Package 'AROC'

October 12, 2022

Type Package

Title Covariate-Adjusted Receiver Operating Characteristic Curve Inference

Version 1.0-4

Date 2022-02-18

Imports stats, grDevices, graphics, splines, np, Matrix, Hmisc, MASS, moments, nor1mix, spatstat.geom

Description
Estimates the covariate-adjusted Receiver Operating Characteristic (AROC) curve and pooled (unadjusted) ROC curve by different methods. Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018) <arXiv:1806.00473>. NOTE: We have created a new package, 'ROCnReg', with more functionalities. It also implements all the methods included in 'AROC'. We, therefore, recommend using 'ROCnReg' ('AROC' will no longer be maintained).

License GPL

NeedsCompilation no

Author Maria Xose Rodriguez-Alvarez [aut, cre] (<https://orcid.org/0000-0002-1329-9238>), Vanda Inacio [aut] (<https://orcid.org/0000-0001-8084-1616>)

Maintainer Maria Xose Rodriguez-Alvarez <mxrodriguez@uvigo.es>

Repository CRAN

Date/Publication 2022-02-21 13:30:22 UTC

R topics documented:

`AROC-package` *Covariate-Adjusted Receiver Operating Characteristic Curve Inference*

Description

Estimates the covariate-adjusted Receiver Operating Characteristic (AROC) curve and pooled (unadjusted) ROC curve by different methods.

Details

Package:	AROC
Type:	Package
Version:	1.0-4
Date:	2022-02-18
License:	GPL

Author(s)

Maria Xose Rodriguez-Alvarez and Vanda Inacio.

Maintainer: Maria Xose Rodriguez-Alvarez <mxrodriguez@uvigo.es>

References

Janes, H., and Pepe, M.S. (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. Biometrika, 96(2), 371 - 382.

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

Rodriguez-Alvarez, M. X., Roca-Pardinas, J., and Cadarso-Suarez, C. (2011). ROC curve and covariates: extending induced methodology to the non-parametric framework. Statistics and Computing, 21(4), 483 - 499.

`AROC.bnp` *Nonparametric Bayesian inference of the covariate-adjusted ROC curve (AROC).*

Description

Estimates the covariate-adjusted ROC curve (AROC) using the nonparametric Bayesian approach proposed by Inacio de Carvalho and Rodriguez-Alvarez (2018).

Usage

```
AROC.bnp(formula.healthy, group, tag.healthy, data, scale = TRUE,
  p = seq(0, 1, l = 101), paauc = paauccontrol(),
  compute.lpml = FALSE, compute.WAIC = FALSE,
  m0, S0, nu, Psi, alpha = 1, a = 2, b = 0.5, L = 10, nsim = 10000, nburn = 2000)
```

Arguments

`formula.healthy`	A `formula` object specifying the B-splines dependent Dirichlet process mixture model for the estimation of the conditional distribution function for the diagnostic test outcome in the healthy population (see Note).
`group`	A character string with the name of the variable that distinguishes healthy from diseased individuals.
`tag.healthy`	The value codifying the healthy individuals in the variable `group`.
`data`	Data frame representing the data and containing all needed variables.
`scale`	A logical value. If TRUE the test outcomes are scaled, i.e., are divided by the standard deviation. The default is TRUE.
`p`	Set of false positive fractions (FPF) at which to estimate the covariate-adjusted ROC curve.
`paauc`	A list of control values to replace the default values returned by the function `paauccontrol`. This argument is used to indicate whether the partial area under the covariate-adjusted ROC curve (pAAUC) should be computed and at which FPF.
`compute.lpml`	A logical value. If TRUE, the log pseudo marginal likelihood (LPML, Geisser and Eddy, 1979) and the conditional predictive ordinates (CPO) are computed.
`compute.WAIC`	A logical value. If TRUE, the widely applicable information criterion (WAIC, Gelman et al., 2014; Watanabe, 2010) is computed.
`m0`	A numeric vector. Hyperparameter; mean vector of the (multivariate) normal prior distribution for the mean of the normal component of the centering distribution. If missing, it is set to a vector of zeros of length `Q` (see Details).

S0	A numeric matrix. Hyperparameter; covariance matrix of the (multivariate) normal prior distribution for the mean of the normal component of the centering distribution. If missing, it is set to a diagonal matrix of dimension QxQ with 100 in the diagonal (see Details).
nu	A numeric value. Hyperparameter; degrees of freedom of the Wishart prior distribution for the precision matrix of the the normal component of the centering distribution. If missing, it is set to Q + 2 (see Details)
Psi	A numeric matrix. Hyperparameter; scale matrix of the Wishart distribution for the precision matrix of the the normal component of the centering distribution. If missing, it is set to an identity matrix of dimension QxQ (see Details).
alpha	A numeric value. Precision parameter of the Dirichlet Process. The default is 1 (see Details).
a	A numeric value. Hyperparameter; shape parameter of the gamma prior distribution for the precision (inverse variance). The default is 2 (scaled data) (see Details).
b	A numeric value. Hyperparameter; rate parameter of the gamma prior distribution for the precision (inverse variance). The default is 0.5 (scaled data) (see Details).
L	A numeric value. Maximum number of mixture components for the B-splines dependent Dirichlet process mixture model. The default is 10 (see Details)
nsim	A numeric value. Total number of Gibbs sampler iterates (including the burn-in). The default is 10000.
nburn	A numeric value. Number of burn-in iterations. The default is 2000.

Details

Estimates the covariate-adjusted ROC curve (AROC) defined as

$$AROC\,(t) = Pr\{1 - F_{\bar{D}}(Y_D|\mathbf{X}_D) \leq t\},$$

where $F_{\bar{D}}(\cdot|\mathbf{X}_{\bar{D}})$ denotes the conditional distribution function for $Y_{\bar{D}}$ conditional on the vector of covariates $\mathbf{X}_{\bar{D}}$. In particular, the method implemented in this function combines a B-splines dependent Dirichlet process mixture model to estimate $F_{\bar{D}}(\cdot|\mathbf{X}_{\bar{D}})$ and the Bayesian bootstrap (Rubin, 1981) to estimate the outside probability. More precisely, and letting $\{(\mathbf{x}_{\bar{D}i}, y_{\bar{D}i})\}_{i=1}^{n_{\bar{D}}}$ be a random sample from the nondiseased population

$$F_{\bar{D}}(y_{\bar{D}i}|\mathbf{X}_{\bar{D}} = \mathbf{x}_{\bar{D}i}) = \sum_{l=1}^{L} \omega_l \Phi(y_{\bar{D}i} \mid \mu_l(\mathbf{x}_{\bar{D}i}), \sigma_l^2),$$

where $\mu_l(\mathbf{x}_{\bar{D}i}) = \mathbf{z}_{\bar{D}i}^T \beta_l$ and L is pre-specified (maximum number of mixture components). The ω_l's result from a truncated version of the stick-breaking construction ($\omega_1 = v_1$; $\omega_l = v_l \prod_{r<l}(1 - v_r)$, $l = 2, \ldots, L$; $v_1, \ldots, v_{L-1} \sim$ Beta $(1, \alpha)$; $v_L = 1$), $\beta_l \sim N_Q(\mathbf{m}, \mathbf{S})$, and $\sigma_l^{-2} \sim \Gamma(a, b)$. It is assumed that $\mathbf{m} \sim N_Q(\mathbf{m}_0, \mathbf{S}_0)$ and $\mathbf{S}^{-1} \sim W(\nu, (\nu\Psi)^{-1})$. Here $W(\nu, (\nu\Psi)^{-1})$ denotes a Wishart distribution with ν degrees of freedom and expectation Ψ^{-1}, and Q denotes the dimension of vector $\mathbf{z}_{\bar{D}i}$. For a detailed description, we refer to Inacio de Carvalho and Rodriguez-Alvarez (2018).

Value

As a result, the function provides a list with the following components:

`call`	The matched call.
`p`	Set of false positive fractions (FPF) at which the pooled ROC curve has been estimated.
`ROC`	Estimated covariate-adjusted ROC curve (AROC) (posterior mean), and 95% pointwise posterior credible band.
`AUC`	Estimated area under the covariate-adjusted ROC curve (AAUC) (posterior mean), and 95% pointwise posterior credible band.
`pAUC`	If required, estimated partial area under the covariate-adjusted ROC curve (pAAUC) (posterior mean), and 95% pointwise posterior credible band.
`lpml`	If required, list with two components: the log pseudo marginal likelihood (LPML) and the conditional predictive ordinates (CPO).
`WAIC`	If required, widely applicable information criterion (WAIC).
`fit`	Results of the fitting process. It is a list with the following components: (1) `mm`: information needed to construct the model matrix associated with the B-splines dependent Dirichlet process mixture model. (2) `beta`: array of dimension `NxLxQ` with the sampled regression coefficients. Here, `N` is the number of Gibbs sampler iterates after burn-in, `L` is the maximum number of mixture components, and `Q` is the dimension of vector $\mathbf{Z}_{\bar{D}}$ (see also Details). (3) `sd`: matrix of dimension `NxL` with the sampled variances. Here, `N` is the number of Gibbs sampler iterates after burn-in, and `L` is the maximum number of mixture components (see also Details). (4) `probs`: matrix of dimension `NxL` with the sampled components' weights. Here, `N` is the number of Gibbs sampler iterates after burn-in and `L` is the maximum number of mixture components (see also Details).
`data_model`	List with the data used in the fit: observed diagnostic test outcome and B-spline design matrices, separately for the healthy and diseased groups.

Note

The input argument `formula.healthy` is similar to that used for the `glm` function, except that flexible specifications can be added by means of function `f()`. For instance, specification $y \sim x1 + f(x2, K = 3)$ would assume a linear effect of `x1` and the effect of `x2` would be modeled using B-splines basis functions. The argument `K = 3` indicates that `3` internal knots will be used, with the quantiles of `x2` used for their location. Categorical variables (factors) can be also incorporated, as well as factor-by-curve interaction terms. For example, to include the interaction between `age` and `gender` we need to specify $y \sim gender + f(age, by = gender, K = 3)$.

References

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

Rubin, D. B. (1981). The Bayesian bootstrap. The Annals of Statistics, 9(1), 130-134.

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

summary(m0)

plot(m0)
```

AROC.bsp *Semiparametric Bayesian inference of the covariate-adjusted ROC curve (AROC).*

Description

Estimates the covariate-adjusted ROC curve (AROC) using the semiparametric Bayesian normal linear regression model discussed in Inacio de Carvalho and Rodriguez-Alvarez (2018).

Usage

```
AROC.bsp(formula.healthy, group, tag.healthy, data, scale = TRUE,
  p = seq(0, 1, l = 101), paauc = paauccontrol(),
  compute.lpml = FALSE, compute.WAIC = FALSE,
  m0, S0, nu, Psi, a = 2, b = 0.5, nsim = 5000, nburn = 1500)
```

Arguments

formula.healthy	A formula object specifying the Bayesian normal linear regression model for the estimation of the conditional distribution function for the diagnostic test outcome in the healthy population (see Details).
group	A character string with the name of the variable that distinguishes healthy from diseased individuals.

tag.healthy	The value codifying the healthy individuals in the variable group.
data	Data frame representing the data and containing all needed variables.
scale	A logical value. If TRUE the test outcomes are scaled, i.e., are divided by the standard deviation. The default is TRUE.
p	Set of false positive fractions (FPF) at which to estimate the covariate-adjusted ROC curve.
compute.lpml	A logical value. If TRUE, the log pseudo marginal likelihood (LPML, Geisser and Eddy, 1979) and the conditional predictive ordinates (CPO) are computed.
paauc	A list of control values to replace the default values returned by the function paauccontrol. This argument is used to indicate whether the partial area under the covariate-adjusted ROC curve (pAAUC) should be computed and at which FPF.
compute.WAIC	A logical value. If TRUE, the widely applicable information criterion (WAIC, Gelman et al., 2014; Watanabe, 2010) is computed.
m0	A numeric vector. Hyperparameter; mean vector of the (multivariate) normal distribution for the mean of the regression coefficients. If missing, it is set to a vector of zeros of length p+1 (see Details).
S0	A numeric matrix. Hyperprior. If missing, it is set to a diagonal matrix of dimension (p+1)x(p+1) with 100 in the diagonal (see Details).
nu	A numeric value. Hyperparameter; degrees of freedom of the Wishart distribution for the precision matrix of the regression coefficients. If missing, it is set to p + 3 (see Details)
Psi	A numeric matrix. Hyperparameter; scale matrix of the Wishart distribution for the precision matrix of the regression coefficients. If missing, it is set to an identity matrix of dimension (p+1)x(p+1) (see Details).
a	A numeric value. Hyperparameter; shape parameter of the gamma distribution for the precision (inverse variance). The default is 2 (scaled data) (see Details).
b	A numeric value. Hyperparameter; rate parameter of the gamma distribution for the precision (inverse variance). The default is 0.5 (scaled data) (see Details).
nsim	A numeric value. Total number of Gibbs sampler iterates (including the burn-in). The default is 5000.
nburn	A numeric value. Number of burn-in iterations. The default is 1500.

Details

Estimates the covariate-adjusted ROC curve (AROC) defined as

$$AROC\,(t) = Pr\{1 - F_{\bar{D}}(Y_D|\mathbf{X}_D) \leq t\},$$

where $F_{\bar{D}}(\cdot|\mathbf{X}_{\bar{D}})$ denotes the conditional distribution function for $Y_{\bar{D}}$ conditional on the vector of covariates $X_{\bar{D}}$. In particular, the method implemented in this function combines a Bayesian normal linear regression model to estimate $F_{\bar{D}}(\cdot|\mathbf{X}_{\bar{D}})$ and the Bayesian bootstrap (Rubin, 1981) to estimate the outside probability. More precisely, and letting $\{(\mathbf{x}_{\bar{D}i}, y_{\bar{D}i})\}_{i=1}^{n_{\bar{D}}}$ be a random sample from the nondiseased population

$$F_{\bar{D}}(y_{\bar{D}i}|\mathbf{X}_{\bar{D}} = \mathbf{x}_{\bar{D}i}) = \Phi(y_{\bar{D}i} \mid \mathbf{x}_{\bar{D}i}^{*T}\beta^*, \sigma^2),$$

where $\mathbf{x}_{\bar{D}i}^{*T} = (1, \mathbf{x}_{\bar{D}i}^T)$, $\beta^* \sim N_{p+1}(\mathbf{m}, \mathbf{S})$ and $\sigma^{-2} \sim \Gamma(a, b)$. It is assumed that $\mathbf{m} \sim N_{p+1}(\mathbf{m}_0, \mathbf{S}_0)$ and $\mathbf{S}^{-1} \sim W(\nu, (\nu\Psi)^{-1})$, where $p+1$ denotes the number of columns of the design matrix $\mathbf{X}_{\bar{D}}^*$. Here $W(\nu, (\nu\Psi)^{-1})$ denotes a Wishart distribution with ν degrees of freedom and expectation Ψ^{-1}. For a detailed description, we refer to Inacio de Carvalho and Rodriguez-Alvarez (2018).

Value

As a result, the function provides a list with the following components:

`call`	The matched call.
`p`	Set of false positive fractions (FPF) at which the pooled ROC curve has been estimated.
`ROC`	Estimated covariate-adjusted ROC curve (AROC) (posterior mean), and 95% pointwise posterior credible band.
`AUC`	Estimated area under the covariate-adjusted ROC curve (AAUC) (posterior mean), and 95% pointwise posterior credible band.
`pAUC`	If required in the call to the function, estimated partial area under the covariate-adjusted ROC curve (pAAUC) (posterior mean), and 95% pointwise posterior credible band.
`lpml`	If required, list with two components: the log pseudo marginal likelihood (LPML) and the conditional predictive ordinates (CPO).
`WAIC`	If required, widely applicable information criterion (WAIC).
`fit`	Results of the fitting process. It is a list with the following components: (1) `mm`: information needed to construct the model matrix associated with the B-splines dependent Dirichlet process mixture model. (2) `beta`: matrix of dimension `Nxp+1` with the sampled regression coefficients. Here, `N` is the number of Gibbs sampler iterates after burn-in, and `p+1` the number of columns of the design matrix (see also Details). (3) `sd`: vector of length `N` with the sampled variances (see also Details).
`data_model`	List with the data used in the fit: observed diagnostic test outcome and B-spline design matrices, separately for the healthy and diseased groups.

References

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

Rubin, D. B. (1981). The Bayesian bootstrap. The Annals of Statistics, 9(1), 130-134.

See Also

`AROC.bnp`, `AROC.bsp`, `AROC.sp`, `AROC.kernel`, `pooledROC.BB` or `pooledROC.emp`.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m1 <- AROC.bsp(formula.healthy = l_marker1 ~ age,
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, nsim = 5000, nburn = 1500)

summary(m1)

plot(m1)
```

`AROC.kernel` *Non parametric kernel-based estimation of the covariate-adjusted ROC curve (AROC).*

Description

Estimates the covariate-adjusted ROC curve (AROC) using the nonparametric kernel-based method proposed by Rodriguez-Alvarez et al. (2011). The method, as it stands now, can only deal with one continuous covariate.

Usage

```
AROC.kernel(marker, covariate, group, tag.healthy, data, p = seq(0, 1, l = 101), B = 1000)
```

Arguments

`marker`	A character string with the name of the diagnostic test variable.
`covariate`	A character string with the name of the continuous covariate.
`group`	A character string with the name of the variable that distinguishes healthy from diseased individuals.
`tag.healthy`	The value codifying the healthy individuals in the variable `group`.
`data`	Data frame representing the data and containing all needed variables.
`p`	Set of false positive fractions (FPF) at which to estimate the covariate-adjusted ROC curve.
`B`	An integer value specifying the number of bootstrap resamples for the construction of the confidence intervals. By default 1000.

Details

Estimates the covariate-adjusted ROC curve (AROC) defined as

$$AROC\,(t) = Pr\{1 - F_{\bar{D}}(Y_D|X_D) \leq t\},$$

where $F_{\bar{D}}(\cdot|X_D)$ denotes the conditional distribution function for $Y_{\bar{D}}$ conditional on the vector of covariates $X_{\bar{D}}$. In particular, the method implemented in this function estimates the outer probability empirically (see Janes and Pepe, 2008) and $F_{\bar{D}}(\cdot|X_{\bar{D}})$ is estimated assuming a nonparametric location-scale regression model for $Y_{\bar{D}}$, i.e.,

$$Y_{\bar{D}} = \mu_{\bar{D}}(X_{\bar{D}}) + \sigma_{\bar{D}}(X_{\bar{D}})\varepsilon_{\bar{D}},$$

where $\mu_{\bar{D}}$ is the regression funcion, $\sigma_{\bar{D}}$ is the variance function, and $\varepsilon_{\bar{D}}$ has zero mean, variance one, and distribution function $F_{\bar{D}}$. As a consequence, and for a random sample $\{(x_{\bar{D}i}, y_{\bar{D}i})\}_{i=1}^{n_{\bar{D}}}$

$$F_{\bar{D}}(y_{\bar{D}i}|X_{\bar{D}} = x_{\bar{D}i}) = F_{\bar{D}}\left(\frac{y_{\bar{D}i} - \mu_{\bar{D}}(x_{\bar{D}i})}{\sigma_{\bar{D}}(x_{\bar{D}i})}\right).$$

Both the regression and variance functions are estimated using the Nadaraya-Watson estimator, and the bandwidth are selected using least-squares cross-validation. Implementation relies on the R-package np. No assumption is made about the distribution of $\varepsilon_{\bar{D}}$, which is empirically estimated on the basis of standardised residuals.

Value

As a result, the function provides a list with the following components:

`call`	The matched call.
`p`	Set of false positive fractions (FPF) at which the pooled ROC curve has been estimated
`ROC`	Estimated covariate-adjusted ROC curve (AROC), and 95% pointwise confidence intervals (if required)
`AUC`	Estimated area under the covariate-adjusted ROC curve (AAUC), and 95% pointwise confidence intervals (if required).
`bw.mean`	An object of class `npregbw` with the selected bandwidth for the nonparametric regression function. For further details, see R-package np.
`bw.var`	An object of class `npregbw` with the selected bandwidth for the nonparametric variance function. For further details, see R-package np.
`fit.mean`	An object of class `npreg` with the nonparametric regression function estimate. For further details, see R-package np.
`fit.var`	An object of class `npreg` with the nonparametric variance function estimate. For further details, see R-package np.

References

Hayfield, T., and Racine, J. S.(2008). Nonparametric Econometrics: The np Package. Journal of Statistical Software 27(5). URL http://www.jstatsoft.org/v27/i05/.

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

Rodriguez-Alvarez, M. X., Roca-Pardinas, J., and Cadarso-Suarez, C. (2011). ROC curve and covariates: extending induced methodology to the non-parametric framework. Statistics and Computing, 21(4), 483 - 499.

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m2 <- AROC.kernel(marker = "l_marker1", covariate = "age",
group = "status", tag.healthy = 0, data = newpsa,
p = seq(0,1,l=101), B = 500)

summary(m2)

plot(m2)
```

AROC.sp *Semiparametric frequentist inference of the covariate-adjusted ROC curve (AROC).*

Description

Estimates the covariate-adjusted ROC curve (AROC) using the semiparametric approach proposed by Janes and Pepe (2009).

Usage

```
AROC.sp(formula.healthy, group, tag.healthy, data,
est.surv.h = c("normal", "empirical"), p = seq(0, 1, l = 101), B = 1000)
```

Arguments

formula.healthy	A formula object specifying the location regression model to be fitted in healthy population (see Details).
group	A character string with the name of the variable that distinguishes healthy from diseased individuals.
tag.healthy	The value codifying the healthy individuals in the variable group.
data	Data frame representing the data and containing all needed variables.
est.surv.h	A character string. It indicates how the conditional distribution function of the diagnostic test in healthy population is estimated. Options are "normal" and "empirical" (see Details). The default is "normal".
p	Set of false positive fractions (FPF) at which to estimate the covariate-adjusted ROC curve.
B	An integer value specifying the number of bootstrap resamples for the construction of the confidence intervals. By default 1000.

Details

Estimates the covariate-adjusted ROC curve (AROC) defined as

$$AROC\,(t) = Pr\{1 - F_{\bar{D}}(Y_D|\mathbf{X}_D) \leq t\},$$

where $F_{\bar{D}}(\cdot|\mathbf{X}_{\bar{D}})$ denotes the conditional distribution function for $Y_{\bar{D}}$ conditional on the vector of covariates $\mathbf{X}_{\bar{D}}$. In particular, the method implemented in this function estimates the outer probability empirically (see Janes and Pepe, 2008) and $F_{\bar{D}}(\cdot|\mathbf{X}_{\bar{D}})$ is estimated assuming a semiparametric location regression model for $Y_{\bar{D}}$, i.e.,

$$Y_{\bar{D}} = \mathbf{X}_{\bar{D}}^T \beta_{\bar{D}} + \sigma_{\bar{D}} \varepsilon_{\bar{D}},$$

such that, for a random sample $\{(\mathbf{x}_{\bar{D}i})\}_{i=1}^{n_{\bar{D}}}$ from the healthy population, we have

$$F_{\bar{D}}(y|\mathbf{X}_{\bar{D}} = \mathbf{x}_{\bar{D}i}) = F_{\bar{D}}\left(\frac{y - \mathbf{x}_{\bar{D}i}^T \beta_{\bar{D}}}{\sigma_{\bar{D}}}\right),$$

where $F_{\bar{D}}$ is the distribution function of $\varepsilon_{\bar{D}}$. In line with the assumptions made about the distribution of $\varepsilon_{\bar{D}}$, estimators will be referred to as: (a) "normal", where Gaussian error is assumed, i.e., $F_{\bar{D}}(y) = \Phi(y)$; and, (b) "empirical", where no assumption is made about the distribution (in this case, the distribution function $F_{\bar{D}}$ is empirically estimated on the basis of standardised residuals).

Value

As a result, the function provides a list with the following components:

call	The matched call.
p	Set of false positive fractions (FPF) at which the pooled ROC curve has been estimated

ROC	Estimated covariate-adjusted ROC curve (AROC), and 95% pointwise confidence intervals (if required)
AUC	Estimated area under the covariate-adjusted ROC curve (AAUC), and 95% pointwise confidence intervals (if required).
fit.h	Object of class lm with the fitted regression model in the healthy population.
est.surv.h	The value of the argument est.surv.h used in the call.

References

Janes, H., and Pepe, M.S. (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. Biometrika, 96(2), 371 - 382.

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m3 <- AROC.sp(formula.healthy = l_marker1 ~ age,
group = "status", tag.healthy = 0, data = newpsa,
p = seq(0,1,l=101), B = 500)

summary(m3)

plot(m3)
```

---

compute.threshold.AROC.bnp

AROC-based threshold values.

Description

Estimates AROC-based threshold values using the nonparametric Bayesian approach proposed by Inacio de Carvalho and Rodriguez-Alvarez (2018).

Usage

```
compute.threshold.AROC.bnp(object, newdata, FPF = 0.5)
```

Arguments

object	An object of class AROC as produced by AROC.bnp.
newdata	Data frame with the covariate values at which threshold values are required.
FPF	Numeric vector with the FPF at which to calculate the AROC-based threshold values. Atomic values are also valid.

Details

Estimation of the covariate-adjusted ROC curve (AROC) using the nonparametric Bayesian approach proposed by Inacio de Carvalho and Rodriguez-Alvarez (2018) involves the estimation of the conditional distribution function for the diagnostic test outcome in the healthy population

$$F_{\bar{D}}(y|\mathbf{X}_{\bar{D}}) = Pr\{Y_{\bar{D}} \leq y|\mathbf{X}_{\bar{D}}\}.$$

This function makes use of this estimate in order to calculate AROC-based threshold values. In particular, for a covariate value $\mathbf{x}$ and a FPF = t, the AROC-based threshold value at the s-th posterior sample ($s = 1, \ldots, S$) is calculated as follows

$$c_{\mathbf{x}}^{(s)} = \hat{F}_{\bar{D}}^{-1(s)}(1 - t|\mathbf{X}_{\bar{D}} = \mathbf{x}).$$

from which the posterior mean can be computed

$$\hat{c}_{\mathbf{x}} = \frac{1}{S}\sum_{s=1}^{S} c_{\mathbf{x}}^{(s)}.$$

Value

As a result, the function provides a list with the following components:

thresholds.est	A matrix with the posterior mean of the AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.
thresholds.ql	A matrix with the posterior 2.5% quantile of the AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.
thresholds.qh	A matrix with the posterior 97.5% quantile of the AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.

References

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

See Also

AROC.bnp

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)


m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

# Compute the threshold values
FPF = c(0.1, 0.3)
newdata <- data.frame(age = seq(52, 80, l = 50))

th_bnp <- compute.threshold.AROC.bnp(m0, newdata, FPF)
names(th_bnp)
```

---

compute.threshold.AROC.bsp

AROC-based threshold values.

Description

Estimates AROC-based threshold values using the semiparametric Bayesian normal linear regression model discussed in Inacio de Carvalho and Rodriguez-Alvarez (2018).

Usage

```
compute.threshold.AROC.bsp(object, newdata, FPF = 0.5)
```

Arguments

object	An object of class AROC as produced by AROC.bsp.
newdata	Data frame with the covariate values at which threshold values are required.
FPF	Numeric vector with the FPF at which to calculate the AROC-based threshold values. Atomic values are also valid.

Details

Estimation of the covariate-adjusted ROC curve (AROC) using the semiparametric Bayesian normal linear regression model discussed in Inacio de Carvalho and Rodriguez-Alvarez (2018) involves the estimation of the conditional distribution function for the diagnostic test outcome in the healthy population

$$F_{\bar{D}}(y|\mathbf{X}_{\bar{D}}) = Pr\{Y_{\bar{D}} \leq y|\mathbf{X}_{\bar{D}}\}.$$

This function makes use of this estimate in order to calculate AROC-based threshold values. In particular, for a covariate value $\mathbf{x}$ and a FPF = t, the AROC-based threshold value at the s-th posterior sample ($s = 1, \ldots, S$) is calculated as follows

$$c_{\mathbf{x}}^{(s)} = \hat{F}_{\bar{D}}^{-1(s)}(1 - t|\mathbf{X}_{\bar{D}} = \mathbf{x}).$$

from which the posterior mean can be computed

$$\hat{c}_{\mathbf{x}} = \frac{1}{S}\sum_{s=1}^{S} c_{\mathbf{x}}^{(s)}.$$

Value

As a result, the function provides a list with the following components:

`thresholds.est` A matrix with the posterior mean of the AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.

`thresholds.ql` A matrix with the posterior 2.5% quantile of the AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.

`thresholds.qh` A matrix with the posterior 97.5% quantile of the AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.

References

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

See Also

`AROC.bsp`

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m1 <- AROC.bsp(formula.healthy = l_marker1 ~ age,
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, nsim = 5000, nburn = 1500)

# Compute the threshold values
FPF = c(0.1, 0.3)
newdata <- data.frame(age = seq(52, 80, l = 50))

th_bsp <- compute.threshold.AROC.bsp(m1, newdata, FPF)
names(th_bsp)
```

`compute.threshold.AROC.kernel` *AROC-based threshold values.*

Description

Estimates AROC-based threshold values using the nonparametric kernel-based method proposed by Rodriguez-Alvarez et al. (2011).

Usage

```
compute.threshold.AROC.kernel(object, newcovariate, FPF = 0.5)
```

Arguments

`object`	An object of class `AROC` as produced by `AROC.kernel`.
`newcovariate`	Numeric vector with the covariate values at which threshold values are required.
`FPF`	Numeric vector with the FPF at which to calculate the AROC-based threshold values. Atomic values are also valid.

Details

Estimation of the covariate-adjusted ROC curve (AROC) using the nonparametric kernel-based method proposed by Rodriguez-Alvarez et al. (2011) involves the estimation of the conditional distribution function for the diagnostic test outcome in the healthy population

$$F_{\bar{D}}(y|X_{\bar{D}}) = Pr\{Y_{\bar{D}} \leq y|X_{\bar{D}}\}.$$

This function makes use of this estimate in order to calculate AROC-based threshold values. In particular, for a covariate value x and a FPF = t, the AROC-based threshold value is calculated as follows

$$\hat{c}_{\mathbf{x}} = \hat{F}^{-1}_{\bar{D}}(1 - t|\mathbf{X}_{\bar{D}} = \mathbf{x}).$$

Value

A matrix with the computed AROC-based threshold values. The matrix has as many columns as different covariate values, and as many rows as different FPFs.

References

Hayfield, T., and Racine, J. S.(2008). Nonparametric Econometrics: The np Package. Journal of Statistical Software 27(5). URL http://www.jstatsoft.org/v27/i05/.

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

Rodriguez-Alvarez, M. X., Roca-Pardinas, J., and Cadarso-Suarez, C. (2011). ROC curve and covariates: extending induced methodology to the non-parametric framework. Statistics and Computing, 21(4), 483 - 499.

See Also

AROC.kernel

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m2 <- AROC.kernel(marker = "l_marker1", covariate = "age",
group = "status", tag.healthy = 0, data = newpsa,
p = seq(0,1,l=101), B = 500)

# Compute the threshold values
cov.values <- seq(52, 80, l = 50)
FPF = c(0.1, 0.3)
```

```
th_np <- compute.threshold.AROC.kernel(m2, cov.values, FPF)
```

`compute.threshold.AROC.sp` *AROC-based threshold values.*

Description

Estimates AROC-based threshold values using the semiparametric approach proposed by Janes and Pepe (2009).

Usage

```
compute.threshold.AROC.sp(object, newdata, FPF = 0.5)
```

Arguments

`object`	An object of class `AROC` as produced by `AROC.sp`.
`newdata`	Data frame with the covariate values at which threshold values are required.
`FPF`	Numeric vector with the FPF at which to calculate the AROC-based threshold values. Atomic values are also valid.

Details

Estimation of the covariate-adjusted ROC curve (AROC) using the semiparametric approach proposed by Janes and Pepe (2009) involves the estimation of the conditional distribution function for the diagnostic test outcome in the healthy population

$$F_{\bar{D}}(y|\mathbf{X}_{\bar{D}}) = Pr\{Y_{\bar{D}} \leq y|\mathbf{X}_{\bar{D}}\}.$$

This function makes use of this estimate in order to calculate AROC-based threshold values. In particular, for a covariate value $\mathbf{x}$ and a FPF = t, the AROC-based threshold value is calculated as follows

$$\hat{c}_{\mathbf{x}} = \hat{F}_{\bar{D}}^{-1}(1 - t|\mathbf{X}_{\bar{D}} = \mathbf{x}).$$

Value

A matrix with the computed AROC-based threshold values. The matrix has as many columns as different covariate vector values, and as many rows as different FPFs.

References

Janes, H., and Pepe, M.S. (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. Biometrika, 96(2), 371 - 382.

See Also

AROC.sp

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m3 <- AROC.sp(formula.healthy = l_marker1 ~ age,
group = "status", tag.healthy = 0, data = newpsa,
p = seq(0,1,l=101), B = 500)

FPF = c(0.1, 0.3)
newdata <- data.frame(age = seq(52, 80, l = 50))

th_sp <- compute.threshold.AROC.sp(m3, newdata, FPF)
names(th_sp)
```

---

compute.threshold.pooledROC.BB

Pooled ROC-based threshold values.

Description

Estimates pooled ROC-based threshold values using the Bayesian bootstrap estimator proposed by Gu et al. (2008).

Usage

```
compute.threshold.pooledROC.BB(object, FPF = 0.5)
```

Arguments

object	An object of class AROC as produced by pooledROC.BB.
FPF	Numeric vector with the FPF at which to calculate the pooled ROC-based threshold values. Atomic values are also valid.

Value

As a result, the function provides a list with the following components:

thresholds	A matrix with the posterior mean and posterior 2.5% and 97.5% quantiles of the pooled ROC-based threshold values. The matrix has as many rows as different FPFs.
FPF	the supplied FPF argument
TPF	TPFs corresponding to the estimated threshold. In addition to the posterior mean, the 95% pointwise credible band is also returned.

References

Gu, J., Ghosal, S., and Roy, A. (2008). Bayesian bootstrap estimation of ROC curve. Statistics in Medicine, **27**, 5407–5420.

See Also

pooledROC.BB

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)


m0_BB <- pooledROC.BB(y0 = newpsa$l_marker1[newpsa$status == 0],
            y1 = newpsa$l_marker1[newpsa$status == 1], p = seq(0,1,l=101), B = 5000)

### Threshold values for a fixed FPF
th_m0_BB <- compute.threshold.pooledROC.BB(m0_BB, FPF = 0.1)

th_m0_BB$threshold
```

---

compute.threshold.pooledROC.emp

Pooled ROC-based threshold values.

Description

Estimates pooled ROC-based threshold values using the empirical estimator proposed by Hsieh and Turnbull (1996).

Usage

```
compute.threshold.pooledROC.emp(object, FPF = 0.5)
```

Arguments

object	An object of class AROC as produced by pooledROC.emp.
FPF	Numeric vector with the FPF at which to calculate the pooled ROC-based threshold values. Atomic values are also valid.

Value

As a result, the function provides a list with the following components:

thresholds	A vector with the estimated pooled ROC-based threshold values, one for each specified FPF.
FPF	the supplied FPF argument
TPF	TPFs corresponding to the estimated threshold.

References

Hsieh, F., and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve, The Annals of Statistics, **24**, 25–40.

See Also

pooledROC.emp

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)


m0_emp <- pooledROC.emp(y0 = newpsa$l_marker1[newpsa$status == 0],
            y1 = newpsa$l_marker1[newpsa$status == 1], p = seq(0,1,l=101), B = 500)

### Threshold values for a fixed FPF
th_m0_emp <- compute.threshold.pooledROC.emp(m0_emp, FPF = 0.1)

th_m0_emp$threshold
```

paauccontrol *Partial area under the covariate-adjusted ROC curve (pAAUC).*

Description

Used to set various parameters controlling the estimation of the partial area under the covariate-adjusted ROC curve (pAAUC).

Usage

```
paauccontrol(compute = FALSE, value = 1)
```

Arguments

compute	Logical value. If TRUE the partial area under the covariate-adjusted ROC curve (pAAUC) is estimated.
value	Numeric value. Pre-specified maximum false positive fraction (FPF) at which to calculate the pAAUC.

Details

The value returned by this function is used as a control argument of the AROC.bnp and AROC.bsp functions.

Value

a list with components for each of the possible arguments.

References

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. Technical report.

See Also

AROC.bnp and AROC.bsp

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
```

```
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), paauc = list(compute = TRUE, value = 0.3),
compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

summary(m0)
```

`plot.AROC` *Default AROC plotting*

Description

Takes a fitted `AROC` object produced by `AROC.bnp()`, `AROC.bsp()`, `AROC.sp()`, `AROC.kernel()`, `pooledROC.BB()` or `pooledROC.emp()` and plots the covariate-adjusted ROC curve (AROC) and associated area under the AROC (AAUC); or the pooled ROC curve and associated AUC.

Usage

```
## S3 method for class 'AROC'
plot(x, ...)
```

Arguments

`x`	an object of class `AROC` as produced by `AROC.bnp()`, `AROC.bsp()`, `AROC.sp()`, `AROC.kernel()`, `pooledROC.BB()` or `pooledROC.emp()`
`...`	further arguments passed to or from other methods

See Also

`AROC.bnp`, `AROC.bsp`, `AROC.sp`, `AROC.kernel`, `pooledROC.BB` or `pooledROC.emp`.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

plot(m0)
```

pooledROC.BB *Bayesian bootstrap estimation of the pooled ROC curve.*

Description

Estimates the pooled ROC curve using the Bayesian bootstrap estimator proposed by Gu et al. (2008).

Usage

```
pooledROC.BB(y0, y1, p = seq(0, 1, l = 101), B = 5000)
```

Arguments

y0	Diagnostic test outcomes in the healthy group.
y1	Diagnostic test outcomes in the diseased group.
p	Set of false positive fractions (FPF) at which to estimate the covariate-adjusted ROC curve.
B	An integer value specifying the number of Bayesian bootstrap resamples. By default 5000.

Value

As a result, the function provides a list with the following components:

call	the matched call.
p	Set of false positive fractions (FPF) at which the pooled ROC curve has been estimated
ROC	Estimated pooled ROC curve, and corresponding 95% credible intervals
AUC	Estimated pooled AUC, and corresponding 95% credible intervals.

References

Gu, J., Ghosal, S., and Roy, A. (2008). Bayesian bootstrap estimation of ROC curve. Statistics in Medicine, 27(26), 5407 - 5420.

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0_BB <- pooledROC.BB(newpsa$l_marker1[newpsa$status == 0],
newpsa$l_marker1[newpsa$status == 1], p = seq(0,1,l=101), B = 5000)

summary(m0_BB)

plot(m0_BB)
```

`pooledROC.emp` *Empirical estimation of the pooled ROC curve.*

Description

Estimates the pooled ROC curve using the empirical estimator proposed by Hsieh and Turnbull (1996).

Usage

```
pooledROC.emp(y0, y1, p = seq(0, 1, l = 101), B = 500,
method = c("ncoutcome", "coutcome"))
```

Arguments

`y0`	Diagnostic test outcomes in the healthy group.
`y1`	Diagnostic test outcomes in the diseased group.
`p`	Set of false positive fractions (FPF) at which to estimate the covariate-adjusted ROC curve.
`B`	An integer value specifying the number of bootstrap resamples for the construction of the confidence intervals. By default 500.
`method`	A character string specifying if bootstrap resampling (for the confidence intervals) should be done with or without regard to the disease status ("coutcome" or "noutcome"). In both cases, a naive bootstrap is used. By default, the resampling is done conditionally on the disease status.

Value

As a result, the function provides a list with the following components:

call	the matched call.
p	Set of false positive fractions (FPF) at which the pooled ROC curve has been estimated
ROC	Estimated pooled ROC curve, and corresponding 95% confidence intervals (if required)
AUC	Estimated pooled AUC, and corresponding 95% confidence intervals (if required).

References

Hsieh, F., and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve, The Annals of Statistics, 24, 25-40.

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0_emp <- pooledROC.emp(newpsa$l_marker1[newpsa$status == 0],
newpsa$l_marker1[newpsa$status == 1], p = seq(0,1,l=101), B = 500)

summary(m0_emp)

plot(m0_emp)
```

---

predictive.checks.AROC.bnp

Posterior predictive checks.

Description

Implements posterior predictive checks. Compares a selected test statistic computed based on the diagnostic test outcome in the nondiseased group against the same test statistics computed based on generated data from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group obtained using a B-splines dependent Dirichlet process mixture model as described in Inacio de Carvalho and Rodriguez-Alvarez (2018).

Usage

```
predictive.checks.AROC.bnp(object,
statistics = c("min", "max", "kurtosis", "skewness"), devnew = TRUE)
```

Arguments

object	An object of class AROC as produced by AROC.bnp.
statistics	Character vector. Statistics to be used for the posterior predictive checking. By default, "min", "max", "kurtosis" and "skewness"
devnew	logical. If TRUE, each plot is depicted in a new graphic device.

Details

The following graphics are depicted: (1) histograms of the desired statistics computed from a number of simulated datasets drawn from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group. In these plots, the estimated statistics from the observed diagnostic test outcome in the nondiseased group are also depicted. (2) Kernel density estimates computed from a number of simulated datasets drawn from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group. In these plots, the kernel density estimate of the observed diagnostic test outcome in the nondiseased group is also depicted. For a detailed discussion about predictive checks, see Gabry et al. (2017).

Value

As a result, the function provides a list with the following components:

yrep	Matrix. Each column corresponds to a dataset generated from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group.
y0	Numeric vector. Observed diagnostic test outcome in the nondiseased group.

References

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2017). Visualization in Bayesian workflow. arXiv preprint arXiv:1709.01449.

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

See Also

AROC.bnp

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
```

```
newpsa$l_marker1 <- log(newpsa$marker1)

m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

predictive.checks.AROC.bnp(m0, statistics = "skewness")
```

`predictive.checks.AROC.bsp` *Posterior predictive checks.*

Description

Implements posterior predictive checks. Compares a selected test statistic computed based on the diagnostic test outcome in the nondiseased group against the same test statistics computed based on generated data from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group obtained using a Bayesian normal linear regression model as discussed in Inacio de Carvalho and Rodriguez-Alvarez (2018).

Usage

```
predictive.checks.AROC.bsp(object,
statistics = c("min", "max", "median", "skewness"), devnew = TRUE)
```

Arguments

`object`	An object of class `AROC` as produced by `AROC.bsp`.
`statistics`	Character vector. Statistics to be used for the posterior predictive checking. By default, "min", "max", "median" and "skewness"
`devnew`	logical. If TRUE, each plot is depicted in a new graphic device.

Details

The following graphics are depicted: (1) histograms of the desired test statistics computed from a number of simulated datasets drawn from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group. In these plots, the estimated statistics from the observed diagnostic test outcome in the nondiseased group are also depicted. (2) Kernel density estimates computed from a number of simulated datasets drawn from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group. In these plots, the kernel density estimate of the observed diagnostic test outcome in the nondiseased group is also depicted. For a detailed discussion about predictive checks, see Gabry et al. (2017).

Value

As a result, the function provides a list with the following components:

yrep	Matrix. Each column corresponds to a dataset generated from the posterior predictive distribution of the diagnostic test outcome in the nondiseased group.
y0	Numeric vector. Observed diagnostic test outcome in the nondiseased group.

References

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2017). Visualization in Bayesian workflow. arXiv preprint arXiv:1709.01449.

Inacio de Carvalho, V., and Rodriguez-Alvarez, M. X. (2018). Bayesian nonparametric inference for the covariate-adjusted ROC curve. arXiv preprint arXiv:1806.00473.

See Also

AROC.bsp

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m1 <- AROC.bsp(formula.healthy = l_marker1 ~ age,
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, nsim = 5000, nburn = 1500)

predictive.checks.AROC.bsp(m1, statistics = "mean")
```

---

print.AROC *Print method for* AROC *objects*

Description

Default print method for objects fitted with AROC.bnp(), AROC.bsp(), AROC.sp(), AROC.kernel(), pooledROC.BB() or pooledROC.emp() functions.

Usage

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

Arguments

x	an object of class AROC as produced by AROC.bnp(), AROC.bsp(), AROC.sp(), AROC.kernel(), pooledROC.BB() or pooledROC.emp()
...	further arguments passed to or from other methods. Not yet implemented.

Details

A short summary is printed including: TODO

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

m0
```

---

psa *Prostate specific antigen (PSA) biomarker study.*

Description

The dataset contains 71 prostate cases and 71 controls who participated in a lung cancer prevention trial (CARET, Beta-carotene and retinol trial). For details, see Etzioni et al. (1999) and Pepe (2003).

Usage

```
data("psa")
```

Format

A data frame with 683 observations on the following 6 variables.

`id` Patient identifier.

`marker1` total prostate specific antigen (PSA).

`marker2` free prostate specific antigen (PSA)

`status` presence/absence of prostate cancer. The non-cancer patients are controls matched to cases on age and number of serum samples available for analysis (see Details).

`age` patient age at blood draw (serum sample).

`t` time (years) relative to prostate cancer diagnosis.

Details

The CARET enrolled 12000 men, aged between 50 and 65 years, at high risk of lung cancer. For each subject on the study, serum samples were drawn at baseline and at two-year intervals after that. The data presented here represent a subsample of the original sample, and it was reported by Etzioni et al. (1999). It contains 71 cases of prostate cancer that occurred during the study. All these cases had, at least, three and up to eight serum samples. As far as controls are concerned, they were selected from the participants of the CARET study verifying that they had not been diagnosed with prostate cancer by the time of the original study, and the selection was done by matching to cases on date of birth and number of serum samples available for analysis.

Source

The dataset can be downloaded from https://research.fredhutch.org/diagnostic-biomarkers-center/en/datasets.html.

References

Pepe, M. S. (2003). The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford Statistical Science Series. Oxford University Press, New York.

Etzioni, R., Pepe, M. S., Longton, G., Hu. C., and Goodman, G. (1999). Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. Medical Decision Making, 19(3), 242-251.

Examples

```
data(psa)
summary(psa)
```

summary.AROC *Summary method for* AROC *objects*

Description

Default summary method for objects fitted with AROC.bnp(), AROC.bsp(), AROC.sp(), AROC.kernel(), pooledROC.BB() or pooledROC.emp() functions.

Usage

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

Arguments

object	an object of class AROC as produced by AROC.bnp(), AROC.bsp(), AROC.sp(), AROC.kernel(), pooledROC.BB() or pooledROC.emp()
...	Further arguments passed to or from other methods. Not yet implemented.

Details

A short summary is printed including the area under the covariate-adjusted ROC curve (AAUC) or the area under the pooled ROC curve (AUC), and if required, the partial area under the covariate-adjusted ROC curve (pAAUC). For the Bayesian methods, and if required, the function also provides the log pseudo marginal likelihood (LPML) and/or widely applicable information criterion (WAIC).

See Also

AROC.bnp, AROC.bsp, AROC.sp, AROC.kernel, pooledROC.BB or pooledROC.emp.

Examples

```
library(AROC)
data(psa)
# Select the last measurement
newpsa <- psa[!duplicated(psa$id, fromLast = TRUE),]

# Log-transform the biomarker
newpsa$l_marker1 <- log(newpsa$marker1)

m0 <- AROC.bnp(formula.healthy = l_marker1 ~ f(age, K = 0),
group = "status", tag.healthy = 0, data = newpsa, scale = TRUE,
p = seq(0,1,l=101), compute.lpml = TRUE, compute.WAIC = TRUE,
a = 2, b = 0.5, L = 10, nsim = 5000, nburn = 1000)

summary(m0)
```

Index