Package: poilog (via r-universe)

August 30, 2024

Type Package

Title Poisson Lognormal and Bivariate Poisson Lognormal Distribution

Version 0.4.2

Date 2022-10-07

Author Vidar Grotan and Steinar Engen

Maintainer Vidar Grotan <`vidar.grotan@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

NeedsCompilation yes

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

RemoteUrl https://github.com/vidargrotan/poilog

RemoteRef HEAD

RemoteSha d5e9eb1bbe02c975705e1ee6b3efd445346833c7

Contents

`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:	0.4.2
Date:	2022-10-07
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 Grotan and Steinar Engen

Maintainer: Vidar Grotan <`vidar.grotan@ntnu.no`>

References

Bulmer, M. G. 1974. On fitting the Poisson lognormal distribution to species abundance data. Biometrics **30**, 651-660.
Engen, S., R. Lande, T. Walla and P. J. DeVries. 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. American Naturalist **160**, 60-73.

`Bivariate Poisson lognormal` *Bivariate Poisson Lognormal Distribution*

Description

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

Usage

```
dbipoilog(n1, n2, mu1, mu2, sig1, sig2, rho)
rbipoilog(S, mu1, mu2, sig1, sig2, rho, nu1=1, nu2=1,
          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
`sig1`	standard deviation of lognormal distribution in sample 1
`sig2`	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
`condS`	logical; if TRUE random deviates are conditional on S
`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 `dpoilog`.

Consider two communities jointly and assume that the log abundances among species have the binormal distribution with parameters (`mu1`,`sig1`,`mu2`,`sig2`,`rho`). If sampling intensities are `nu1` = `nu2` = 1, samples from the communities will have the bivariate Poisson lognormal distribution

$$P(N_1 = \mathtt{n1}, N_2 = \mathtt{n2};\ \mathtt{mu1}, \mathtt{sig1}, \mathtt{mu2}, \mathtt{sig2}, \mathtt{rho}) =$$
$$q(\mathtt{n1}, \mathtt{n2};\ \mathtt{mu1}, \mathtt{sig1}, \mathtt{mu2}, \mathtt{sig2}, \mathtt{rho}) =$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{\mathtt{n1}}(\mathtt{mu1}, \mathtt{sig1}, u) g_{\mathtt{n2}}(\mathtt{mu2}, \mathtt{sig2}, v)\, \phi(u, v; \mathtt{rho})\, du dv,$$

where $\phi(u, v; \mathtt{rho})$ 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

$$\frac{q(\mathtt{n1}, \mathtt{n2};\ \mathtt{mu1}, \mathtt{sig1}, \mathtt{mu2}, \mathtt{sig2}, \mathtt{rho})}{1 - q(0, 0;\ \mathtt{mu1}, \mathtt{sig1}, \mathtt{mu2}, \mathtt{sig2}, \mathtt{rho})}$$

Value

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

Author(s)

Vidar Grotan <`vidar.grotan@ntnu.no`> and Steinar Engen

References

Engen, S., R. Lande, T. Walla and P. J. DeVries. 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. American Naturalist **160**, 60-73.

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,sig=1,sig2=1,rho=0.0),
              dbipoilog(n1=rep(10,21),n2=0:20,mu1=0,mu2=0,sig=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")
```

`MLE for bivariate Poisson lognormal distribution` *Maximum Likelihood Estimation for Bivariate Poisson Lognormal Distribution*

Description

`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 ν_1 and ν_2, the estimates of the means are $\texttt{mu1} + \ln \nu_1$ and $\texttt{mu2} + \ln \nu_2$. 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: $1 - q(0; \texttt{mu1}, \texttt{sig1})$ and $1 - q(0; \texttt{mu2}, \texttt{sig2})$.

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`	Goodness 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 Grotan <`vidar.grotan@ntnu.no`>, Steinar Engen

References

Engen, S., R. Lande, T. Walla and P. J. DeVries. 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. American Naturalist **160**, 60-73.

See Also

`optim`, `dbipoilog`, `rbipoilog`

Examples

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

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

## change start values and request tracing information
## from optimization procedure
## Not run: est <- bipoilogMLE(xy,startVals=c(2,2,4,4,0.3),
       control=list(maxit=1000,trace=1, REPORT=1))
## End(Not run)

## 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)
## Not run: est <- bipoilogMLE(xy)
## the expected estimates of mu1 and mu2 are now 1-log(0.5) = 0.3 (approximately)
```

`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 `sig` in the lognormal distribution

Usage

```
poilogMLE(n, startVals = c(mu=1, sig=2),
          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 `sig`. In cases of incomplete sampling the estimate of `mu` will be confounded with the sampling intensity (see `rpoilog`). Assuming sampling intensity ν, the estimates of the mean is $\texttt{mu} + \ln(\nu)$. Parameter `sig` 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 $1 - q(0; \texttt{mu}, \texttt{sig})$.

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

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`	Goodness 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 Grotan <`vidar.grotan@ntnu.no`>, Steinar Engen

References

Bulmer, M. G. 1974. On fitting the Poisson lognormal distribution to species abundance data. Biometrics **30**, 651-660.
Engen, S., R. Lande, T. Walla and P. J. DeVries. 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. American Naturalist **160**, 60-73.

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` *Poisson lognormal distribution*

Description

Density and random generation for the Poisson lognormal distribution with parameters `mu` and `sig`.

Usage

```
dpoilog(n, mu, sig)
rpoilog(S, mu, sig, nu=1, condS=FALSE, keep0=FALSE)
```

Arguments

`n`	vector of observed individuals for each species
`S`	number of species in the community
`mu`	mean of lognormal distribution
`sig`	standard deviation of lognormal distribution
`nu`	sampling intensity, defaults to 1
`condS`	logical; if TRUE random deviates are conditional on S
`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 $\mathsf{nu}\, y$ where the parameter `nu` expresses the sampling intensity. If ln y is normally distributed with mean `mu` and standard deviation `sig` 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`, `sig`), where `mu` and `sig` are the mean and standard deviation of the log abundances. For `nu` = 1, this is the Poisson lognormal distribution with parameters (`mu`,`sig`) which may be written in the form

$$P(N = \mathsf{n}; \mathsf{mu}, \mathsf{sig}) = q(\mathsf{n}; \mathsf{mu}, \mathsf{sig}) = \int_{-\infty}^{\infty} g_{\mathsf{n}}(\mathsf{mu}, \mathsf{sig}, u)\phi(u)\ du,$$

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

$$g_{\mathsf{n}}(\mathsf{mu}, \mathsf{sig}, u) = \frac{\exp(u\, \mathsf{sig}\, \mathsf{n} + \mathsf{mu}\, \mathsf{n} - \exp(u\, \mathsf{sig} + \mathsf{mu}))}{\mathsf{n}!}$$

Since `S` is usually unknown, we only consider the observed number of individuals for the observed species. With a general sampling intensity `nu`, the distribution of the number of individuals then follows the zero-truncated Poisson lognormal distribution

$$\frac{q(\texttt{n};\texttt{mu},\texttt{sig})}{1-q(0;\texttt{mu},\texttt{sig})}$$

Value

`dpoilog` returns the density
`rpoilog` returns random deviates

Author(s)

Vidar Grotan <`vidar.grotan@ntnu.no`> and Steinar Engen

References

Engen, S., R. Lande, T. Walla & P. J. DeVries. 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. American Naturalist 160: 60-73.

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