Package ‘bbefkr’

February 19, 2015

Type Package

Title Bayesian bandwidth estimation and semi-metric selection for the functional kernel regression with unknown error density

Version 4.2

Date 2014-04-29

Depends R (>= 3.0.3), splines

Author Han Lin Shang

Maintainer Han Lin Shang <hanlin.shang@anu.edu.au>

Description Estimating optimal bandwidths for the regression mean function approximated by the functional Nadaraya-Watson estimator and the error density approximated by a kernel density of residuals simultaneously in a scalar-on-function regression. As a by-product of Markov chain Monte Carlo, the optimal choice of semi-metric is selected based on largest marginal likelihood.

License GPL (>= 2)

LazyLoad yes

LazyData yes

URL https://sites.google.com/site/hanlinshangswebsite/

NeedsCompilation no

Repository CRAN

Date/Publication 2014-04-29 07:59:07

R topics documented:

`bbefkr-package` *Bayesian bandwidth estimation for the functional kernel regression with unknown error density*

Description

This package aims to estimate bandwidths in the regression function and kernel-form error density simultaneously, using a Bayesian sampling algorithm. We demonstrate this Bayesian sampling algorithm using a functional nonparametric regression and a semi-functional partial linear model

Details

The regression function is approximated by the functional Nadaraya-Watson estimator, while the unknown error density is approximated by a kernel density of residuals. In both regression function and error density, they depend crucially on the selection of the optimal bandwidths.

Author(s)

Han Lin Shang

Maintainer: Han Lin Shang <H.Shang@soton.ac.uk>

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

X. Zhang and R. D. Brooks and M. L. King (2009) A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation, *Journal of Econometrics*, **153**, 21-32.

See Also

`bayMCMC_np_global`, `bayMCMC_np_local`, `bayMCMC_semi_global`, `bayMCMC_semi_local` `bayMCMC_np_global` *Bayesian bandwidth estimation for a functional nonparametric regression with homoscedastic errors*

Description

Estimate the bandwidths in the regression function approximated by the functional Nadaraya-Watson estimator and kernel-form error density with one global bandwidth, in a functional nonparametric regression

Usage

```
bayMCMC_np_global(data_x, data_y, data_xnew, warm = 1000, M = 1000,
mutprob = 0.44, errorprob = 0.44, mutsizp = 1, errorsizp = 1,
prior_alpha = 1, prior_beta = 0.05, err_int = c(-10, 10),
err_ngrid = 10001, num_batch = 20, step = 10, alpha = 0.95, ...)
```

Arguments

`data_x`	An (n by p) matrix of discretised data points of functional curves
`data_y`	A scalar-valued response of length n
`data_xnew`	A matrix of discretised data points of new functional curve(s)
`warm`	Number of iterations for the burn-in period
`M`	Number of iterations for the Markov chain Monte Carlo (MCMC)
`mutprob`	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function
`errorprob`	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density
`mutsizp`	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function. Its value will be updated at each iteration to achieve the optimal acceptance rate, given the MCMC converges to its target distribution
`errorsizp`	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density. Its value will be updated at each iteration to achieve the optimal acceptance rate, given the MCMC converges to its target distribution
`prior_alpha`	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
`prior_beta`	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
`err_int`	Range of the error-density grid for computing its probability density function and cumulative probability density function
`err_ngrid`	Number of the error-density grid points

num_batch	Number of batches to assess the convergence of the MCMC
step	Thinning parameter. For example, when step=10, it keeps every 10th iteration of the MCMC output
alpha	The nominal coverage probability of the prediction interval, customarily 95 percent
...	Other arguments used to define semi-metric. For a set of smoothed functional data, the semi-metric based on derivative is suggested. For a set of rough functional data, the semi-metric based on the functional principal component analysis is suggested

Details

The Bayesian method estimates the bandwidths in the regression function and kernel-form error density in the context of the functional nonparametric regression with homoscedastic errors. It performs better than the functional cross validation in terms of estimation accuracy, since the latter one does not utilise the information about the unknown error density

Value

xpfinalres	Estimated bandwidths
mhat	Estimated regression function
sif_value	Simulation inefficiency factor
mlikeres	Marginal likelihood calculated using the Chib's (1995) method
acceptnwMCMC	Acceptance rate for sampling bandwidth in the regression function
accepterroMCMC	Acceptance rate for sampling bandwidth in the kernel-form error density
fore.den.mkr	Estimated probability density function of the error
fore.cdf.mkr	Estimated cumulative density function of the error
point forecast	Predicted response
PI	Prediction interval of response

Note

It can be time-consuming when the sample size is large, say above 250

Author(s)

Han Lin Shang

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

X. Zhang and R. D. Brooks and M. L. King (2009) A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation, *Journal of Econometrics*, **153**, 21-32.

F. Ferraty, I. Van Keilegom and P. Vieu (2010) On the validity of the bootstrap in non-parametric functional regression, *Scandinavian Journal of Statistics*, **37**, 286-306.

R. Meyer and J. Yu (2000) BUGS for a Bayesian analysis of stochastic volatility models, *Econometircs Journal*, **3**(2), 198-215.

S. Chib (1995) Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

bayMCMC_np_local, bayMCMC_semi_global

Examples

```
htm = proc.time()
dum = bayMCMC_np_global(data_x = simcurve_smooth_normerr, data_y = simresp_np_normerr,
data_xnew = simcurve_smooth_normerr,warm = 50, M = 100, range.grid=c(0,pi), q=2,
nknot=20)
proc.time() - htm
```

---

bayMCMC_np_local	*Bayesian bandwidth estimation for a functional nonparametric regression with homoscedastic errors*

Description

Estimate the bandwidths in the regression function approximated by the functional Nadaraya-Watson estimator and kernel-form error density with localised bandwidths, in a functional nonparametric regression

Usage

```
bayMCMC_np_local(data_x, data_y, data_xnew, warm = 1000, M = 1000,
mutprob = 0.44, errorprob = 0.44, epsilonprob = 0.44, mutsizp = 1,
errorsizp = 1, epsilonsizp = 1, prior_alpha = 1, prior_beta = 0.05,
err_int = c(-10, 10), err_ngrid = 10001, num_batch = 20,
step = 10, alpha = 0.95, ...)
```

Arguments

data_x	An (n by p) matrix of discretised data points of functional curves
data_y	A scalar-valued response of length n
data_xnew	A matrix of discretised data points of new functional curve(s)
warm	Number of iterations for the burn-in period

M	Number of iterations for the Markov chain Monte Carlo (MCMC)
mutprob	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function
errorprob	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density
epsilonprob	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth adjust factor in the kernel-form error density
mutsizp	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function. Its value will be updated at each iteration to achieve the optimal acceptance rate, given the MCMC converges to its target distribution
errorsizp	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density. Its value will be updated at each iteration to achieve the optimal acceptance rate, given the MCMC converges to its target distribution
epsilonsizp	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth adjustment factor in the kernel-form error density. Its value will be updated at each iteration to achieve the optimal acceptance rate
prior_alpha	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
prior_beta	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
err_int	Range of the error-density grid for computing the probability density function and cumulative probability density function
err_ngrid	Number of the error-density grid points
num_batch	Number of batches to assess the convergence of the MCMC
step	Thinning parameter. For example, when `step=10`, it keeps every 10th iteration of the MCMC output
alpha	The nominal coverage probability of the prediction interval, customarily 95 percent
...	Other arguments used to define semi-metric. For a set of smoothed functional data, the semi-metric based on derivative is suggested. For a set of rough functional data, the semi-metric based on the functional principal component analysis is suggested

Details

The Bayesian method estimates the bandwidths in the regression function and kernel-form error density. It performs better than the functional cross validation in terms of estimation accuracy, since the latter one does not utilise the information about the unknown error density. Furthermore, it can estimate error density more accurate than the Bayesian method with a global bandwidth

Value

`xpfinalres`	Estimated bandwidths
`mhat`	Estimated regression function
`sif_value`	Simulation inefficiency factor
`mlikeres`	Marginal likelihood calculated using the Chib's (1995) method
`acceptnwMCMC`	Acceptance rate for sampling bandwidth in the regression function
`accepterroMCMC`	Acceptance rate for sampling bandwidth in the kernel-form error density
`acceptepsilonMCMC`	Acceptance rate for sampling bandwidth adjustment factor in the kernel-form error density
`fore.den.mkr`	Estimated probability density function of the error
`fore.cdf.mkr`	Estimated cumulative density function of the error
`point forecast`	Predicted response
`PI`	Prediction interval of response

Note

It can be time-consuming when the sample size is large, say above 250

Author(s)

Han Lin Shang

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a functional nonparametric regression model with unknown error density, *Computational Statistics and Data Analysis*, 67, 185-198.

X. Zhang and R. D. Brooks and M. L. King (2009) A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation, *Journal of Econometrics*, **153**, 21-32.

F. Ferraty, I. Van Keilegom and P. Vieu (2010) On the validity of the bootstrap in non-parametric functional regression, *Scandinavian Journal of Statistics*, **37**, 286-306.

R. Meyer and J. Yu (2000) BUGS for a Bayesian analysis of stochastic volatility models, *Econometircs Journal*, **3**(2), 198-215.

S. Chib (1995) Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

`bayMCMC_np_global`, `bayMCMC_semi_local` bayMCMC_semi_global *Bayesian bandwidth estimation for a semi-functional partial linear model*

Description

Estimate the bandwidths in both the regression function containing real-valued and function-valued predictors and kernel-form error density with a global bandwidth, in a semi-functional partial linear model

Usage

```
bayMCMC_semi_global(data_x, data_y, data_xnew, Xvar, Xvarpred, warm = 1000, M = 1000,
mutprob = 0.44, errorprob = 0.44, mutsizp = 1, errorsizp = 1, prior_alpha = 1,
prior_beta = 0.05, err_int = c(-10, 10), err_ngrid = 10001,
num_batch = 20, step = 10, alpha, ...)
```

Arguments

data_x	An (n by p) matrix of discretised data points of functional curves
data_y	A scalar-valued response of length n
data_xnew	A matrix of discretised data points of new functional curve(s)
Xvar	Real-valued predictors. For example, n by 2 matrix
Xvarpred	Real-valued predictors for prediction.
warm	Number of iterations for the burn-in period
M	Number of iterations for the Markov chain Monte Carlo (MCMC)
mutprob	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function
errorprob	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density
mutsizp	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function. Its value will be updated at each iteration to achieve the optimal acceptance rate
errorsizp	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density. Its value will be updated at each iteration to achieve the optimal acceptance rate
prior_alpha	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
prior_beta	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
err_int	Range of the error-density grid for computing the probability density function and cumulative probability density function
err_ngrid	Number of the error-density grid points

num_batch	Number of batches to assess the convergence of the MCMC
step	Thinning parameter. For example, when step=10, it keeps every 10th iteration of the MCMC output
alpha	The nominal coverage probability of the prediction interval, customarily 95 percent
...	Other arguments used to define semi-metric. For a set of smoothed functional data, the semi-metric based on derivative is suggested. For a set of rough functional data, the semi-metric based on the functional principal component analysis is suggested

Details

The Bayesian method estimates the bandwidths in the regression function and kernel-form error density in the context of semi-functional partial linear model with homoscedastic errors. It performs better than the functional cross validation in terms of estimation accuracy, since the latter one does not utilise the information about the unknown error density

Value

xpfinalres	Estimated bandwidths
mhat	Estimated regression function for the nonparametric part of the regression function
betahat	Estimated regression coefficient for the parametric part of the regression function
sif_value	Simulation inefficiency factor
mlikeres	Marginal likelihood calculated using the Chib's (1995) method
acceptnwMCMC	Acceptance rate for sampling bandwidth in the regression function
accepterroMCMC	Acceptance rate for sampling bandwidth in the kernel-form error density
fore.den.mkr	Estimated probability density function of the error
fore.cdf.mkr	Estimated cumulative density function of the error
point forecast	Predicted response
PI	Prediction interval of response

Note

It can be time-consuming when the sample size is large, say above 250

Author(s)

Han Lin Shang

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

X. Zhang and R. D. Brooks and M. L. King (2009) A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation, *Journal of Econometrics*, **153**, 21-32.

F. Ferraty, I. Van Keilegom and P. Vieu (2010) On the validity of the bootstrap in non-parametric functional regression, *Scandinavian Journal of Statistics*, **37**, 286-306.

R. Meyer and J. Yu (2000) BUGS for a Bayesian analysis of stochastic volatility models, *Econometircs Journal*, **3**(2), 198-215.

S. Chib (1995) Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

`bayMCMC_np_global`, `bayMCMC_np_local`, `bayMCMC_semi_local`

Examples

```
htm = proc.time()
dum = bayMCMC_semi_global(data_x = simcurve_smooth_normerr, data_y = simresp_semi_normerr,
data_xnew = simcurve_smooth_normerr, Xvar = Xvar, warm = 50, M = 100, range.grid=c(0,pi),
q=2, nknot=20)
proc.time() - htm
```

`bayMCMC_semi_local` *Bayesian bandwidth estimation for a semi-functional partial linear model*

Description

Estimate the bandwidths in both the regression function containing real-valued and function-valued predictors and kernel-form error density with localised bandwidths, in a semi-functional partial linear model

Usage

```
bayMCMC_semi_local(data_x, data_y, data_xnew, Xvar, Xvarpred, warm = 1000, M = 1000,
mutprob = 0.44, errorprob = 0.44, epsilonprob = 0.44, mutsizp = 1, errorsizp = 1,
epsilonsizp = 1, prior_alpha = 1, prior_beta = 0.05,
err_int = c(-10, 10), err_ngrid = 10001, num_batch = 20,
step = 10, alpha, ...)
```

Arguments

`data_x`	An (n by p) matrix of discretised data points of functional curves
`data_y`	A scalar-valued response of length n
`data_xnew`	A matrix of discretised data points of new functional curve(s)
`Xvar`	Real-valued predictors. For example, n by 2 matrix
`Xvarpred`	Real-valued predictors for prediction.
`warm`	Number of iterations for the burn-in period
`M`	Number of iterations for the Markov chain Monte Carlo (MCMC)
`mutprob`	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function
`errorprob`	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density
`epsilonprob`	Optimal acceptance rate of the random-walk Metropolis algorithm for sampling the bandwidth adjustment factor in the kernel-from error density
`mutsizp`	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the regression function. Its value will be updated at each iteration to achieve the optimal acceptance rate
`errorsizp`	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth in the kernel-form error density. Its value will be updated at each iteration to achieve the optimal acceptance rate
`epsilonsizp`	Initial step length of the random-walk Metropolis algorithm for sampling the bandwidth adjustment factor in the kernel-form error density. Its value will be updated at each iteration to achieve the optimal acceptance rate
`prior_alpha`	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
`prior_beta`	Hyperparameter of the inverse gamma prior distribution for the squared bandwidths
`err_int`	Range of the error-density grid for computing the probability density function and cumulative probability density function
`err_ngrid`	Number of the error-density grid points
`num_batch`	Number of batches to assess the convergence of the MCMC
`step`	Thinning parameter. For example, when `step=10`, it keeps every 10th iteration of the MCMC output
`alpha`	The nominal coverage probability of the prediction interval, customarily 95 percent
`...`	Other arguments used to define semi-metric. For a set of smoothed functional data, the semi-metric based on derivative is suggested. For a set of rough functional data, the semi-metric based on the functional principal component analysis is suggested

Details

The Bayesian method estimates the bandwidths in the regression function and kernel-form error density in the context of semi-functional partial linear model with homoscedastic errors. It performs better than the functional cross validation in terms of estimation accuracy, since the latter one does not utilise the information about the unknown error density

Value

`xpfinalres`	Estimated bandwidths
`mhat`	Estimated regression function for the nonparametric part of the regression function
`betahat`	Estimated regression coefficient for the parametric part of the regression function
`sif_value`	Simulation inefficiency factor
`mlikeres`	Marginal likelihood calculated using the Chib's (1995) method
`acceptnwMCMC`	Acceptance rate for sampling bandwidth in the regression function
`accepterroMCMC`	Acceptance rate for sampling bandwidth in the kernel-form error density
`acceptepsilonMCMC`	Acceptance rate for sampling bandwidth adjustment factor in the kernel-form error density
`fore.den.mkr`	Estimated probability density function of the error
`fore.cdf.mkr`	Estimated cumulative density function of the error
`point forecast`	Predicted response
`PI`	Prediction interval of response

Note

It can be time-consuming when the sample size is large, say above 250

Author(s)

Han Lin Shang

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

X. Zhang and R. D. Brooks and M. L. King (2009) A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation, *Journal of Econometrics*, **153**, 21-32.

F. Ferraty, I. Van Keilegom and P. Vieu (2010) On the validity of the bootstrap in non-parametric functional regression, *Scandinavian Journal of Statistics*, **37**, 286-306.

R. Meyer and J. Yu (2000) BUGS for a Bayesian analysis of stochastic volatility models, *Econometircs Journal*, **3**(2), 198-215.

S. Chib (1995) Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

`bayMCMC_np_global`, `bayMCMC_np_local`, `bayMCMC_semi_global`

`error.den` *Compute the probability density function and cumulative probability density function of the error, using a global bandwidth of residuals*

Description

With the estimated bandwidth of residuals, error density can be approximated by the kernel density estimator.

Usage

```
error.den(band, eps, res.data)
```

Arguments

`band`	Bandwidth of residuals
`eps`	Grid point
`res.data`	Residuals obtained from the estimated conditional mean

Value

Numerical values

Author(s)

Han Lin Shang

References

H. L. Shang (2013) ‘Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density’, *Computational Statistics and Data Analysis*, **67**, 185-198.

H. L. Shang (2013) ‘Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density’, *Computational Statistics*, in press.

See Also

`error.denadj`

`error.denadj` *Compute the probability density function and cumulative probability density function of error, using localised bandwidths of residuals*

Description

With the estimated bandwidths of residuals, error density can be approximated by the kernel density estimator.

Usage

```
error.denadj(ban, badj, eps, res.data)
```

Arguments

`ban`	Bandwidth of residuals
`badj`	Bandwidth adjustment factor
`eps`	Grid point
`res.data`	Residuals obtained from the estimated conditional mean

Value

Numerical values

Author(s)

Han Lin Shang

References

H. L. Shang (2013) 'Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density', *Computational Statistics and Data Analysis*, **67**, 185-198.

H. L. Shang (2013) 'Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density', *Computational Statistics*, in press.

See Also

`error.den`

funopare.kernel *Functional Nadaraya-Watson estimator*

Description

It implements the functional Nadaraya-Watson estimator to estimate the regression function. It depends on the type of semi-metric used as well as the optimal selection of bandwidth parameter

Usage

```
funopare.kernel(Response, CURVES, PRED, bandwidth, ..., kind.of.kernel = "quadratic",
semimetric = "deriv")
```

Arguments

Response	A real-valued scalar response of length n
CURVES	An (n by p) matrix of discretised data of functional curves
PRED	An (n by k) matrix of discretised data of functional curves. PRED can be the same as the CURVES or the discretised data points of a new functional curve
bandwidth	A real-valued bandwidth parameter
...	Other arguments
kind.of.kernel	Type of kernel function. By default, it is the Epanechnikov kernel
semimetric	Type of semi-metric. By default, it is the semi-metric based on the qth order derivative, where q is an integer

Details

The functional NW estimator of the conditional mean can be expressed as a weighted average of response variable: $\sum_{i=1}^{n} K_h(d(x_i, x))y_i / \sum_{i=1}^{n} K_h(d(x_i, x))$, where $K(\cdot)$ is a kernel function which integrates to one, it has continuous derivative on the function support range. The semi-metric d is used to measure distances among curves. For a set of smooth curves, the semi-metric based on derivative should be considered. For a set of rough curves, the semi-metric based on functional principal components should be used. The bandwidth h controls the tradeoff between squared bias and variance in the mean squared error

Value

NWweit	Estimated Nadaraya-Watson weights
Estimated.values	Estimated values of the regression function
Predicted.values	Predicted values of the regression function
band	Bandwidth of the functional NW estimator
Mse	In-sample mean squared error

Author(s)

Han Lin Shang

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

X. Zhang and R. D. Brooks and M. L. King (2009) A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation, *Journal of Econometrics*, **153**, 21-32.

F. Ferraty, I. Van Keilegom and P. Vieu (2010) On the validity of the bootstrap in non-parametric functional regression, *Scandinavian Journal of Statistics*, **37**, 286-306.

F. Ferraty and P. Vieu (2006) Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York.

F. Ferraty and P. Vieu (2002) The functional nonparametric model and application to spectrometric data, *Computational Statistics*, **17**, 545-564.

See Also

`bayMCMC_np_global`, `bayMCMC_np_local`, `bayMCMC_semi_global`, `bayMCMC_semi_local`

Examples

```
funopare.kernel(Response = simresp_np_normerr, CURVES = simcurve_smooth_normerr,
PRED = simcurve_smooth_normerr, bandwidth = 2.0, range.grid=c(0,pi), q=2, nknot=20)
```

`logdensity_admkr` *Compute the marginal likelihood using Chib's (1995) method*

Description

The log marginal likelihood can be computed as the log likelihood + log prior - log posterior. The latter one can be estimated from the MCMC.

Usage

```
logdensity_admkr(tau2, cpost)
```

Arguments

`tau2`	Squared bandwidths recorded at each iteration
`cpost`	Squared bandwidths recorded at only the iteration defined by the variable `step`

Value

The value of the estimated posterior density

Author(s)

Han Lin Shang

References

S. Chib (1995) 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

`logpriors_admkr`, `loglikelihood_global_admkr`

`loglikelihood_global_admkr` *Compute the marginal likelihood using Chib's (1995) method*

Description

The log marginal likelihood can be computed as the log likelihood + log prior - log posterior. The latter one can be estimated from the MCMC.

Usage

```
loglikelihood_global_admkr(h, resid)
```

Arguments

`h`	Estimated bandwidth
`resid`	Estimated residuals

Value

The value of log likelihood

Author(s)

Han Lin Shang

References

S. Chib (1995) 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

logpriors_admkr, logdensity_admkr

---

logpriorh2 *Prior density of the squared bandwidth parameters*

Description

Inverse gamma distribution is used as the prior density of bandwidth parameters

Usage

```
logpriorh2(h2, prior_alpha, prior_beta)
```

Arguments

h2	Squared bandwidths
prior_alpha	Hyper-parameter of the inverse gamma distribution
prior_beta	Hyper-parameter of the inverse gamma distribution

Value

Prior density

Author(s)

Han Lin Shang

References

H. L. Shang (2013) 'Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density', *Computational Statistics and Data Analysis*, **67**, 185-198.

H. L. Shang (2013) 'Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density', *Computational Statistics*, in press.

`logpriors_admkr` *Compute the marginal likelihood using Chib's (1995) method*

Description

The log marginal likelihood can be computed as the log likelihood + log prior - log posterior. The latter one can be estimated from the MCMC.

Usage

```
logpriors_admkr(h2, prior_alpha, prior_beta)
```

Arguments

`h2`	Squared bandwidths
`prior_alpha`	Hyper-parameter of the inverse gamma distribution
`prior_beta`	Hyper-parameter of the inverse gamma distribution

Value

The value of log prior density in the marginal likelihood

Author(s)

Han Lin Shang

References

S. Chib (1995) 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association*, **90**(432), 1313-1321.

See Also

`logdensity_admkr`, `loglikelihood_global_admkr`

`SIF` *Simulation inefficiency factor*

Description

Calculate the simulation inefficiency factor, as a means of checking the convergence of MCMC

Usage

```
SIF(BAND_MATRIX, NUM_ITERATIONS, NUM_BATCH)
```

Arguments

`BAND_MATRIX`	Bandwidth parameters drawn from the MCMC
`NUM_ITERATIONS`	Number of total iterations
`NUM_BATCH`	Number of batch samples

Value

Numerical values

Author(s)

Han Lin Shang

References

S. Kim, N. Shephard and S. Chib (1998) 'Stochastic volatility: likelihood inference and comparison with ARCH models', *The Review of Economic Studies*, **65**(3), 361-393.

R. Meyer and J. Yu (2000) 'BUGS for a Bayesian analysis of stochastic volatility models', *Econometrics Journal*, **3**(2), 198-215.

Y. K. Tse, X. Zhang and J. Yu (2004) 'Estimation of hyperbolic diffusion using the Markov chain Monte Carlo method', *Quantitative finance*, **4**(2), 158-169.

X. Zhang, R. D. Brooks and M. L. King (2009) 'A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation', *Journal of Econometrics*, **153**(1), 21-32.

---

simcurve_smooth_normerr *Simulated data set*

Description

Simulated data

Usage

```
data(simcurve_smooth_normerr)
```

Format

`simcurve_smooth_normerr`: 50 by 100

`simcurve_rough_normerr`: 50 by 100

`simresp_normerr`: 50 by 1

`tau_normerr`: 50 by 1

Details

The simulated discretised curves are defined as $x_i(t_j) = a_i cos(2t_j) + b_i sin(4t_j) + c_i(t_j^2 - \pi \times t_j + 2/9\pi^2)$, where t represents the function support range and $0 \leq t_1 \leq t_2 \ldots \leq \pi$ are equispaced points within the function support range, a_i, b_i and c_i are independently drawn from a uniform distribution on [0,1], and n represents the sample size. For simulating a set of rough curves, we add one extra term d_j generated from $U(-0.1, 0.1)$. Having defined functional curves, we then construct the regression mean function $\tau = 10 \times (a_i^2 - b_i^2)$. Then the response variable is obtained by adding the regression mean function with a set of errors generated from a standard normal distribution

Source

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

F. Ferraty, I. Van Keilegom, P. Vieu (2010) On the validity of the bootstrap in non-parametric functional regression, *Scandinavian Journal of Statistics*, **37**(2), 286-306.

Examples

```
data(simcurve_normerr)
```

`simulate_error` *Simulate errors*

Description

Simulate errors which has the specified error density

Usage

```
simulate_error(samplesize, errordensity)
```

Arguments

`samplesize`	Sample size
`errordensity`	Type of error density

Value

A vector of simulated error, which follows the specific error density

Author(s)

Han Lin Shang

References

J. S. Marron and M. P. Wand (1992) ‘Exact Mean Integrated Squared Error’, Annals of Statistics, 20(2), 712-736.

Examples

```
simulate_error(samplesize = 100, errordensity = "normal")
```

---

specurves — *Spectroscopy tecator data*

Description

This data set is a part of the original one which can be found at http://lib.stat/cmu.edu/datasets/tecator.

Usage

```
data(specurves)
```

Format

specurves: 215 by 100 matrix fat: a vector of length 215 protein: a vector of length 215 moisture: a vector of length 215

Details

For each unit, we observe one spectrometric curve which corresponds to the absorbance measured at 100 wavelengths (from 852 to 1050 in step of 2nm). For each measurement, we have at hand its fat content obtained by an analytic chemical processing

Source

Nonparametric Functional Data Analysis website at http://www.lsp.ups-tlse.fr/staph/npfda/

References

C. Goutis (1998) "Second-derivative functional regression with applications to near infra-red spectroscopy", *Journal of the Royal Statistical Society: Series B*, **60**(1), 103-114.

F. Ferraty and P. Vieu (2002) "The functional nonparametric model and application to spectrometric data", *Computational Statistics*, **17**(4), 545-564.

F. Ferraty and P. Vieu (2003) "Curve discrimination: A nonparametric functional approach", *Computational Statistics and Data Analysis*, **44**(1-2), 161-173.

F. Ferraty and P. Vieu (2003) "Functional nonparametric statistics: A double infinite dimensional framework", Recent advances and trends in nonparametric statistics, Ed M. G. Akritas and D. N. Politis, Amsterdam, The Netherlands, 61-76.

F. Ferraty, I. Van Keilegom, P. Vieu (2010) "On the validity of the bootstrap in non-parametric functional regression", *Scandinavian Journal of Statistics*, **37**, 286-306.

F. Rossi and N. Delannay and B. Conan-Guez and M. Verleysen (2005) "Representation of functional data in neural networks", *Neurocomputing*, **64**, 183-210.

F. Ferraty and P. Vieu (2007) Nonparametric functional data analysis, New York: Springer.

H. Matsui and Y. Araki and S. Konishi (2008) "Multivariate regression modeling for functional data", *Journal of Data Science*, **6**, 313-331.

Examples

```
data(specurves)
```

`Xvar` *Simulated real-valued predictors in the semi-functional partial linear model*

Description

Simulated real-valued predictors in the semi-functional partial linear model. It is a 50 by 2 matrix, where the column variables are generated from $U[0,1]$ and the true regression coefficients are (-1,2). Note that the estimation of the regression coefficient for these predictors depends crucially on the bandwidth parameter estimated in the functional Nadaraya-Watson estimator of the regression function

Usage

```
data(Xvar)
```

Source

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

References

H. L. Shang (2013) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Computational Statistics*, in press.

H. L. Shang (2013) Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density, *Computational Statistics and Data Analysis*, **67**, 185-198.

Examples

```
data(Xvar)
data(tau_semierr)
data(simresp_semi_normerr)
```

Index