Package 'exPrior'

November 15, 2019

Type Package

Title Prior Distributions Using a Bayesian Hierarchical Framework

Version 1.0.1

Maintainer Falk Hesse <falk.hesse@ufz.de>

Description The aim of this package is to provide practitioners of statistics in geology, hydrology, etc. a tool to derive prior distributions for Bayesian inference. Prior distributions summarize knowledge from studies at similar sites. The main features of the package are to (i) generate prior distributions based on external data only; (ii) to account for possible autocorrelation in the data, and (iii) to account for available soft data, say, in the form of expert information on bounds and moments.

Depends R (>= 3.1.0), nimble (>= 0.7.0)

Imports plyr, stats, ggplot2, reshape2, gtable, grid, coda

License MIT + file LICENSE

LazyData TRUE

RoxygenNote 6.1.1

Suggests knitr, rmarkdown, devtools, testthat

VignetteBuilder knitr

NeedsCompilation no

Author Falk Hesse [cre],
Nura Kawa [aut],
Karina Cucchi [aut]

Repository CRAN

Date/Publication 2019-11-15 00:00:03 UTC

R topics documented:

`cap_prior` *calculates the prior according to the Carsel and Parrish methodology*

Description

`cap_prior` calculates the prior according to the Carsel and Parrish methodology

Usage

```
cap_prior(meas, theta)
```

Arguments

`meas`	a vector of measurements
`theta`	values for which to calculate the pdf

Value

the corresponding pdf

`df_porosity` *566 porosity values*

Description

A dataset containing 566 porosity values from 6 different sites

Usage

```
df_porosity
```

Format

A data frame with 566 rows and 2 variables:

site_id id of the site, categorical

val porosity value, non-dimensional fraction ...

Source

https://github.com/GeoStat-Bayesian/geostatDB

`genExPrior` *prior using from external data*

Description

`genExPrior` generates priors from a set of data from multiple sites

Usage

```
genExPrior(exdata, theta, niter = 10^5, range_alpha = NULL,
  seed = NULL, hierarchicalSigma = F, spatialCoordinates = F,
  verbose = F)
```

Arguments

`exdata`	a dataframe containing external data to assimilate, with fields val and site_id (see example)
`theta`	a vector of numerical values of informative prior evaluation points
`niter`	(optional) an integer for the number of samples to use in the MCMC
`range_alpha`	(optional) a vector of two values corresponding to the lower and the upper bounds of the uniform distribution for alpha
`seed`	(optional) a value for the seed to make calls of genExPrior deterministic
`hierarchicalSigma`	(optional) a boolean specifying whether the site-specific variance is defined hierarchically by an inverse-gamma distribution (T) or by a prior (F)
`spatialCoordinates`	(optional) a boolean specifying whether spatial coordinates are provided as covariates to numerical external data. If T, the spatial autocorrelation of external data is accounted for, assuming that the spatial covariance has an exponential form.
`verbose`	(optional) boolean indicating whether R should print information from the progress

Value

the pdf at values corresponding to theta

Examples

```
theta <- seq(from=-5,to=5,by=1)
exdata <- data.frame(val=c(c(2,3,4),c(2,1),c(6,7,2,3)),
                  site_id=c(rep("a",3),rep("b",2),rep("c",4)),
                  x = c(c(2,3,4),c(2,3),c(2,2,3,3)),
                  y = c(c(2,2,3),c(3,2),c(2,3,2,3)))
genExPrior(exdata=exdata,theta=theta)
```

---

goodness_of_fit *tests goodness of fit to normal distribution*

Description

goodness_of_fit tests the closeness to the normal distribution

Usage

```
goodness_of_fit(y)
```

Arguments

y	a vector of samples for which to test the normality

Value

the Kolmogorov-Smirnov statistics

---

johnson_ln *log transform*

Description

johnson_ln is the log-transform (1st Johnson transform)

Usage

```
johnson_ln(x, a, b)
```

Arguments

x	vector of original dataset to transform
a	useless argument
b	useless argument

Value

the transformed sample

`johnson_sb` *log ratio transform*

Description

`johnson_sb` is the second Johnson transform

Usage

```
johnson_sb(x, a, b)
```

Arguments

`x`	vector of original dataset to transform
`a`	useless argument
`b`	useless argument

Value

the transformed sample

`johnson_su` *hyperbolic arcsine transform*

Description

`johnson_su` is the third Johnson transform

Usage

```
johnson_su(x, a, b)
```

Arguments

`x`	vector of original dataset to transform
`a`	useless argument
`b`	useless argument

Value

the transformed sample

KL_divergence *Kullback-Leibler divergence*

Description

KL_divergence calculates the Kullback-Leibler divergence

Usage

```
KL_divergence(theta, p_theta, q_theta)
```

Arguments

theta	a vector of numerics containing values of the RV θ
p_theta	a vector of numerics containing values of the pdf p at locations theta
q_theta	a vector of numerics containing values of the pdf q at locations theta

Value

The Kullback-Leibler divergence from q to p

Examples

```
theta=seq(from=-5,to=5,by=0.1)
p_theta = dnorm(theta,mean = 0.2,sd = 1)
q_theta = dnorm(theta,mean = 0.25,sd = 0.5)
```

KS_distance *Kolmogorov-Smirnov distance*

Description

KS_distance calculates the Kolmogorov-Smirnov distance between two pdfs.

Usage

```
KS_distance(theta, p_theta, q_theta)
```

Arguments

theta	a vector of numerics containing values of the RV θ
p_theta	a vector of numerics containing values of the pdf p at locations theta
q_theta	a vector of numerics containing values of the pdf q at locations theta

Value

The Kolmogorov-Smirnov distance between p and q

Examples

```
theta=seq(from=-5,to=5,by=0.1)
p_theta = dnorm(theta,mean = 0.2,sd = 1)
q_theta = dnorm(theta,mean = 0.25,sd = 0.5)
```

`multiplot` *multiplot function*

Description

`multiplot` allows to combines multiple plots

Usage

```
multiplot(..., plotlist = NULL, file, cols = 1, layout = NULL)
```

Arguments

`...`	ggplot objects
`plotlist`	list of the ggplot objects
`file`	filename
`cols`	Number of columns in layout
`layout`	A matrix specifying the layout. If present, 'cols' is ignored

Value

a plot

`normalize_pdf` *Normalize a pdf*

Description

`normalize_pdf` normalizes a pdf so that the integral of the pdf is equal to 1

Usage

```
normalize_pdf(x, p_x)
```

Arguments

x	a vector corresponding to values of a random variable X (length strictly greater than 1)
p_x	a vector containing the density of the RV X at locations x

Value

The normalized pdf

Examples

```
x <- seq(from=-5,to=5,by=0.1)
p_x <- 2*dnorm(x)
res <- normalize_pdf(x,p_x)
plot(x,p_x)
lines(x,res$p_x)
lines(x,dnorm(x),col='red',lty=2)
```

---

plotExData *plot histogram of measurements*

Description

plotExData plots histogram of ex-situ data as provided to genExPrior

Usage

```
plotExData(exdata, bindwidth = NULL, xrange = NULL, ymax = NULL,
  showLegend = T)
```

Arguments

exdata	a dataframe containing ex-situ data, as provided to the function genExPrior
bindwidth	a numeric specifying the width of the bins (optional)
xrange	a vector with the limits for the x-axis site
ymax	is a numeric specifying the maximum value on the y-axis
showLegend	a boolean indicating whether to show the legend with the names of sites (optional, defaults to true)

Value

a plot

plotExPrior *plot ex-situ prior*

Description

plotExPrior plot informative and non-informative priors

Usage

```
plotExPrior(resExPrior, plotExData = F)
```

Arguments

resExPrior	output from the genExPrior function
plotExData	boolean asking whether to additionally plot the ex-situ data

Value

a plot

Examples

```
exdata <- data.frame(val=c(c(2,3,4),c(2,1),c(6,7,2,3)),
                  site_id=c(rep("a",3),rep("b",2),rep("c",4)),
                  x = c(c(2,3,4),c(2,3),c(2,2,3,3)),
                  y = c(c(2,2,3),c(3,2),c(2,3,2,3)))
ex_prior <- genExPrior(exdata=exdata,theta=seq(from=-5,to=5,by=1))
plotExPrior(ex_prior)
```

plotHyperDist *plot prior and posterior distribution of hyperparameters*

Description

plotHyperDist plot prior and posterior distribution of hyperparameters

Usage

```
plotHyperDist(resExPrior)
```

Arguments

resExPrior	output from the genExPrior function

Value

a plot

Examples

```
exdata <- data.frame(val=c(c(2,3,4),c(2,1),c(6,7,2,3)),
                  site_id=c(rep("a",3),rep("b",2),rep("c",4)),
                  x = c(c(2,3,4),c(2,3),c(2,2,3,3)),
                  y = c(c(2,2,3),c(3,2),c(2,3,2,3)))
ex_prior <- genExPrior(exdata=exdata,theta=seq(from=-5,to=5,by=1))
plotHyperDist(ex_prior)
```

---

smooth_pdf *Smooths a pdf*

Description

smooth_pdf smooths a pdf using convolution with a kernel

Usage

```
smooth_pdf(x, p_x)
```

Arguments

x	a vector corresponding to values of a random variable X (length strictly greater than 1)
p_x	a vector containing the density of the RV X at locations x

Value

The normalized pdf

Examples

```
x <- seq(from=-5,to=5,by=0.1)
p_x <- dnorm(x) + rnorm(length(x),mean=0,sd=0.02)
p_x <- pmax(p_x,0)
plot(x,p_x,type='l')
res <- smooth_pdf(x,p_x)
lines(x,res$p_x,col='red')
```

Index