Package 'modelIntegration'

June 20, 2016

Type Package

Title Integration of Probability Distributions

Version 1.0.0

Date 2016-06-17

Author Anna Shchiptsova [aut, cre]

Maintainer Anna Shchiptsova <shchipts@iiasa.ac.at>

Depends R (>= 3.2.5)

Description Combination of prior distributions into an aggregated (synthetic) one. The package implements the posterior integration method, reported in Kryazhimskiy, A.V. (2013). Posterior integration of independent stochastic estimates. IIASA Interim Report. IR-13-006. For comparison, an implementation of simple averaging of the input distributions is added.

License MIT + file LICENSE

LazyData true

RoxygenNote 5.0.1

Suggests testthat, knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

average *Average distribution*

Description

Combines source distributions into the resultant integrated distribution using the average of distribution probabilities. The latter one is called an *average* probability distribution.

Usage

```
average(x, ...)
```

Arguments

x	an R object with source distributions
...	other arguments passed to methods

Details

The probability of the true value to be equal z is computed as *(p1[z]+p2[z]+...+pn[z])/n*, where *p1...pn* are probabilities from x and n is the number of source distributions.

Value

A data frame with the 'table-based' distribution, where every row contains an outcome in the random variable range and its probability in the integrated distribution.

Examples

```
# Average distribution of two 'table-based' distributions
# S3 method for class "distribution_table"
pdf1 <- c(0.25, 0.75)
pdf2 <- c(0.5, 0.5)
bins <- c(1, 2)
average(integrate(bins, list(pdf1, pdf2)))
```

forest_npp *NPP distribution tables*

Description

Probability distribution tables for net primary production (NPP) of the forest ecosystems in seven bioclimatic zones in Russia. A dataset contains distributions estimated with the landscape-ecosystem approach (LEA) and from the ensemble of dynamic global vegetation models (DGVMs).

Usage

```
forest_npp
```

Format

A data frame with 1131 rows and 17 variables

- npp: NPP values
- LEA_Tundra: LEA-based probability for the Tundra climatic zone
- LEA_Tundra_Northern_Taiga: LEA-based probability for the Forest tundra and Northern Taiga climatic zone
- LEA_Middle_Taiga: LEA-based probability for the Middle taiga climatic zone
- LEA_Southern_Taiga: LEA-based probability for the Southern taiga climatic zone
- LEA_Temperate: LEA-based probability for the Temperate climatic zone
- LEA_Steppe: LEA-based probability for the Steppe climatic zone
- LEA_Deserts: LEA-based probability for the Semi-deserts and deserts climatic zone
- LEA_Total: LEA-based probability taken over all climatic zones
- DGVM_Tundra: DGVM-based probability for the Tundra climatic zone
- DGVM_Tundra_Northern_Taiga: DGVM-based probability for the Forest tundra and Northern Taiga climatic zone
- DGVM_Middle_Taiga: DGVM-based probability for the Middle taiga climatic zone
- DGVM_Southern_Taiga: DGVM-based probability for the Southern taiga climatic zone
- DGVM_Temperate: DGVM-based probability for the Temperate climatic zone
- DGVM_Steppe: DGVM-based probability for the Steppe climatic zone
- DGVM_Deserts: DGVM-based probability for the Semi-deserts and deserts climatic zone
- DGVM_Total: DGVM-based probability taken over all climatic zones

References

Kryazhimskiy, A., Rovenskaya, E., Shvidenko, A., Gusti, M. Shchepashchenko, D. & Veshchinskaya, V. (2015). Towards harmonizing competing models: Russian forests' net primary production case study. Technological Forecasting & Social Change, 98: 245-254.

`forest_npp90` *NPP distribution tables*

Description

Probability distribution tables for aggregated net primary production (NPP) of the forest ecosystems in seven bioclimatic zones in Russia. A dataset contains distributions estimated with the landscape-ecosystem approach (LEA).

Usage

```
forest_npp90
```

Format

A data frame with 15 rows and 9 variables

- npp: NPP values aggregated into size classes of 90 gC/m^2 per year
- LEA_Tundra: LEA-based probability for the Tundra climatic zone
- LEA_Tundra_Northern_Taiga: LEA-based probability for the Forest tundra and Northern Taiga climatic zone
- LEA_Middle_Taiga: LEA-based probability for the Middle taiga climatic zone
- LEA_Southern_Taiga: LEA-based probability for the Southern taiga climatic zone
- LEA_Temperate: LEA-based probability for the Temperate climatic zone
- LEA_Steppe: LEA-based probability for the Steppe climatic zone
- LEA_Deserts: LEA-based probability for the Semi-deserts and deserts climatic zone
- LEA_Total: LEA-based probability taken over all climatic zones

References

Kryazhimskiy, A., Rovenskaya, E., Shvidenko, A., Gusti, M. Shchepashchenko, D. & Veshchinskaya, V. (2015). Towards harmonizing competing models: Russian forests' net primary production case study. Technological Forecasting & Social Change, 98: 245-254.

`integrate` *Integration of distributions*

Description

Combines distributions into an aggregated discrete distribution using several methods.

Usage

```
integrate(vals, pdfs, cdfs)
```

Arguments

`vals`	a discrete ordered range of the random variables
`pdfs`	probability distributions in the 'table-based' format
`cdfs`	probability distributions in the 'cdf-based' format

Details

Integration is done for the discrete prior stochastic estimates. Every prior has the same range specified in *vals*. Prior distribution can be specified in the 'table-based' format or in the 'cdf-based' format. A 'table-based' distribution gives a probability to every outcome in the range. A 'cdf-based' distribution assigns a probability for every bin, whose middle point is an outcome in the range and bin's borders are defined between the range points.

Value

An object of the distribution_table class.

Note

Source distributions are integrated using the posterior integration method [Kryazhimskiy, 2013] and using simple averaging of all input distributions.

References

Kryazhimskiy, A.V. (2013). Posterior integration of independent stochastic estimates. IIASA Interim Report. IR-13-006.

See Also

product, average, summary.distribution_table

Examples

```
# Integration of the two 'table-based' distributions
range <- c(1, 2, 3)
result <- integrate(range, list("m1" = c(0.3, 0.6, 0.1), "m2" = c(0.2, 0.4, 0.4)))
summary(result)
print(result)

# Integration of the 'table-based' and 'cdf-based' distributions
integrate(
   forest_npp90[, 1],
   as.list(forest_npp90["LEA_Tundra"]),
   list("DGVM_Tundra" = function(x)(pnorm(x, mean = 202, sd = 52))))

# Integration of the 'cdf-based' distributions
integrate(
   c(1, 2),
   cdfs = list(function(x)(punif(x, min = 0.5, max = 2.5)),
               function(x)(punif(x, min = 0.5, max = 2.5))))
```

---

is.distribution_table *Distribution tables*

Description

Test of an object being interpretable as a distribution table.

Usage

```
is.distribution_table(obj)
```

Arguments

obj	an object to be tested

Value

The method returns TRUE if its argument inherits from the class "distribution_table", and FALSE otherwise.

`modelIntegration` *Overview of the modelIntegration package*

Description

Combination of prior distributions into an aggregated (synthetic) one. The package implements the posterior integration method [Kryazhimskiy, 2013]. For comparison, an implementation of simple averaging of the input distributions is added.

Note

A posteriori integration is understood as an improvement of data given by a priori probabilities. The approach is based on the concept of an a posteriori event in the product of probability spaces associated with a priori probabilities. The conditional probability on the product space that is specified by an a posteriori event (which reflects the fact that all the prior stochastic estimates represent the same deterministic element) determines in a natural way the probability on the set of initial elementary events; the latter is recognized as the result of a posteriori integration of a priori models [Kryazhimskiy, 2016].
E.g., the probability of the true value to be equal z is computed as *p1[z]*p2[z]*...*pn[z]/P(E*)*, where *p1...pn* are probabilities from x and E^* denotes a posteriori event.

Author(s)

Anna Shchiptsova, IIASA

References

Kryazhimskiy, A.V. (2013). Posterior integration of independent stochastic estimates. IIASA Interim Report. IR-13-006.
Kryazhimskiy, A.V. (2016). Posteriori integration of probabilities. Elementary theory. Theory of Probability and its Applications, 60(1): 62-87.

See Also

`integrate`

Examples

```
library(modelIntegration)

pdf1 <- c(0.75, 0.25)
pdf2 <- c(0.75, 0.25)
bins <- c(1, 2)

result <- integrate(bins, list(pdf1, pdf2))

summary(result)
product(result)
```

print.distribution_table *Print method*

Description

Prints an object of the distribution_table class.

Usage

```
## S3 method for class 'distribution_table'
print(x, ...)
```

Arguments

x	an R object of the distribution_table class
...	other arguments passed to methods

Details

The print format matches a format for the data.frame class. The output contains discrete probability distributions of the sources and their product. The column *x* returns a random variable range. The *Product* column has probabilities from the product distribution.

See Also

product

Examples

```
# S3 method for class "distribution_table"
pdf1 <- c(0.75, 0.25)
pdf2 <- c(0.75, 0.25)
bins <- c(1, 2)
print(integrate(bins, list(pdf1, pdf2)))
```

product *Product distribution*

Description

Combines source distributions into the resultant integrated distribution using the posterior integration method [Kryazhimskiy, 2013]. The latter one is called a *product* probability distribution.

Usage

```
product(x, ...)
```

Arguments

x	an R object with source distributions
...	other arguments passed to methods

Details

A posteriori integration is understood as an improvement of data given by a priori probabilities in *x*. The approach is based on the concept of an a posteriori event in the product of probability spaces associated with a priori probabilities. The conditional probability on the product space that is specified by an a posteriori event (which reflects the fact that all the prior stochastic estimates represent the same deterministic element) determines in a natural way the probability on the set of initial elementary events; the latter is recognized as the result of a posteriori integration of a priori models [Kryazhimskiy, 2016].
E.g, the probability of the true value to be equal *z* is computed as *p1[z]*p2[z]*...*pn[z]/P(E*)*, where *p1...pn* are probabilities from *x* and *E** denotes a posteriori event.

Value

A data frame with the 'table-based' distribution, where every row contains an outcome in the random variable range and its probability in the integrated distribution.

References

Kryazhimskiy, A.V. (2013). Posterior integration of independent stochastic estimates. IIASA Interim Report. IR-13-006.
Kryazhimskiy, A.V. (2016). Posteriori integration of probabilities. Elementary theory. Theory of Probability and its Applications, 60(1): 62-87.

Examples

```
# Product distribution of two 'table-based' distributions
# S3 method for class "distribution_table"
pdf1 <- c(0.75, 0.25)
pdf2 <- c(0.75, 0.25)
bins <- c(1, 2)
product(integrate(bins, list(pdf1, pdf2)))
```

---

statistics *Descriptive statistics*

Description

Provides descriptive statistics of the source and integrated distributions.

Usage

```
statistics(x, ...)
```

Arguments

x	an R object with source distributions
...	other arguments passed to methods

Details

Descriptive statistics include distribution mean and standard deviation. Source distributions are combined into the product and average integrated distributions.

Value

A data frame with means and standard deviations. The the last two columns "Product" and "Average" contain estimates for the corresponding integrated distributions.

See Also

product, average

Examples

```
# S3 method for class "distribution_table"
pdf1 <- c(0.75, 0.25)
pdf2 <- c(0.75, 0.25)
bins <- c(1, 2)
statistics(integrate(bins, list(pdf1, pdf2)))
```

---

summary.distribution_table

Summary of integrated distributions

Description

Provides summary statistics of integrated combinations of the source distributions.

Usage

```
## S3 method for class 'distribution_table'
summary(object, ...)
```

Arguments

object	an R object of the distribution_table class
...	other arguments passed to methods

Details

Distributions are summarized by their mean and standard deviation. Source distributions are combined into the product and average integrated distributions.

Value

A data frame with means and standard deviations. The columns "Product" and "Average" contain estimates for the corresponding integrated distributions.

See Also

statistics, product, average

Examples

```
# S3 method for class "distribution_table"
pdf1 <- c(0.75, 0.25)
pdf2 <- c(0.75, 0.25)
bins <- c(1, 2)
summary(integrate(bins, list(pdf1, pdf2)))
```

Index