Package 'hermiter'

September 10, 2020

Title Efficient Sequential and Batch Estimation of Probability Density
Functions, Cumulative Distribution Functions and Quantiles

Version 1.0.0

Author Michael Stephanou [aut, cre],
Melvin Varughese [ctb]

Maintainer Michael Stephanou <michael.stephanou@gmail.com>

Description Facilitates estimation of the full probability density function,
cumulative distribution function and quantile function using Hermite series
based estimators. These estimators are particularly useful in the sequential
setting (both stationary and non-stationary) and one-pass batch estimation
setting for large data sets.
Based on: Stephanou, Michael, Varughese, Melvin and Iain Macdonald. ``Sequential quan-
tiles via Hermite series density estimation.'' Electronic Journal of Statistics 11.1 (2017): 570-
607 <doi:10.1214/17-EJS1245> and
Stephanou, Michael and Varughese, Melvin. ``On the properties of Hermite series based distribu-
tion function estimators.'' Metrika (2020) <doi:10.1007/s00184-020-00785-z>.

License MIT + file LICENSE

Imports Rcpp (>= 1.0.5), methods

LinkingTo Rcpp, BH

RoxygenNote 7.1.1

Suggests testthat, magrittr, knitr, rmarkdown, dplyr, data.table,
ggplot2, DT

VignetteBuilder knitr

URL https://github.com/MikeJaredS/hermiter

BugReports https://github.com/MikeJaredS/hermiter/issues

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-09-10 14:50:02 UTC

R topics documented:

---

hermiter-package	*Efficient Sequential and Batch Estimation of Probability Density Functions, Cumulative Distribution Functions and Quantiles*

Description

Facilitates estimation of the full probability density function, cumulative distribution function and quantile function using Hermite series based estimators. These estimators are particularly useful in the sequential setting (both stationary and non-stationary) and one-pass batch estimation setting for large data sets. Based on: Stephanou, Michael, Varughese, Melvin and Iain Macdonald. "Sequential quantiles via Hermite series density estimation." Electronic Journal of Statistics 11.1 (2017): 570-607 <doi:10.1214/17-EJS1245> and Stephanou, Michael and Varughese, Melvin. "On the properties of Hermite series based distribution function estimators." Metrika (2020) <doi:10.1007/s00184-020-00785-z>.

Package Content

Index of help topics:

```
combine_hermite        Combines a list of Hermite estimators
combine_pair           Combines two Hermite estimators
cum_prob               Estimates the cumulative probability for a
                       vector of x values
cum_prob_quantile_helper
```

	Estimates the cumulative probability for quantile estimation
dens	Estimates the probability density for a vector of x values
get_coefficients	Returns Hermite series expansion coefficients
hermite_estimator	A class to sequentially estimate pdfs, cdfs and quantile functions
hermite_function	Outputs orthonormal Hermite functions
hermite_integral_val	Outputs a definite integral of the orthonormal Hermite functions
hermite_integral_val_quantile_adap	
	Outputs a definite integral of the orthonormal Hermite functions
hermite_normalization	Outputs Hermite normalization factors
hermiter-package	Efficient Sequential and Batch Estimation of Probability Density Functions, Cumulative Distribution Functions and Quantiles
quant	Estimates the quantiles at a vector of probability values
quantile_helper	Estimates the quantile at a single probability value
standardizeInputs	Standardizes the observation x and updates the online moment inputs
standardizeInputsEW	Standardizes the observation x and updates the online moment inputs
standardize_value	Standardize a vector of observations x
update_batch	Updates the Hermite series based estimator with a batch of data
update_sequential	Updates the Hermite series based estimator sequentially

Maintainer

Michael Stephanou <michael.stephanou@gmail.com>

Author(s)

Michael Stephanou [aut, cre], Melvin Varughese [ctb]

---

combine_hermite	*Combines a list of Hermite estimators*

Description

This method allows a list of Hermite based estimators of class hermite_estimator to be consistently combined.

Usage

```
combine_hermite(hermite_estimators)
```

Arguments

```
hermite_estimators
```
>   A list of hermite_estimator objects.

Details

Note that the N and standardize arguments must be the same for all estimators in order to combine them. In addition, note that exponentially weighted estimators cannot be combined. If the Hermite estimators are not standardized, the combined estimator will be exactly equivalent to constructing a single estimator on the data set formed by combining the data sets used to update the respective hermite_estimator inputs. If the input Hermite estimators are standardized however, then the equivalence will be approximate.

Value

An object of class hermite_estimator.

Examples

```
hermite_est_1 <- hermite_estimator(N = 10, standardize = FALSE)
hermite_est_1 <- update_batch(hermite_est_1, rnorm(30))
hermite_est_2 <- hermite_estimator(N = 10, standardize = FALSE)
hermite_est_2 <- update_batch(hermite_est_2, rnorm(30))
hermite_combined <- combine_hermite(list(hermite_est_1, hermite_est_2))
```

---

combine_pair   *Combines two Hermite estimators*

Description

This method allows a pair of Hermite based estimators of class hermite_estimator to be consistently combined.

Usage

```
combine_pair(this, hermite_estimator_other)
```

Arguments

```
this                A hermite_estimator object. The first Hermite series based estimator.
hermite_estimator_other
```
>   A hermite_estimator object. The second Hermite series based estimator.

Details

Note that the N and standardize arguments must be the same for the two estimators in order to combine them. In addition, note that exponentially weighted estimators cannot be combined. If the Hermite estimators are not standardized, the combined estimator will be exactly equivalent to constructing a single estimator on the data set formed by combining the data sets used to update the respective hermite_estimator inputs. If the input Hermite estimators are standardized however, then the equivalence will be approximate.

Value

An object of class hermite_estimator.

Examples

```
hermite_est_1 <- hermite_estimator(N = 10, standardize = FALSE)
hermite_est_1 <- update_batch(hermite_est_1, rnorm(30))
hermite_est_2 <- hermite_estimator(N = 10, standardize = FALSE)
hermite_est_2 <- update_batch(hermite_est_2, rnorm(30))
hermite_combined <- combine_pair(hermite_est_1, hermite_est_2)
```

---

cum_prob	*Estimates the cumulative probability for a vector of x values*

Description

This method calculates the cumulative probability values at a vector of x values using the hermite_estimator object (this).

Usage

```
cum_prob(this, x, clipped)
```

Arguments

this	A hermite_estimator object.
x	A numeric vector. Values at which to estimate the cumulative probability
clipped	A boolean value. This value determines whether cumulative probabilities are clipped to lie within the range [0,1].

Details

The object must be updated with observations prior to the use of this method.

Value

A numeric vector of cumulative probability values.

Examples

```
hermite_est <- hermite_estimator(N = 10, standardize = TRUE)
hermite_est <- update_batch(hermite_est, rnorm(30))
cdf_est <- cum_prob(hermite_est, c(0, 0.5, 1))
```

---

cum_prob_quantile_helper

Estimates the cumulative probability for quantile estimation

Description

This helper method uses a modified distribution function estimator which differs from the cum_prob.hermite_estimator.

Usage

```
cum_prob_quantile_helper(this, x)
```

Arguments

this	A hermite_estimator object.
x	A numeric vector.

Details

The modified distribution function estimator appears more accurate for quantile estimation as validated empirically in:

https://projecteuclid.org/euclid.ejs/1488531636

This method is intended for internal use by the hermite_estimator class.

Value

A numeric vector of cumulative probability values.

---

dens

Estimates the probability density for a vector of x values

Description

This method calculates the probability density values at a vector of x values using the hermite_estimator object (this).

Usage

```
dens(this, x, clipped)
```

Arguments

`this`	A hermite_estimator object.
`x`	A numeric vector. Values at which to estimate the probability density.
`clipped`	A boolean value. This value determines whether probability densities are clipped to be bigger than zero.

Details

The object must be updated with observations prior to the use of the method.

Value

A numeric vector of probability density values.

Examples

```
hermite_est <- hermite_estimator(N = 10, standardize = TRUE)
hermite_est <- update_batch(hermite_est, rnorm(30))
pdf_est <- dens(hermite_est, c(0, 0.5, 1))
```

---

get_coefficients   *Returns Hermite series expansion coefficients*

Description

Returns Hermite series expansion coefficients

Usage

```
get_coefficients(this)
```

Arguments

`this`	A hermite_estimator object.

Value

Numeric vector of length N+1

Examples

```
hermite_est_1 <- hermite_estimator(N = 10, standardize = FALSE)
hermite_est_1 <- update_batch(hermite_est_1, rnorm(30))
coefficient_vec <- get_coefficients(hermite_est_1)
```

---

hermite_estimator   *A class to sequentially estimate pdfs, cdfs and quantile functions*

Description

This method constructs an S3 object with associated methods for univariate nonparametric estimation of pdfs, cdfs and quantiles.

Usage

```
hermite_estimator(N = 10, standardize = FALSE, exp_weight_lambda = NA)
```

Arguments

N	An integer between 0 and 100. The Hermite series based estimator is truncated at N+1 terms.
standardize	A boolean value. Determines whether the observations are standardized, a transformation which often improves performance.
exp_weight_lambda	
	A numerical value between 0 and 1. This parameter controls the exponential weighting of the Hermite series based estimator. If this parameter is NA, no exponential weighting is applied.

Details

The hermite_estimator class allows the sequential or one-pass batch estimation of the full probability density function, cumulative distribution function and quantile function. It is well suited to streaming data (both stationary and non-stationary) and to efficient estimation in the context of massive or distributed data sets. Indeed, estimators constructed on different subsets of a distributed data set can be consistently combined.

Value

An S3 object of class hermite_estimator, with methods for density function, distribution function and quantile function estimation.

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

Examples

```
hermite_est <- hermite_estimator(N = 10, standardize = TRUE)
```

---

hermite_function   *Outputs orthonormal Hermite functions*

Description

The method calculates the orthonormal Hermite functions, $h_k(x)$ from $k = 0, \ldots, N$ for the vector of values, x.

Usage

```
hermite_function(N, x, normalization)
```

Arguments

N	An integer number.
x	A numeric vector.
normalization	A numeric vector of normalization factors generated by the hermite_normalization function.

Value

A numeric matrix with N+1 rows and length(x) columns.

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

---

hermite_integral_val   *Outputs a definite integral of the orthonormal Hermite functions*

Description

The method calculates $\int_{-\infty}^{x} h_k(t)dt$ for $k = 0, \ldots, N$ and the vector of values x.

Usage

```
hermite_integral_val(N, x, hermite_function_mat)
```

Arguments

N	An integer number.
x	A numeric vector.
hermite_function_mat	
	A numeric matrix of Hermite function values generated by the function hermite_function.

Value

A numeric matrix with N+1 rows and length(x) columns.

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

---

hermite_integral_val_quantile_adap

Outputs a definite integral of the orthonormal Hermite functions

---

Description

The method calculates $\int_x^\infty h_k(t)dt$ for $k = 0, \ldots, N$ and the vector of values x.

Usage

```
hermite_integral_val_quantile_adap(N, x, hermite_function_mat)
```

Arguments

N   An integer number.

x   A numeric vector.

hermite_function_mat

                A numeric matrix of Hermite function values generated by the function her-
                mite_function.

Value

A numeric matrix with N+1 rows and length(x) columns.

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

---

hermite_normalization   *Outputs Hermite normalization factors*

Description

The method returns numeric normalization factors that, when multiplied by the physicist Hermite polynomials $H_k(x)$, yield orthonormal Hermite functions $h_k(x)$ for $k = 0, \ldots, N$.

Usage

```
hermite_normalization(N)
```

Arguments

N   An integer number.

Value

A numeric vector of length N+1

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

---

quant   *Estimates the quantiles at a vector of probability values*

Description

Estimates the quantiles at a vector of probability values

Usage

```
quant(this, p)
```

Arguments

this	A hermite_estimator object.
p	A numeric vector. A vector of probability values.

Value

A numeric vector. The vector of quantile values associated with the probabilities p.

Examples

```
hermite_est <- hermite_estimator(N = 10, standardize = TRUE)
hermite_est <- update_batch(hermite_est, rnorm(30))
quant_est <- quant(hermite_est, c(0.25, 0.5, 0.75))
```

---

quantile_helper   *Estimates the quantile at a single probability value*

Description

This helper method is intended for internal use by the hermite_estimator class.

Usage

```
quantile_helper(this, p)
```

Arguments

this	A hermite_estimator object.
p	A numeric value. The probability at which to calculate the associated quantile.

Value

A numeric value. The value of the quantile associated with p.

---

standardizeInputs   *Standardizes the observation x and updates the online moment inputs*

Description

Standardizes the observation x and updates the online moment inputs

Usage

```
standardizeInputs(x, n_obs, current_mean, current_var)
```

Arguments

x	A numeric value.
n_obs	A numeric value. The number of observations.
current_mean	A numeric value.
current_var	A numeric value.

Value

A numeric vector. The first element is the updated mean. The second element is the updated variance times n_obs. The third element is the updated, standardized value of x.

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

---

standardizeInputsEW *Standardizes the observation x and updates the online moment inputs*

Description

The online moments are updated via exponential weighting.

Usage

```
standardizeInputsEW(x, n_obs, lambda, current_mean, current_var)
```

Arguments

x	A numeric value.
n_obs	A numeric value. The number of observations.
lambda	A numeric value.
current_mean	A numeric value.
current_var	A numeric value.

Value

A numeric vector. The first element is the updated mean. The second element is the updated variance times n_obs. The third element is the updated, standardized value of x.

Author(s)

Michael Stephanou <michael.stephanou@gmail.com>

---

standardize_value   *Standardize a vector of observations x*

Description

This helper method standardizes the observations x using the online mean and online standard deviation contained in the hermite_estimator object (this).

Usage

```
standardize_value(this, x)
```

Arguments

| this | A hermite_estimator object. |
| x | A numeric vector of observations. |

Value

An object of class hermite_estimator.

---

update_batch   *Updates the Hermite series based estimator with a batch of data*

Description

This method can be applied in one-pass batch estimation settings. This method cannot be used with an exponentially weighted estimator.

Usage

```
update_batch(this, x)
```

Arguments

| this | A hermite_estimator object. |
| x | A numeric vector. A vector of observations to be incorporated into the estimator. |

Value

An object of class hermite_estimator.

Examples

```
hermite_estimator <- hermite_estimator(N = 10, standardize = TRUE)
hermite_estimator <- update_batch(hermite_estimator, x = c(1, 2))
```

update_sequential

Updates the Hermite series based estimator sequentially

Description

This method can be applied in sequential estimation settings.

Usage

```
update_sequential(this, x)
```

Arguments

this	A hermite_estimator object.
x	A numeric value. An observation to be incorporated into the estimator.

Value

An object of class hermite_estimator.

Examples

```
hermite_estimator <- hermite_estimator(N = 10, standardize = TRUE)
hermite_estimator <- update_sequential(hermite_estimator, x = 2)
```

Index