Package 'lira'

October 13, 2022

Type Package

Title LInear Regression in Astronomy

Version 2.0.1

Date 2018-02-05

Author Mauro Sereno

Maintainer Mauro Sereno <mauro.sereno@unibo.it>

Description
Performs Bayesian linear regression and forecasting in astronomy. The method accounts for heteroscedastic errors in both the independent and the dependent variables, intrinsic scatters (in both variables) and scatter correlation, time evolution of slopes, normalization, scatters, Malmquist and Eddington bias, upper limits and break of linearity. The posterior distribution of the regression parameters is sampled with a Gibbs method exploiting the JAGS library.

License GPL-2

Depends R (>= 2.14.0), coda, rjags

SystemRequirements JAGS (>= 3.0.0) (see
http://mcmc-jags.sourceforge.net)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-02-06 14:19:28 UTC

R topics documented:

1

---

distance.LCDM *Angular diameter distance in a LCDM universe*

Description

Angular diameter distance in a LCDM (Cold Dark Matter universe with a cosmological constant) universe between an observer at 'zd' and a source at 'zs'. The distance is in units of the Hubble distance c/H0, where 'c' is the speed of light and 'H0' is the Hubble constant.

Usage

```
distance.LCDM(zo=0.0,z,Omega.M0=0.3,Omega.L0=0.7)
```

Arguments

zo	redshift of the observer.
z	redshift of the source.
Omega.M0	present day matter density in units of the critical density.
Omega.L0	renormalized cosmological constant in units of the critical density.

Value

The estimate of the distance.

Author(s)

Mauro Sereno <mauro.sereno@unibo.it>

References

Sereno M., Covone G., Piedipalumbo E., de Ritis R., (2001) Monthly Notices of the Royal Astronomical Society 327, 517-530.

See Also

See also Hubble.LCDM.

Examples

```
# Distance to a source at z=1 in a standard cosmology,
distance.LCDM(z=1.0)
# Distance between a lens at zl=0.3 and a source at z=1 in an open cosmology,
distance.LCDM(0.3,1,Omega.M0=0.2,Omega.L0=0.7)
```

Hubble.LCDM *Hubble parameter*

Description

Hubble parameter in a LCDM (Cold Dark Matter universe with a cosmological constant) universe in units of the present day Hubble constant.

Usage

```
Hubble.LCDM(z,Omega.M0=0.3,Omega.L0=0.7)
```

Arguments

z	the redshift
Omega.M0	present day matter density in units of the critical density in a LCDM universe.
Omega.L0	renormalized cosmological constant in units of the critical density. See 'Omega.M0'.

Value

The Hubble parameter in units of H0.

Author(s)

Mauro Sereno <mauro.sereno@unibo.it>

References

Sereno M., Covone G., Piedipalumbo E., de Ritis R., (2001) Monthly Notices of the Royal Astronomical Society 327, 517-530.

See Also

See also Hubble.LCDM.

Examples

```
# Hubbel parameter at redshit z=1 in a standard cosmology,
Hubble.LCDM(1)
# Hubbel parameter at redshit z=1 in an open cosmology,
Hubble.LCDM(1,Omega.M0=0.2,Omega.L0=0.7)
```

---

lira   *LInear Regression in Astronomy*

Description

Performs Bayesian linear regression analysis and forecasting optimized for astronomy. The method accounts for heteroscedastic and correlated errors in both the independent and the dependent variables, intrinsic scatters (in both variables) and scatter correlation, time evolution of normalization, slope, scatter, Malmquist and Eddington bias, upper limits and break of linearity ('knee'). The posterior distribution of the parameters is sampled with a JAGS (Just Another Gibbs Sampler) powered Gibbs sampler.

Usage

```
lira(x, y,
delta.x, delta.y, covariance.xy,
y.threshold, delta.y.threshold, x.threshold, delta.x.threshold,
y.upperlimit, delta.y.upperlimit, x.upperlimit, delta.x.upperlimit,
z, z.ref=0.01, time.factor="Ez", distance=c("luminosity","angular"),
Omega.M0=0.3, Omega.L0=0.7,
alpha.YIZ="dunif", beta.YIZ="dt", gamma.YIZ="dt", delta.YIZ=0.0,
sigma.YIZ.0="prec.dgamma", gamma.sigma.YIZ.Fz=0.0, gamma.sigma.YIZ.D=0.0,
alpha.XIZ=0.0, beta.XIZ=1.0, gamma.XIZ=0.0, delta.XIZ=0.0,
sigma.XIZ.0=0.0, gamma.sigma.XIZ.Fz=0.0, gamma.sigma.XIZ.D=0.0,
rho.XYIZ.0=0.0,  gamma.rho.XYIZ.Fz=0.0, gamma.rho.XYIZ.D=0.0,
n.mixture=1, pi="ddirch",
mu.Z.0="dunif", gamma.mu.Z.Fz="dt", gamma.mu.Z.D="dt",
sigma.Z.0="prec.dgamma", gamma.sigma.Z.Fz=0.0, gamma.sigma.Z.D=0.0,
mu.Z.0.mixture="dunif", sigma.Z.0.mixture="prec.dgamma",
Z.knee="dunif", l.knee=1e-04,
beta.YIZ.knee="beta.YIZ", delta.YIZ.knee="delta.YIZ", sigma.YIZ.0.knee = "sigma.YIZ.0",
mu.Z.min.0="-n.large", gamma.mu.Z.min.Fz="dt", gamma.mu.Z.min.D="dt",
sigma.Z.min.0=0.0, gamma.sigma.Z.min.Fz=0.0, gamma.sigma.Z.min.D=0.0,
Z.max= "n.large",
n.large=1e04,
inits, n.chains=1, n.adapt=1e03, quiet=TRUE,
n.iter=1e04, thin=1,
print.summary=TRUE, print.diagnostic=TRUE,
export=FALSE, export.mcmc="", export.jags="",
X.monitored=FALSE, export.X="",
XZ.monitored=FALSE, export.XZ="",
Z.monitored=FALSE, export.Z="",
Y.monitored=FALSE, export.Y="",
YZ.monitored=FALSE, export.YZ="")
```

Arguments

x   the vector containing the observed values "x" of the independent variable.

y	the vector containing the observed values "y" of the dependent variable. For forecasting, set the missing data values to NA. The posterior distribution of the node will be sampled.
delta.x	optional vector containing the measurement 1-sigma uncertainties in "x".
delta.y	optional vector containing the measurement 1-sigma uncertainties in "y".
covariance.xy	optional vector containing the measurement uncertainties covariances.
y.threshold	optional vector containing the minimum observable values of "y". These threshold values are used to correct for the Malmquist bias in truncated samples. To be used if undetected objects (y<y.threshold) are entirely missing from the dataset.
delta.y.threshold	optional vector containing the 1-sigma uncertainties in "y.threshold".
x.threshold	optional vector containing the minimum observable values of "x". See "y.threshold".
delta.x.threshold	optional vector containing the 1-sigma uncertainties in "x.threshold".
y.upperlimit	optional vector containing the upper limit of "y" for left-censored data points. To be used if the dataset contains the full sample of interest, but some objects have upper limits and others have detections. For detected objects, set the upper limit to NA or to very large values.
delta.y.upperlimit	optional vector containing the 1-sigma uncertainties in "y.upperlimit".
x.upperlimit	optional vector containing the upper limit of "x". See "y.upperlimit".
delta.x.upperlimit	optional vector containing the 1-sigma uncertainties in "x.upperlimit".
z	optional vector containing the measured redshifts.
z.ref	reference redshift.
time.factor	character string indicating the form of the "Fz" time factor. "Fz" is renormalized such that Fz=1 at the reference redshift z.ref. If time.factor="Ez", the evolution is proportional to the Hubble factor; if "1+z", the evolution is proportional to (1+z).
distance	character string denoting the cosmological distances to use. If distance="luminosity" or "angular", luminosity or angular diameter distances are considered, respectively.
Omega.M0	present day matter density in units of the critical density in a LCDM universe. This is used to compute the distances and the time factor "Ez", if applicable.
Omega.L0	renormalized cosmological constant in units of the critical density. See "Omega.M0".
alpha.YIZ	a character string or a number giving the prior on the intercept of the Y-Z regression. By default, a uniform distribution is assumed. See Details.
beta.YIZ	a character string or a number giving the prior on the slope of the Y-Z relation. By default, a Student's t-distribution is assumed. See Details.
gamma.YIZ	a character string or a number giving the prior on slope of the time evolution of the Y-Z relation. If redshifts are provided, a default Student's t-distribution is assumed. Otherwise, the parameter is set to 0. See Details.

delta.YIZ	a character string or a number giving the prior on the slope tilt with time of the Y-Z relation. By default, this parameter is set to 0. See Details.
sigma.YIZ.0	a character string or a number giving the prior on the standard deviation of the intrinsic scatter of the Y-Z relation at z.ref. By default, the precision, i.e., sigma.YIZ.0^-2, is assumed to follow a Gamma distribution. See Details.
gamma.sigma.YIZ.Fz	
	a character string or a number giving the prior on the evolution with "Fz" of the scatter sigma.YIZ. See Details. By default, this parameter is set to 0. See Details.
gamma.sigma.YIZ.D	
	a character string or a number giving the prior on the evolution with the distance "D" of the scatter sigma.YIZ. By default, a delta prior is assumed and the parameter is set to 0. See Details.
alpha.XIZ	a character string or a number giving the prior on the intercept of the X-Z relation. By default, this parameter is set to 0. See Details.
beta.XIZ	a character string or a number giving the prior on the slope of the X-Z relation. By default, this parameter is fixed to 1. See Details.
gamma.XIZ	a character string or a number giving the prior on the slope of the time evolution of the X-Z relation. By default, this parameter is set to 0. See Details.
delta.XIZ	a character string or a number giving the prior on the slope tilt with time of the X-Z relation. By default, this parameter is set to 0. See Details.
sigma.XIZ.0	a character string or a number giving the prior on the standard deviation of the intrinsic scatter of the X-Z relation at z.ref. By default, this parameter is set to 0. See Details.
gamma.sigma.XIZ.Fz	
	a character string or a number giving the prior on the evolution with "Fz" of the scatter sigma.XIZ. By default, this parameter is set to 0. See Details.
gamma.sigma.XIZ.D	
	a character string or a number giving the prior on the evolution with the distance "D" of the scatter sigma.XIZ. By default, a delta prior is assumed and the parameter is set to 0. See Details.
rho.XYIZ.0	a character string or a number giving the prior on the correlation between the intrinsic scatters at z.ref. By default, this parameter is set to 0. See Details.
gamma.rho.XYIZ.Fz	
	a character string or a number giving the prior on the evolution with "Fz" of the scatter-correlation rho.XYIZ. By default, this parameter is set to 0. See Details.
gamma.rho.XYIZ.D	
	a character string or a number giving the prior on the evolution with the distance "D" of the scatter-correlation rho.XYIZ. By default, a delta prior is assumed and the parameter is set to 0. See Details.
n.mixture	the number of Gaussian distributions in the mixture modelling the Z-distribution p(Z). See Details.
pi	the prior on the probability coefficients of the mixture components. By default, a Dirichelet distribution is assumed. See Details.

mu.Z.0	a character string or a number giving the prior on the mean of the first mixture component at z.ref. By default, a uniform distribution is assumed. See Details.
gamma.mu.Z.Fz	a character string or a number giving the prior on the evolution with "Fz" of the mean mu.Z. By default, a Student's t-distribution is assumed. See Details.
gamma.mu.Z.D	a character string or a number giving the prior on the evolution with the distance "D" of the mean mu.Z. By default, a Student's t-distribution is assumed. See Details.
sigma.Z.0	a character string or a number giving the prior on the dispersion of the first mixture component at z.ref. By default, the precision follows a priori a Gamma distribution. See Details.
gamma.sigma.Z.Fz	
	a character string or a number giving the prior on the evolution with "Fz" of the dispersion sigma.Z. By default, this parameter is set to 0. See Details.
gamma.sigma.Z.D	
	a character string or a number giving the prior on the evolution with the distance "D" of the dispersion sigma.Z. By default, a delta prior is assumed and the parameter is set to 0. See Details.
mu.Z.0.mixture	a character string or a number giving the prior on the mean of the additional mixture components at z.ref. By default, a uniform distribution is assumed. See Details.
sigma.Z.0.mixture	
	a character string or a number giving the prior on the dispersion of the additional mixture components at z.ref. By default, the precisions follow a priori a Gamma distribution. See Details.
Z.knee	a character string or a number giving the prior on the Z value where linearity breaks down. If applicable, the default prior is the uniform distribution. See Details.
l.knee	a character string or a number giving the prior on the width of the transition region where linearity breaks down. By default, a delta prior is assumed. See Details.
beta.YIZ.knee	a character string or a number giving the prior on the slope of the Y-Z relation before the break. If beta.YIZ.knee="beta.YIZ", no breaking occurs. See Details.
delta.YIZ.knee	a character string or a number giving the prior on the time evolution of the slope of the Y-Z relation before the break. If delta.YIZ.knee="delta.YIZ", the tilt does not change before the break. See Details.
sigma.YIZ.0.knee	
	a character string or a number giving the prior on the intercept of the Y-Z relation before the break. If sigma.YIZ.0.knee="sigma.0.YIZ", the intrinsic scatter does not change before the break. See Details.
mu.Z.min.0	minimum value of the Z distribution at z.ref. The Gaussian distribution of the covariate is truncated below "mu.Z.Min.0". If "n.mixture">1, "mu.Z.Min.0" is automatically set to -n.large. If "mu.Z.Min.0" is set to -n.large, the Gaussian is not truncated.
gamma.mu.Z.min.Fz	
	a character string or a number giving the prior on the evolution with "Fz" of mu.Z.min. By default, a Student's t-distribution is assumed. See Details.

gamma.mu.Z.min.D

> a character string or a number giving the prior on the evolution with the distance "D" of the mean mu.Z.min. By default, a Student's t-distribution is assumed. See Details.

sigma.Z.min.0   a character string or a number giving the prior on the dispersion of the truncation "mu.Z.min.0" at z.ref. By default, "sigma.Z.min.0" is set to 0.0 and the covariate distribution is sharply truncated. If "sigma.Z.min.0" is left free, the truncation is modelled with a complementary error function with dispersion "sigma.Z.min.0". See Details.

gamma.sigma.Z.min.Fz

> a character string or a number giving the prior on the evolution with "Fz" of the dispersion sigma.Z.min. By default, this parameter is set to 0. See Details.

gamma.sigma.Z.min.D

> a character string or a number giving the prior on the evolution with the distance "D" of the dispersion sigma.Z.min. By default, a delta prior is assumed and the parameter is set to 0. See Details.

Z.max   maximum value of the Z distribution. The Gaussian distribution of Z values is truncated for Z > Z.max. If Z.max=n.large, the distribution is not truncated. If "n.mixture">1, Z.max is automatically set to n.large.

n.large   large number used to define standard prior distributions. n.large has to be much larger than max(abs(x)).

inits   optional specification of initial values in the form of a list or a function. See jags.model for details.

n.chains   the number of parallel chains for the model. See jags.model for details.

n.adapt   the number of iterations for adaptation. See jags.model for details.

quiet   logical value. If TRUE then messages generated during compilation will be suppressed. See jags.model for details.

n.iter   number of iterations to monitor. See coda.samples for details.

thin   thinning interval for monitors. See coda.samples for details.

print.summary   logical value. If TRUE a summary of the chains properties is printed on the standard output.

print.diagnostic

> logical value. If TRUE a diagnostic of the chains properties is printed on the standard output.

export   logical value. If TRUE two output files are written, one text file with the merged thinned chains and one text file with the JAGS script used to perform the regression.

export.mcmc   character string indicating the file name where to write the chains. If "" and export=TRUE, a file named "mcmc_all.dat" is created in the working directory.

export.jags   character string indicating the file name where to write the JAGS model. If "" and export=TRUE, a file named "lira.jags" is created in the working directory.

X.monitored   logical value. If TRUE the X variables are monitored.

export.X   character string indicating the file name where to write the chains for the X variables. If "", there is no export.

`XZ.monitored`	logical value. If TRUE the XZ variables are monitored.
`export.XZ`	character string indicating the file name where to write the chains for the XZ variables. If "", there is no export.
`Z.monitored`	logical value. If TRUE the Z variables are monitored.
`export.Z`	character string indicating the file name where to write the chains for the Z variables. If "", there is no export.
`Y.monitored`	logical value. If TRUE the Y variables are monitored.
`export.Y`	character string indicating the file name where to write the chains for the Y variables. If "", there is no export.
`YZ.monitored`	logical value. If TRUE the YZ variables are monitored.
`export.YZ`	character string indicating the file name where to write the chains for the YZ variables. If "", there is no export.

Details

The linear regression assumes:

Y = alpha.YIZ + beta.YIZ*Z + gamma.YIZ*T + delta.YIZ*T*Z +-sigma.YIZ,

where T=log10(Fz), alpha.YIZ is the intercept, beta.YIZ is the slope, gamma.YIZ is the time-evolution of the normalization, delta.YIZ is the time evolution of the slope, and sigma.YIZ is the time dependent dispersion of the normally-distributed intrinsic scatter. 'Fz' is the time factor. If the covariate variable "Z" cannot be mesured, we may consider X as a scattered proxy of Z,

X = alpha.XIZ + beta.XIZ*Z + gamma.XIZ*T + delta.XIZ*T*Z +-sigma.XIZ.

The input parameters x and y are the measured values of X and Y, respectively. The values delta.x and delta.x are the 1-sigma uncertainties,

x = X +- delta.x,

y = Y +- delta.y.

The measurement error covariances is covariance.xy.

The scatters sigma.YIZ and sigma.XIZ evolve with time as

sigma.YIZ = sigma.YIZ.0*(Fz^gamma.sigma.YIZ.Fz)*(D^gamma.sigma.YIZ.D),

sigma.XIZ = sigma.XIZ.0*(Fz^gamma.sigma.XIZ.Fz)*(D^gamma.sigma.XIZ.D),

where D is the normalized cosmological distance. The scatters sigma.YIZ and sigma.XIZ may be correlated, with correlation

rho.XYIZ = rho.XYIZ.0*(Fz^gamma.rho.XYIZ.Fz)*(D^gamma.rho.XYIZ.D).

The intrinsic distribution of Z is modelled as a mixture of normal distributions. The mean and the dispersion of the first component evolve as

mu.Z = mu.Z.0 + gamma.mu.Z.D*log10(D) + gamma.mu.Z.Fz*T,

sigma.Z = sigma.Z.0*(Fz^gamma.sigma.Z.Fz)*(D^gamma.sigma.Z.D).

When modeled as a single Gaussian, the distribution of the covariate can be truncated at low values with a complementary error function centered on

mu.Z.min = mu.Z.min.0 + gamma.mu.Z.min.D*log10(D) + gamma.mu.Z.min.Fz*T,

and with dispersion

sigma.Z.min = sigma.Z.min.0*(Fz^gamma.sigma.Z.min.Fz)*(D^gamma.sigma.Z.min.D). By default, the distribution is not truncated.

At the high values end, the distribution can be truncated with an hard cut at Z.max. By default, the distribution is not truncated.

Breaking of the linearity is allowed too. At Z.knee, the linear relation abruptly changes to

Y = alpha.YIZ.knee + beta.YIZ.knee*Z + gamma.YIZ.knee*T + delta.YIZ.knee*Z*T +-sigma.YIZ.0.knee.

The transition function is

fknee= 1/(1+exp((Z-Zknee)/lknee),

where lknee is the transition length.

Priors on regression parameters can be expressed either as numbers or as character strings. If numbers, delta priors are adopted, the parameters are fixed to the input values, and they are fixed during the the regression. If strings, the prior distribution of the parameter is modeled after the argument. The syntax follows JAGS, see the JAGS user manual at http://sourceforge.net/projects/mcmc-jags/files/Manuals/ for the complete list of distributions. The following shortcuts can be used: "dt" for the Student's" t-distribution "dt(0,1,1)", "dgamma" for the gamma distribution "dgamma(1.0/n.large,1.0/n.large)"; "dunif" for the uniform distribution "dunif(-n.large,n.large)". The distribution "prec.dgamma" is only applicable to dispersions and standard deviations. It means that the precision, i.e., the inverse of the variance, follows a "dgamma" distribution. "ddirch" is the Dirichlet distribution "ddirch(alpha)", with alpha=c(1,..,1).

If beta.YIZ.knee="beta.YIZ", no breaking occurs. Any prior on delta.YIZ.knee or sigma.YIZ.0.knee is superseded, and the conditions delta.YIZ.knee="delta.YIZ" and sigma.YIZ.knee="sigma.YIZ" are assumed.

If redshifts are not provided, all gamma and delta parameters are set to 0.

Value

lira return a LIST

`first` component

A data frame containing the merged chains of all parameters.

`second` component

An mcmc.list object containing the MCMC chains of the regression parameters.

Warning

The dependence packages do not include a copy of the JAGS library, which must be installed separately. For instructions on downloading JAGS, see http://mcmc-jags.sourceforge.net. C++ compilers are also needed.

Note

The vectors x and y can contain missing values NA. The posterior distribution for the corresponding nodes, i.e. the statistical relations, is sampled. Missing values are best suited to the y vector. The samples at the NA locations in the response vector are samples from the posterior predictive distribution and can be summarized just like other model parameters. They can be used to forecast the missing data.

Author(s)

Mauro Sereno <mauro.sereno@oabo.inaf.it>

References

Sereno M., "A Bayesian approach to linear regression in astronomy" (2016), MNRAS 455, 2149-2162; electronic version at http://adsabs.harvard.edu/abs/2016MNRAS.455.2149S.

Sereno M., Ettori S., "CoMaLit-V. Mass forecasting with proxies. Method and application to weak lensing calibrated samples" (2017), MNRAS 468, 3322; electronic version at http://adsabs.harvard.edu/abs/2017MNRAS.468.3322S.

Preprints at http://pico.oabo.inaf.it/~sereno/LIRA/.

See Also

See also distance.LCDM, Hubble.LCDM, jags.model, coda.samples.

Examples

```
## EXAMPLE 1
## sample length
n.sample <- 20
# intrinsic p(Z) distribution of the independet variable
mu.Z.0 <- 0.0; sigma.Z.0 <- 1.0;
Z <- rnorm(n.sample, mean = mu.Z.0, sd = sigma.Z.0)
# X is unscattered
X <- Z
# observational uncertainties in X
delta.x.sd <- 0.04
delta.x <- rnorm(n.sample, sd = delta.x.sd)
x <- X + delta.x
# scaling relation Y-Z
alpha.YIZ <- 0.0; beta.YIZ <- 1.0; sigma.YIZ.0 <- 0.1;
# dependent variable
Y <- alpha.YIZ + beta.YIZ*Z + rnorm(n.sample, sd = sigma.YIZ.0)
# observational uncertainties in Y
delta.y.sd <- 0.05
delta.y <- rnorm(n.sample, sd = delta.y.sd)
y <- Y + delta.y
data.all <- data.frame(x=x, y=y, delta.x=delta.x.sd, delta.y=delta.y.sd)
# regression
samples <- lira(x, y, delta.x=delta.x, delta.y=delta.y, n.adapt=100, n.iter=1000)

## EXAMPLE 2
## Malmquist bias: only objects with y > y.th are selected.
y.th <- -0.5
# the sample is truncated
data.sel <- data.all[which(data.all$y>y.th),]
n.sel <- nrow(data.sel)
# regression
samples.MB <- lira(data.sel$x, data.sel$y, data.sel$delta.x, data.sel$delta.y,
y.threshold = rep(y.th,n.sel), n.adapt=100, n.iter=1000, print.summary=FALSE)
```

```
## EXAMPLE 3
## Upper limits: objects with y < y.th are left-censored
# censoring
data.ul <- data.all
index.ul <- which(data.ul$y<y.th)
data.ul[index.ul,]$y <- NA
data.ul[index.ul,]$delta.y <- 2
# upper limits for undetected (detected) objects are set to the threshold value (NA).
y.upperlimit <- rep(NA, n.sample)
y.upperlimit[index.ul] <- y.th
# the regression recovers the parameters of the scaling relation and
# forecasts the Y-values of the censored objects.
samples.UL <- lira(data.ul$x, data.ul$y, data.ul$delta.x, data.ul$delta.y,
y.upperlimit = y.upperlimit, n.adapt=100, n.iter=1000, Y.monitored=TRUE)


## Further examples can be found at http://pico.oabo.inaf.it/~sereno/LIRA/.
```

Index