Package 'GSM'

April 17, 2009

Title Gamma Shape Mixture

Description This package implements a Bayesian approach for estimation of a mixture of gamma distributions in which the mixing occurs over the shape parameter. This family provides a flexible and novel approach for modeling heavy-tailed distributions, it is computationally efficient, and it only requires to specify a prior distribution for a single parameter.

Version 0.1-2

Date 2007-08-10

Author Sergio Venturini

Maintainer Sergio Venturini <sergio.venturini@unibocconi.it>

Depends gtools

License GPL (>= 2)

URL http://www.bepress.com/jhubiostat/paper124

Repository CRAN

Date/Publication 2007-08-09 16:48:51

R topics documented:

GSM-package *Estimation of a Gamma Shape Mixture Model*

Description

This package implements a Bayesian approach for estimation of a mixture of gamma distributions in which the mixing occurs over the shape parameter. This family provides a flexible and novel approach for modeling heavy-tailed distributions, it is computationally efficient, and it only requires to specify a prior distribution for a single parameter. See Venturini et al. (2006).

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

References

Venturini, S., Dominici, F., and Parmigiani, G., "Gamma Shape Mixtures for Heavy-Tailed Distributions" (December 2006). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 124. http://www.bepress.com/jhubiostat/paper124

See Also

gsm, gsm.theta.

allcurves.q *Utility function*

Description

Utility function for plotting a Gamma Shape Mixture Model density.

Usage

```
allcurves.q(postdata,perc)
```

Arguments

postdata	matrix containing of a mixture's density posterior draws.
perc	percentile.

Details

This is a utility function used to generate the credibility bands for a Gamma Shape Mixture density within gsm.plot.

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

See Also

gsm.plot.

---

gsm *Estimation of a Gamma Shape Mixture Model (GSM)*

Description

This function provides the inferential algorithm to estimate a mixture of gamma distributions in which the mixing occurs over the shape parameter. It implements the collapsing approach for the GSM model, as discussed in Venturini et al. (2006).

Usage

```
gsm(y,J,G,M,a,b,alpha)
```

Arguments

y	vector of data.
J	number of mixture components.
G	number of points where to evaluate the GSM density.
M	number of MCMC runs.
a	hyperparameter of the rate parameter prior distribution.
b	hyperparameter of the rate parameter prior distribution.
alpha	hyperparameter of the mixture's weights prior distribution.

Details

Suggestions on how to choose J, a and b are provided in Venturini et al. (2006). In that work the alpha vector is always set at (1/J,...,1/J), but here one is free to choose the value of the generic element of alpha.

Value

List with the following components:

J	number of mixture components used in the GSM model.
a	hyperparameter of the rate parameter prior distribution used in the GSM model.
b	hyperparameter of the rate parameter prior distribution used in the GSM model.
alpha	hyperparameter of the mixture's weights prior distribution used in the GSM model.

`ff`	matrix containing the posterior draws for the mixture's density.
`y.grid`	vector of values used to evalute the GSM density.
`theta`	vector containing the posterior draws for the mixture's rate parameter.
`label`	matrix containing the posterior draws for the mixture's hidden label.
`weight`	matrix containing the posterior draws for the mixture's weights.

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

References

Venturini, S., Dominici, F., and Parmigiani, G., "Gamma Shape Mixtures for Heavy-Tailed Distributions" (December 2006). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 124. http://www.bepress.com/jhubiostat/paper124

See Also

`gsm.theta`, `gsm.plot`.

Examples

```
set.seed(2040)
y <- rgsm(500,c(.1,.3,.4,.2),1)
burnin <- 100
J <- 250
gsm.out <- gsm(y,J,300,burnin+500,6500,340,1/J)
gsm.plot(gsm.out,y,ndens=0,nbin=20,histogram=TRUE)
```

`gsm.plot` *Plot of a Gamma Shape Mixture Model*

Description

This function plots the output of a Gamma Shape Mixture estimation procedure.

Usage

```
gsm.plot(v,y,ndens=5,xlim=c(min(y),max(y)),ylim=c(0,max(v$ff)),xlab="x",ylab="de
```

Arguments

v	list returned by the gsm or gsm.theta functions.
y	vector of data to use in the analysis.
ndens	number of simulated density curves to plot.
xlim	the x limits (x1, x2) of the plot.
ylim	the y limits of the plot.
xlab	a title for the x axis.
ylab	a title for the y axis.
nbin	number of bins for the histogram.
histogram	logical; if TRUE the histogram is plotted on the figure.
bands	logical; if TRUE the 95% credibility bands are overimposed on the density graph.

Details

To produce a standard histogram with the estimated density curve superimposed on it, simply set ndens to 0 and histogram to TRUE.

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

References

Venturini, S., Dominici, F., and Parmigiani, G., "Gamma Shape Mixtures for Heavy-Tailed Distributions" (December 2006). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 124. http://www.bepress.com/jhubiostat/paper124

See Also

gsm, gsm.theta, rgsm.

Examples

```
set.seed(2040)
y <- rgsm(500,c(.1,.3,.4,.2),1)
burnin <- 100
J <- 250
gsm.out <- gsm(y,J,300,burnin+500,6500,340,1/J)
par(mfrow=c(2,2))
gsm.plot(gsm.out,y,ndens=5,nbin=20)
gsm.plot(gsm.out,y,ndens=0,nbin=20,histogram=TRUE)
gsm.plot(gsm.out,y,ndens=0,nbin=20,histogram=TRUE,bands=TRUE)
gsm.plot(gsm.out,y,ndens=5,nbin=20,histogram=TRUE,bands=TRUE)
```

gsm.theta *Estimation of a Gamma Shape Mixture Model (GSM)*

Description

This function provides the inferential algorithm to estimate a mixture of gamma distributions in which the mixing occurs over the shape parameter. It implements the standard approach for the GSM model, as discussed in Venturini et al. (2006).

Usage

```
gsm.theta(y,J,G,M,a,b,alpha)
```

Arguments

y	vector of data.
J	number of mixture components.
G	number of points where to evaluate the GSM density.
M	number of MCMC runs.
a	hyperparameter of the rate parameter prior distribution.
b	hyperparameter of the rate parameter prior distribution.
alpha	hyperparameter of the mixture's weights prior distribution.

Details

Suggestions on how to choose J, a and b are provided in Venturini et al. (2006). In that work the alpha vector is always set at (1/J,...,1/J), but here one is free to choose the value of the generic element of alpha.

Value

List with the following components:

J	number of mixture components used in the GSM model.
a	hyperparameter of the rate parameter prior distribution used in the GSM model.
b	hyperparameter of the rate parameter prior distribution used in the GSM model.
alpha	hyperparameter of the mixture's weights prior distribution used in the GSM model.
ff	matrix containing the posterior draws for the mixture's density.
y.grid	vector of values used to evalute the GSM density.
theta	vector containing the posterior draws for the mixture's rate parameter.
label	matrix containing the posterior draws for the mixture's hidden label.
weight	matrix containing the posterior draws for the mixture's weights.

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

References

Venturini, S., Dominici, F., and Parmigiani, G., "Gamma Shape Mixtures for Heavy-Tailed Distributions" (December 2006). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 124. http://www.bepress.com/jhubiostat/paper124

See Also

gsm, gsm.plot.

Examples

```
set.seed(2040)
y <- rgsm(500,c(.1,.3,.4,.2),1)
burnin <- 100
J <- 250
gsm.out <- gsm.theta(y,J,300,burnin+500,6500,340,1/J)
gsm.plot(gsm.out,y,ndens=0,nbin=20,histogram=TRUE)
```

GSMDist *Utility function*

Description

Density function evaluation for a Gamma Shape Mixture Model.

Usage

```
dgsm(x,weight,rateparam)
pgsm(q,weight,rateparam)
rgsm(n,weight,rateparam)
```

Arguments

x, q	vector of quantiles.
n	number of observations.
weight	vector of mixture weights.
rateparam	reciprocal of the shape parameter, as in GammaDist.

Details

The parametrisation implemented in this function is described in Venturini et al. (2006).

Value

dgsm gives the density, pgsm gives the distribution function, and rgsm generates random deviates.

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

References

Venturini, S., Dominici, F., and Parmigiani, G., "Gamma Shape Mixtures for Heavy-Tailed Distributions" (December 2006). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 124. http://www.bepress.com/jhubiostat/paper124

See Also

dgamma, pgamma, rgamma.

`prob.predict` *Tail probability estimation for a Gamma Shape Mixture Model*

Description

This function allows to estimate the tail probability of a Gamma Shape Mixture Model using the output of the gsm or gsm.theta procedures.

Usage

```
prob.predict(mcmc.w,mcmc.theta,thresh)
```

Arguments

mcmc.w	matrix of the mixture's weights posterior draws; it is part of the output of the gsm or gsm.theta functions.
mcmc.theta	vector of the mixture's rate parameter posterior draws; it is part of the output of the gsm or gsm.theta functions.
thresh	threshold value.

Details

The tail probability is estimated by applying the standard Rao-Blackwellized estimator on the Gibbs sampling realizations obtained through the gsm or gsm.theta procedures.

Value

A numerical vector containing the posterior draws for the tail probability exceeding the value of thresh.

Author(s)

Sergio Venturini ⟨sergio.venturini@unibocconi.it⟩

References

Venturini, S., Dominici, F., and Parmigiani, G., "Gamma Shape Mixtures for Heavy-Tailed Distributions" (December 2006). Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 124. http://www.bepress.com/jhubiostat/paper124

See Also

gsm, gsm.plot.

Examples

```
set.seed(2040)
y <- rgsm(500,c(.1,.3,.4,.2),1)
burnin <- 100
J <- 250
gsm.out <- gsm(y,J,300,burnin+500,6500,340,1/J)
thresh <- c(0.1,0.5,0.75,1,2)
tail.prob.est <- rep(NA,length(thresh))
tail.prob.true <- rep(NA,length(thresh))
for (i in 1:length(thresh)){
   tail.prob.est[i] <- mean(prob.predict(gsm.out$weight[(burnin+1):600,],gsm.out$theta[(burn
   tail.prob.true[i] <- sum(y>thresh[i])/length(y)
}
qqplot(tail.prob.true,tail.prob.est,main="Q-Q plot of true vs. estimated tail probability")
abline(0,1,lty=2)
```

Index