n1}(\mu1,\sigma1,u)g_{n2}(\mu2,\sigma2,v) \phi(u,v;\rho) dudv,
\]
where \( \phi(u,v;\rho) \) here denotes the binormal distribution with zero means, unit variances and correlation \( \rho \). In the general case with sampling intensities \(\nu1\) and \(\nu2\), \(\mu1\) and \(\mu2\) should be replaced by \(\mu1 + \ln \nu1\) and \(\mu2 + \ln \nu2\) respectively. In this case, some species will be missing from both samples. The number of individuals for observed species will then have the truncated distribution

\[
q(n1,n2; \mu1,\sigma1,\mu2,\sigma2,\rho)
\]
\[
1 - q(0,0; \mu1,\sigma1,\mu2,\sigma2,\rho)
\]
Bivariate Poisson lognormal

Value

- `dbipoilog` returns the density
- `rbipoilog` returns random deviates

Author(s)

Vidar Grøtan <vidar.grotan@bio.ntnu.no> and Steinar Engen

References


See Also

- `bipoilogMLE` for maximum likelihood estimation

Examples

### change in density of `n2` for two different values of `rho` (given `n1=10`)
```
barplot(rbind(dbipoilog(n1=rep(10,21),n2=0:20,mu1=0,mu2=0,sig1=1,sig2=1,rho=0.0),
            dbipoilog(n1=rep(10,21),n2=0:20,mu1=0,mu2=0,sig1=1,sig2=1,rho=0.8)),
          beside=TRUE,space=c(0,0.2),names.arg=0:20,xlab="n2",col=1:2)
legend(35,0.0012,c("rho=0","rho=0.8"),fill=1:2)
```

### draw random deviates from a community of 50 species
```
rbipoilog(S=50,mu1=0,mu2=0,sig1=1,sig2=2,rho=0.7)
```

### draw random deviates including zeros
```
rbipoilog(S=50,mu1=0,mu2=0,sig1=1,sig2=2,rho=0.7,keep0=TRUE)
```

### draw random deviates with sampling intensities `nu1=0.5` and `nu2=0.7`
```
rbipoilog(S=50,mu1=0,mu2=0,sig1=1,sig2=2,rho=0.7,nu1=0.5,nu2=0.7)
```

### draw random deviates conditioned on a certain number of species
```
rbipoilog(S=50,mu1=0,mu2=0,sig1=1,sig2=2,rho=0.7,nu1=0.5,nu2=0.7,conds=TRUE)
```

### how many species are likely to be observed in at least one of the samples

#### (given `S,mu1,mu2,sig1,sig2,rho`)
```
hist(replicate(1000,nrow(rbipoilog(S=50,mu1=0,mu2=0,sig1=1,sig2=2,rho=0.7)))),
     main="", xlab = "Number of species observed in at least one of the samples")
```

### how many individuals are likely to be observed

#### (given `S,mu1,mu2,sig1,sig2,rho`)
```
hist(replicate(1000,sum(rbipoilog(S=50,mu1=0,mu2=0,sig1=1,sig2=2,rho=0.7)))),
     main="", xlab="sum nr of individuals in both samples")
```
bipoilogMLE fits the bivariate Poisson lognormal distribution to data

Usage

bipoilogMLE(n1, n2 = NULL,
    startVals = c(mu1=1, mu2=1, sig1=2, sig2=2, rho=0.5),
    nboot = 0, zTrunc = TRUE, file = NULL,
    method = "BFGS", control = list(maxit=1000))

Arguments

n1 a vector or a matrix with two columns of pairwise counts of observed individuals for each species
n2 if n1 is not given as a matrix, a vector of counts with same ordering of species as in argument n1
startVals starting values of parameters
nboot number of parametric bootstraps, defaults to zero
zTrunc logical; if TRUE (default) the zero-truncated distribution is fitted
file text file to hold copies of bootstrap estimates
method method to use during optimization, see details
control a list of control parameters for the optimization routine, see details

Details

The function estimates the parameters mu1, sig1, mu2, sig2 and rho. In cases of incomplete sampling the estimates of mu1 and mu2 will be confounded with the sampling intensities (see rbipoilog). Assuming sampling intensities ν₁ and ν₂, the estimates of the means are mu1 + ln ν₁ and mu2 + ln ν₂. Parameters sig1, sig2 and rho can be estimated without any knowledge of sampling intensities. The parameters must be given starting values for the optimization procedure (default starting values are used if starting values are not specified in the function call).

A zero-truncated distribution (see dbipoilog) is assumed by default (zTrunc = TRUE). In cases where the number of zeros is known the zTrunc argument should be set to FALSE.

The function uses the optimization procedures in optim to obtain the maximum likelihood estimate. The method and control arguments are passed to optim, see the help page for this function for additional methods and control parameters.

The approximate fraction of species revealed by each sample is estimated as 1 minus the zero term of the univariate Poisson lognormal distribution: 1 − q(0; mu1, sig1) and 1 − q(0; mu2, sig2).
Parametric bootstrapping could be time consuming for large data sets. If argument file is specified, e.g. file = 'C:\myboots.txt', the matrix with bootstrap estimates are copied into a tab-separated text-file providing extra backup. Bootstrapping is done by simulating new sets of observations conditioned on the observed number of species (see rbipoilog).

Value

- **par**: Maximum likelihood estimates of the parameters
- **p**: Approximate fraction of species revealed by the samples for sample 1 and 2 respectively
- **logLval**: Log likelihood of the data given the estimated parameters
- **gof**: Godness of fit measure obtained by checking the rank of logLval against logLval’s obtained during the bootstrap procedure, (gof < 0.05) or (gof > 0.95) indicates lack of fit
- **boot**: A data frame containing the bootstrap replicates of parameters and logLval

Author(s)

Vidar Grøtan <vidar.grotan@bio.ntnu.no>, Steinar Engen

References


See Also

- optim, dbipoilog, rbipoilog

Examples

```r
## simulate observations
xy <- rbipoilog(S=30, mu1=1, mu2=1, sig1=2, sig2=2, rho=0.5)

## obtain estimates of parameters
est <- bipoilogMLE(xy)

## similar, but now with bootstrapping
## Not run: est <- bipoilogMLE(xy, nboot=10)

## change start values and request tracing information
## from optimization procedure
est <- bipoilogMLE(xy, startVals=c(2, 2, 4, 4, 0.3),
               control=list(maxit=10000, trace=1, REPORT=1))

## effect of sampling intensity
xy <- rbipoilog(S=100, mu1=1, mu2=1, sig1=2, sig2=2, rho=0.5, nu1=0.5, nu2=0.5)
est <- bipoilogMLE(xy)
```
MLE for Poisson lognormal distribution

## Maximum Likelihood Estimation for Poisson Lognormal Distribution

### Description

`poilogMLE` fits the Poisson lognormal distribution to data and estimates parameters mean $\mu$ and standard deviation $\sigma$ in the lognormal distribution.

### Usage

```r
poilogMLE(n, startVals = c(mu=1L, sig=2L), nboot = 0, zTrunc = TRUE, method = "BFGS", control = list(maxit=1000))
```

### Arguments

- `n`: A vector of counts
- `startVals`: Starting values of parameters, see details
- `nboot`: Number of parametric bootstraps, defaults to zero
- `zTrunc`: Zero-truncated distribution, defaults to TRUE
- `method`: Method to use during optimization, see details
- `control`: A list of control parameters for the optimization routine, see details

### Details

The function estimates parameters mean $\mu$ and standard deviation $\sigma$. In cases of incomplete sampling the estimate of $\mu$ will be confounded with the sampling intensity (see `rpoilog`). Assuming sampling intensity $\nu$, the estimates of the mean is $\mu + \ln(\nu)$. Parameter $\sigma$ can be estimated without any knowledge of sampling intensity.

The parameters must be given starting values for the optimization procedure (default starting values are used if starting values are not specified in the function call).

The function uses the optimization procedures in `optim` to obtain the maximum likelihood estimate. The method and control arguments are passed to `optim`, see the help page for this function for additional methods and control parameters.

A zero-truncated distribution (see `dpoilog`) is assumed by default ($zTrunc = TRUE$). In cases where the number of zeros is known the `zTrunc` argument should be set to FALSE.

The approximate fraction of species revealed by the sample is $1 - q(0; \mu, \sigma)$.

Parametric bootstrapping is done by simulating new sets of observations using the estimated parameters (see `rbipoilog`).

---

### Example

```r
## the expected estimates of mu1 and mu2 are now 1-log(0.5) = 0.3 (approximately)
```

---

# the expected estimates of mu1 and mu2 are now 1-log(0.5) = 0.3 (approximately)
Value
par     Maximum likelihood estimates of the parameters
p       Approximate fraction of species revealed by the sample
logLval Log likelihood of the data given the estimated parameters
gof     Godness of fit measure obtained by checking the rank of logLval against logLval’s obtained during the bootstrap procedure, (gof<0.05) or (gof>0.95) indicates lack of fit
boot    A data frame containing the bootstrap replicates of parameters and logLval

Author(s)
Vidar Grøtøn <vidar.grotan@bio.ntnu.no>, Steinar Engen

References

See Also
optim, dpoilog, rpoilog

Examples

### simulate observations
n <- rpoilog(S=80,mu=1,sig=2)

### obtain estimates of parameters
est <- poilogMLE(n)

### similar, but now with bootstrapping ###
## Not run: est <- poilogMLE(n,nboot=10)

### change start values and request tracing information
### from optimization procedure
est <- poilogMLE(n,startVals=c(2,3),
                  control=list(maxit=1000,trace=1, REPORT=1))
Poisson lognormal

Description

Density and random generation for the Poisson lognormal distribution with parameters \( \mu \) and \( \sigma \).

Usage

\[
dpoilog(n, \mu, \sigma) \\
rpoilog(S, \mu, \sigma, nu=1, \text{condS}=\text{FALSE}, \text{keep0}=\text{FALSE})
\]

Arguments

- \( n \): vector of observed individuals for each species
- \( S \): number of species in the community
- \( \mu \): mean of lognormal distribution
- \( \sigma \): standard deviation of lognormal distribution
- \( nu \): sampling intensity, defaults to 1
- \( \text{condS} \): logical; if TRUE random deviates are conditional on \( S \)
- \( \text{keep0} \): logical; if TRUE species with count 0 are included in the random deviates

Details

The following is written from the perspective of using the Poisson lognormal distribution to describe community structure (the distribution of species when sampling individuals from a community of several species).

Under the assumption of random sampling, the number of individuals sampled from a given species with abundance \( y \), say \( N \), is Poisson distributed with mean \( \nu y \) where the parameter \( \nu \) expresses the sampling intensity. If \( \ln y \) is normally distributed with mean \( \mu \) and standard deviation \( \sigma \) among species, then the vector of individuals sampled from all \( S \) species then constitutes a sample from the Poisson lognormal distribution with parameters \( (\mu + \ln \nu, \sigma) \), where \( \mu \) and \( \sigma \) are the mean and standard deviation of the log abundances. For \( \nu = 1 \), this is the Poisson lognormal distribution with parameters \( (\mu, \sigma) \) which may be written in the form

\[
P(N = n; \mu, \sigma) = q(n; \mu, \sigma) = \int_{-\infty}^{\infty} g_n(\mu, \sigma, u) \phi(u) \, du,
\]

where \( \phi(u) \) is the standard normal distribution and

\[
g_n(\mu, \sigma, u) = \frac{\exp(u \sigma n + \mu n - \exp(u \sigma + \mu))}{n!}
\]
Since $S$ is usually unknown, we only consider the observed number of individuals for the observed species. With a general sampling intensity $n_u$, the distribution of the number of individuals then follows the zero-truncated Poisson lognormal distribution

$$q(n; \mu, \sigma) \over 1 - q(0; \mu, \sigma)$$

Value

dpoilog returns the density
rpoilog returns random deviates

Author(s)

Vidar Grøtan <vidar.grotan@bio.ntnu.no> and Steinar Engen

References


See Also

poilogMLE for ML estimation

Examples

### plot density for given parameters
barplot(dpoilog(n=0:20, mu=2, sig=1), names.arg=0:20)

### draw random deviates from a community of 50 species
rpoilog(S=50, mu=2, sig=1)

### draw random deviates including zeros
rpoilog(S=50, mu=2, sig=1, keep0=TRUE)

### draw random deviates with sampling intensity = 0.5
rpoilog(S=50, mu=2, sig=1, nu=0.5)

### how many species are likely to be observed
### (given S, mu, sig2 and nu)?
hist(replicate(1000, length(rpoilog(S=30, mu=0, sig=3, nu=0.7))))

### how many individuals are likely to be observed
### (given S, mu, sig2 and nu)?
hist(replicate(1000, sum(rpoilog(S=30, mu=0, sig=3, nu=0.7))))
Index

*Topic distribution
  Bivariate Poisson lognormal, 2
  MLE for bivariate Poisson lognormal distribution, 5
  MLE for Poisson lognormal distribution, 7
  Poisson lognormal, 9

*Topic package
  poilog-package, 2

bipoilogMLE, 2, 4
bipoilogMLE (MLE for bivariate Poisson lognormal distribution), 5
Bivariate Poisson lognormal, 2

dbipoilog, 2, 5, 6
dbipoilog (Bivariate Poisson lognormal), 2
dpoilog, 2, 3, 7, 8
dpoilog (Poisson lognormal), 9

MLE for bivariate Poisson lognormal distribution, 5
MLE for Poisson lognormal distribution, 7

optim, 5–8

poilog (poilog-package), 2
poilog-package, 2
poilogMLE, 2, 10
poilogMLE (MLE for Poisson lognormal distribution), 7
Poisson lognormal, 9

rbipoilog, 2, 5–7
rbipoilog (Bivariate Poisson lognormal), 2
rpoilog, 2, 7, 8
rpoilog (Poisson lognormal), 9