Package 'fipp'

October 13, 2022

Title Induced Priors in Bayesian Mixture Models

Version 1.0.0

Description Computes implicitly induced quantities from prior/hyperparameter
specifications of three Mixtures of Finite Mixtures models: Dirichlet
Process Mixtures (DPMs; Escobar and West (1995)
<doi:10.1080/01621459.1995.10476550>), Static Mixtures of Finite Mixtures
(Static MFMs; Miller and Harrison (2018)
<doi:10.1080/01621459.2016.1255636>), and Dynamic Mixtures of Finite
Mixtures (Dynamic MFMs; Frühwirth-Schnatter, Malsiner-Walli and Grün (2020)
<arXiv:2005.09918>). For methodological details, please refer to
Greve, Grün, Malsiner-Walli and Frühwirth-Schnatter (2020)
<arXiv:2012.12337>) as well as the package vignette.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Depends R (>= 3.5.0)

LinkingTo Rcpp, RcppArmadillo

Imports Rcpp, stats, matrixStats

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation yes

Author Jan Greve [aut, cre],
Bettina Grün [ctb] (<https://orcid.org/0000-0001-7265-4773>),
Gertraud Malsiner-Walli [ctb] (<https://orcid.org/0000-0002-1213-4749>),
Sylvia Frühwirth-Schnatter [ctb]
(<https://orcid.org/0000-0003-0516-5552>)

Maintainer Jan Greve <jan.greve@wu.ac.at>

Repository CRAN

Date/Publication 2021-02-11 10:30:02 UTC

R topics documented:

---

dbnb   *Probability density function of the BNB distribution*

Description

Evaluates the probability density function of the beta-negative-binomial (BNB) distribution with a mean parameter and two shape parameters.

Usage

```
dbnb(x, mu, a, b, log = FALSE)
```

Arguments

x	vector of quantiles.
mu	mean parameter.
a	1st shape parameter.
b	2nd shape parameter.
log	logical; if TRUE, density values p are given as log(p).

Details

The BNB distribution has density

$$f(x) = \frac{\Gamma(\mu + x)B(\mu + a, x + b)}{\Gamma(\mu)\Gamma(x + 1)B(a, b)},$$

where μ is the mean parameter and a and b are the first and second shape parameter.

Value

Numeric vector of density values.

References

Frühwirth-Schnatter, S., Malsiner-Walli, G., and Grün, B. (2020) Generalized mixtures of finite mixtures and telescoping sampling https://arxiv.org/abs/2005.09918

Examples

```
## Similar to other d+DISTRIBUTION_NAME functions such as dnorm, it
## evaluates the density of a distribution (in this case the BNB distri)
## at point x
##
## Let's try with the density of x = 1 for BNB(1,4,3)
x <- 1
dbnb(x, mu = 1, a = 4, b = 3)

## The primary use of this function is in the closures returned from
## fipp() or nCluststers() as a prior on K-1
pmf <- nClusters(Kplus = 1:15, N = 100, type = "static",
gamma = 1, maxK = 150)

## Now evaluate above when K-1 ~ BNB(1,4,3)
pmf(priorK = dbnb, priorKparams = list(mu = 1, a = 4, b = 3))

## Compare the result with the case when K-1 ~ Pois(1)
pmf(priorK = dpois, priorKparams = list(lambda = 1))

## Although both BNB(1,4,3) and Pois(1) have 1 as their mean, the former
## has a fatter rhs tail. We see that it is reflected in the induced prior
## on K+ as well
```

---

fipp	*Moments of symmetric additive functional computed over the induced prior partitions (static/dynamic MFMs and DPM)*

Description

`fipp` is a closure which returns a function that computes moments of a user-specified functional over the induced prior partitions. Required arguments are: prior distribution of the number of mixture components and its parameters (see examples for details). Optional arguments are: the number of moments to be evaluated (currently only up to 2 are implemented) and whether the mean/variance or 1st/2nd moments should be printed out as a result of computing the first two moments (default is set to print out mean/variance).

Usage

```
fipp(
  lfunc,
  Kplus,
  N,
  type = c("DPM", "static", "dynamic"),
  alpha = NULL,
  gamma = NULL,
  maxK = NULL,
  log = FALSE
)
```

Arguments

`lfunc`	a logged version of the additive symmetric functional intended to compute over the prior partition. The function should only accept one argument N_j (= number of observations in each partition).
`Kplus`	a numeric value that represents the number of filled clusters in data
`N`	the number of observation in data
`type`	the type of model considered. Three models (static/dynamic MFMs and DPM) are supported.
`alpha, gamma`	hyperparameters for the Dirichlet prior. For static MFM, gamma should be specified, while alpha should be specified for all other models (that is, for dynamic MFM and DPM).
`maxK`	the maximum number of K (= the number of mixture components) considered. Only needed for static/dynamic MFMs.
`log`	logical, indicating whether the probability should be logged or not

Value

`fipp` returns a function which takes two required arguments (required only for static/dynamic MFMs) and 2 optional arguments:

priorK a function with support on the positive integers. The function serves as a prior of K (default = NULL which is for DPM).

priorKparams a named list of prior parameters for the function supplied in argument `priorK` (default = NULL which is for DPM).

order maximum number of moments to be evaluated by the function (default = 2)

replace2ndwvar replace 2nd moment with variance (default = TRUE)

References

Greve, J., Grün, B., Malsiner-Walli, G., and Frühwirth-Schnatter, S. (2020) Spying on the Prior of the Number of Data Clusters and the Partition Distribution in Bayesian Cluster Analysis. https://arxiv.org/abs/2012.12337

Escobar, M. D., and West, M. (1995) Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* **90** (430), Taylor & Francis: 577-–88. https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476550

Miller, J. W., and Harrison, M. T. (2018) Mixture Models with a Prior on the Number of Components. *Journal of the American Statistical Association* **113** (521), Taylor & Francis: 340-–56. https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1255636

Frühwirth-Schnatter, S., Malsiner-Walli, G., and Grün, B. (2020) Generalized mixtures of finite mixtures and telescoping sampling https://arxiv.org/abs/2005.09918

Examples

```
## Determine mean/variance of the number of singleton clusters for dynamic
## MFM model conditional on K+ = 5, alpha = 1 with a sample size N = 100.
## We assume that K will be smaller than 30 by setting maxK = 30, please
## increase this value for more realistic analysis.
##
## First create the function singletons():
singletons <- fipp(lfunc = function(n) log(n==1), Kplus = 5, N = 100,
  type = "dynamic", alpha = 1, maxK = 30)

## Then evaluate it using a Geom(0.1) prior:
singletons(dgeom, list(prob = 0.1))

## Try a different prior, the Poisson prior Pois(1):
singletons(dpois, list(lambda = 1))

## If mean is the only thing you are interested in, try the following:
singletons(dpois, list(lambda = 1), order = 1)

## Also, if you want 1st/2nd moments instead of mean/variance, try:
singletons(dpois, list(lambda = 1), replace2ndwvar = FALSE)
```

---

nClusters	*Prior pmf of the number of data clusters for three model types (static/dynamic MFMs and DPM)*

Description

nClusters is a closure that returns a function which computes a table of probability masses for specified K+s. Arguments needed for the returned function to evaluate are: prior distribution of the number of mixture components and its parameters (see examples for details).

Usage

```
nClusters(
  Kplus,
  N,
  type = c("DPM", "static", "dynamic"),
  alpha = NULL,
  gamma = NULL,
  maxK = NULL,
  log = FALSE
)
```

Arguments

Kplus	a numeric value or vector. All values must be positive integers (that is 1,2,...). It specifies the range of the number of data clusters the user wants to evaluate the prior probabilities on.
N	the number of observations in data
type	the type of model considered. Three models (static/dynamic MFMs and DPM) are supported.
alpha, gamma	hyperparameters for the symmetric Dirichlet prior. For static MFM, gamma should be specified, while alpha should be specified for all other models (that is, dynamic MFM and DPM).
maxK	the maximum number of K (= the number of mixture components) considered. Only needed for static/dynamic MFMs.
log	logical, indicating whether the returned probability should be logged or not

Value

nClusters returns a function which takes two arguments:

priorK a function with support on the positive integers. The function serves as a prior on K (default = NULL which is for the DPM).

priorKparams a named list of prior parameters for the function supplied in argument priorK (default = NULL which is for the DPM).

References

Greve, J., Grün, B., Malsiner-Walli, G., and Frühwirth-Schnatter, S. (2020) Spying on the Prior of the Number of Data Clusters and the Partition Distribution in Bayesian Cluster Analysis. https://arxiv.org/abs/2012.12337

Escobar, M. D., and West, M. (1995) Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* **90** (430), Taylor & Francis: 577-–88. https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476550

Miller, J. W., and Harrison, M. T. (2018) Mixture Models with a Prior on the Number of Components. *Journal of the American Statistical Association* **113** (521), Taylor & Francis: 340-–56. https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1255636

Frühwirth-Schnatter, S., Malsiner-Walli, G., and Grün, B. (2020) Generalized mixtures of finite mixtures and telescoping sampling https://arxiv.org/abs/2005.09918

Examples

```
## first, create the function pmf() for the dynamic MFM
## with N = 100, K+ evaluated between 1 and 15 with alpha = 1,
## we assume that K will be smaller than 30 by setting maxK  = 30,
## please increase this value for more realistic analysis.
pmf <- nClusters(Kplus = 1:15, N = 100, type = "dynamic",
alpha = 1, maxK = 30)

## then, specifiy the prior for K so that the pmf can be evaluated
```

```
## between K+ = 1 and K+ = 15
pmf(dgeom, list(prob = 0.1))

## we can also compare this result with a different prior setting
pmf(dpois, list(lambda = 1))
```

Index