Package ‘eco’

August 1, 2017

Version 4.0-1
Date 2017-7-26
Title Ecological Inference in 2x2 Tables
Maintainer Ying Lu <ying.lu@nyu.edu>
Depends R (>= 2.0), MASS, utils
Description Implements the Bayesian and likelihood methods proposed in Imai, Lu, and Strauss (2008 <DOI: 10.1093/pan/mpm017>) and (2011 <DOI:10.18637/jss.v042.i05>) for ecological inference in 2 by 2 tables as well as the method of bounds introduced by Duncan and Davis (1953). The package fits both parametric and nonparametric models using either the Expectation-Maximization algorithms (for likelihood models) or the Markov chain Monte Carlo algorithms (for Bayesian models). For all models, the individual-level data can be directly incorporated into the estimation whenever such data are available. Along with in-sample and out-of-sample predictions, the package also provides a functionality which allows one to quantify the effect of data aggregation on parameter estimation and hypothesis testing under the parametric likelihood models.

LazyLoad yes
LazyData yes
License GPL (>= 2)
URL https://github.com/kosukeimai/eco
BugReports https://github.com/kosukeimai/eco/issues
RoxygenNote 6.0.1
NeedsCompilation yes
Author Kosuke Imai [aut],
Ying Lu [aut, cre],
Aaron Strauss [aut],
Hubert Jin [ctb]
Repository CRAN
Date/Publication 2017-08-01 05:24:50 UTC
Description

This data set contains the proportion of the residents who are black, the proportion of those who can read, the total population as well as the actual black literacy rate and white literacy rate for 1040 counties in the US. The dataset was originally analyzed by Robinson (1950) at the state level. King (1997) recoded the 1910 census at county level. The data set only includes those who are older than 10 years of age.

Format

A data frame containing 5 variables and 1040 observations

X numeric the proportion of Black residents in each county
Y numeric the overall literacy rates in each county
N numeric the total number of residents in each county
W1 numeric the actual Black literacy rate
W2 numeric the actual White literacy rate
References


**eco**

*Fitting the Parametric Bayesian Model of Ecological Inference in 2x2 Tables*

**Description**

eco is used to fit the parametric Bayesian model (based on a Normal/Inverse-Wishart prior) for ecological inference in $2 \times 2$ tables via Markov chain Monte Carlo. It gives the in-sample predictions as well as the estimates of the model parameters. The model and algorithm are described in Imai, Lu and Strauss (2008, 2011).

**Usage**

```
eco(formula, data = parent.frame(), N = NULL, supplement = NULL,
     context = FALSE, mu0 = 0, tau0 = 2, nu0 = 4, S0 = 10,
     mu.start = 0, Sigma.start = 10, parameter = TRUE, grid = FALSE,
     n.draws = 5000, burnin = 0, thin = 0, verbose = FALSE)
```

**Arguments**

- **formula**: A symbolic description of the model to be fit, specifying the column and row margins of $2 \times 2$ ecological tables. $Y \sim X$ specifies $Y$ as the column margin (e.g., turnout) and $X$ as the row margin (e.g., percent African-American). Details and specific examples are given below.
- **data**: An optional data frame in which to interpret the variables in formula. The default is the environment in which eco is called.
- **N**: An optional variable representing the size of the unit; e.g., the total number of voters. $N$ needs to be a vector of same length as $Y$ and $X$ or a scalar.
- **supplement**: An optional matrix of supplemental data. The matrix has two columns, which contain additional individual-level data such as survey data for $W_1$ and $W_2$, respectively. If NULL, no additional individual-level data are included in the model. The default is NULL.
- **context**: Logical. If TRUE, the contextual effect is also modeled, that is to assume the row margin $X$ and the unknown $W_1$ and $W_2$ are correlated. See Imai, Lu and Strauss (2008, 2011) for details. The default is FALSE.
- **mu0**: A scalar or a numeric vector that specifies the prior mean for the mean parameter $\mu$ for $(W_1, W_2)$ (or for $(W_1, W_2, X)$ if context=TRUE). When the input of mu0 is a scalar, its value will be repeated to yield a vector of the length of $\mu$, otherwise, it needs to be a vector of same length as $\mu$. When context=TRUE, the length of $\mu$ is 3, otherwise it is 2. The default is 0.
 tau0    A positive integer representing the scale parameter of the Normal-Inverse Wishart prior for the mean and variance parameter \((\mu, \Sigma)\). The default is 2.

 nu0    A positive integer representing the prior degrees of freedom of the Normal-Inverse Wishart prior for the mean and variance parameter \((\mu, \Sigma)\). The default is 4.

 S0    A positive scalar or a positive definite matrix that specifies the prior scale matrix of the Normal-Inverse Wishart prior for the mean and variance parameter \((\mu, \Sigma)\). If it is a scalar, then the prior scale matrix will be a diagonal matrix with the same dimensions as \(\Sigma\) and the diagonal elements all take value of \(S0\), otherwise \(S0\) needs to have same dimensions as \(\Sigma\). When context=TRUE, \(\Sigma\) is a 3 \times 3 matrix, otherwise, it is 2 \times 2. The default is 10.

 mu.start    A scalar or a numeric vector that specifies the starting values of the mean parameter \(\mu\). If it is a scalar, then its value will be repeated to yield a vector of the length of \(\mu\), otherwise, it needs to be a vector of same length as \(\mu\). When context=FALSE, the length of \(\mu\) is 2, otherwise it is 3. The default is 0.

 Sigma.start    A scalar or a positive definite matrix that specified the starting value of the variance matrix \(\Sigma\). If it is a scalar, then the prior scale matrix will be a diagonal matrix with the same dimensions as \(\Sigma\) and the diagonal elements all take value of \(S0\), otherwise \(S0\) needs to have same dimensions as \(\Sigma\). When context=TRUE, \(\Sigma\) is a 3 \times 3 matrix, otherwise, it is 2 \times 2. The default is 10.

 parameter    Logical. If TRUE, the Gibbs draws of the population parameters, \(\mu\) and \(\Sigma\), are returned in addition to the in-sample predictions of the missing internal cells, \(W\). The default is TRUE.

 grid    Logical. If TRUE, the grid method is used to sample \(W\) in the Gibbs sampler. If FALSE, the Metropolis algorithm is used where candidate draws are sampled from the uniform distribution on the tomography line for each unit. Note that the grid method is significantly slower than the Metropolis algorithm. The default is FALSE.

 n.draws    A positive integer. The number of MCMC draws. The default is 50000.

 burnin    A positive integer. The burnin interval for the Markov chain; i.e. the number of initial draws that should not be stored. The default is 0.

 thin    A positive integer. The thinning interval for the Markov chain; i.e. the number of Gibbs draws between the recorded values that are skipped. The default is 0.

 verbose    Logical. If TRUE, the progress of the Gibbs sampler is printed to the screen. The default is FALSE.

 Details

 An example of 2 \times 2 ecological table for racial voting is given below:

<table>
<thead>
<tr>
<th></th>
<th>black voters (W_{1i})</th>
<th>white voters (W_{2i})</th>
<th>(Y_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>vote</td>
<td>(1 - W_{1i})</td>
<td>(1 - W_{2i})</td>
<td>(1 - Y_i)</td>
</tr>
<tr>
<td>not vote</td>
<td>(X_i)</td>
<td>(1 - X_i)</td>
<td></td>
</tr>
</tbody>
</table>

Details

An example of 2 \times 2 ecological table for racial voting is given below:
where \( Y_i \) and \( X_i \) represent the observed margins, and \( W_1 \) and \( W_2 \) are unknown variables. In this example, \( Y_i \) is the turnout rate in the \( i \)th precinct, \( X_i \) is the proportion of African American in the \( i \)th precinct. The unknowns \( W_{1i} \) and \( W_{2i} \) are the black and white turnout, respectively. All variables are proportions and hence bounded between 0 and 1. For each \( i \), the following deterministic relationship holds, \( Y_i = X_i W_{1i} + (1 - X_i) W_{2i} \).

**Value**

An object of class eco containing the following elements:

- **call**: The matched call.
- **X**: The row margin, \( X \).
- **y**: The column margin, \( Y \).
- **N**: The size of each table, \( N \).
- **burnin**: The number of initial burnin draws.
- **thin**: The thinning interval.
- **nu0**: The prior degrees of freedom.
- **tau0**: The prior scale parameter.
- **mu0**: The prior mean.
- **S0**: The prior scale matrix.
- **W**: A three dimensional array storing the posterior in-sample predictions of \( W \). The first dimension indexes the Monte Carlo draws, the second dimension indexes the columns of the table, and the third dimension represents the observations.
- **wmin**: A numeric matrix storing the lower bounds of \( W \).
- **wmax**: A numeric matrix storing the upper bounds of \( W \).

The following additional elements are included in the output when parameter = TRUE.

- **mu**: The posterior draws of the population mean parameter, \( \mu \).
- **Sigma**: The posterior draws of the population variance matrix, \( \Sigma \).

**Author(s)**

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University, <ying.lu@nyu.Edu>

**References**


See Also

ecomL, ecoNP, predict.eco, summary.eco

Examples

```r
## load the registration data
## Not run: data(reg)

## NOTE: convergence has not been properly assessed for the following
## examples. See Imai, Lu and Strauss (2008, 2011) for more
## complete analyses.

## fit the parametric model with the default prior specification
res <- eco(Y ~ X, data = reg, verbose = TRUE)
## summarize the results
summary(res)

## obtain out-of-sample prediction
out <- predict(res, verbose = TRUE)
## summarize the results
summary(out)

## load the Robinson's census data
data(census)

## fit the parametric model with contextual effects and N
## using the default prior specification
res1 <- eco(Y ~ X, N = N, context = TRUE, data = census, verbose = TRUE)
## summarize the results
summary(res1)

## obtain out-of-sample prediction
out1 <- predict(res1, verbose = TRUE)
## summarize the results
summary(out1)

## End(Not run)
```

---

ecoBD  

*Calculating the Bounds for Ecological Inference in RxC Tables*

Description

ecoBD is used to calculate the bounds for missing internal cells of $R \times C$ ecological table. The data can be entered either in the form of counts or proportions.
Usage

ecoBD(formula, data = parent.frame(), N = NULL)

Arguments

formula A symbolic description of ecological table to be used, specifying the column and row margins of \( R \times C \) ecological tables. Details and specific examples are given below.
data An optional data frame in which to interpret the variables in formula. The default is the environment in which ecoBD is called.
N An optional variable representing the size of the unit; e.g., the total number of voters. If formula is entered as counts and the last row and/or column is omitted, this input is necessary.

Details

The data may be entered either in the form of counts or proportions. If proportions are used, formula may omit the last row and/or column of tables, which can be calculated from the remaining margins. For example, \( Y \sim X \) specifies \( Y \) as the first column margin and \( X \) as the first row margin in \( 2 \times 2 \) tables. If counts are used, formula may omit the last row and/or column margin of the table only if \( N \) is supplied. In this example, the columns will be labeled as \( X \) and \( Y \), and the rows will be labeled as \( X \) and \( Y \).

For larger tables, one can use cbind() and +. For example, cbind(Y1, Y2, Y3) ~ X1 + X2 + X3 + X4) specifies \( 3 \times 4 \) tables.

An \( R \times C \) ecological table in the form of counts:

\[
\begin{array}{cccc}
n_{i11} & n_{i12} & \cdots & n_{i1C} \\
n_{i21} & n_{i22} & \cdots & n_{i2C} \\
\vdots & \vdots & \ddots & \vdots \\
n_{iR1} & n_{iR2} & \cdots & n_{iRC} \\
n_{i,1} & n_{i,2} & \cdots & n_{i,C} \\
\end{array}
\]

where \( n_{ir} \) and \( n_{ic} \) represent the observed margins, \( N_i \) represents the size of the table, and \( n_{irc} \) are unknown variables. Note that for each \( i \), the following deterministic relationships hold: \( n_{ir} = \sum_{c=1}^{C} n_{irc} \) for \( r = 1, \ldots, R \), and \( n_{ic} = \sum_{r=1}^{R} n_{irc} \) for \( c = 1, \ldots, C \). Then, each of the unknown inner cells can be bounded in the following manner,

\[
\max(0, n_{ir} + n_{ic} - N_i) \leq n_{irc} \leq \min(n_{ir}, n_{ic}).
\]

If the size of tables, \( N \), is provided,

An \( R \times C \) ecological table in the form of proportions:

\[
\begin{array}{cccc}
W_{i11} & W_{i12} & \cdots & W_{i1C} \\
W_{i21} & W_{i22} & \cdots & W_{i2C} \\
\vdots & \vdots & \ddots & \vdots \\
W_{iR1} & W_{iR2} & \cdots & W_{iRC} \\
X_{i1} & X_{i2} & \cdots & X_{iC} \\
\end{array}
\]
where \( Y_{ir} \) and \( X_{ic} \) represent the observed margins, and \( W_{irc} \) are unknown variables. Note that for each \( i \), the following deterministic relationships hold: 
\[
Y_{ir} = \sum_{c=1}^{C} X_{ic} W_{irc} \quad \text{for} \quad r = 1, \ldots, R, \quad \text{and} \quad \sum_{r=1}^{R} W_{irc} = 1 \quad \text{for} \quad c = 1, \ldots, C.
\]
Then, each of the inner cells of the table can be bounded in the following manner,
\[
\max(0, (X_{ic} + Y_{ir} - 1)/X_{ic}) \leq W_{irc} \leq \min(1, Y_{ir}/X_{ir}).
\]

**Value**

An object of class `ecoBD` containing the following elements (When three dimensional arrays are used, the first dimension indexes the observations, the second dimension indexes the row numbers, and the third dimension indexes the column numbers):

- `call` The matched call.
- `X` A matrix of the observed row margin, \( X \).
- `Y` A matrix of the observed column margin, \( Y \).
- `N` A vector of the size of ecological tables, \( N \).
- `aggWmin` A three dimensional array of aggregate lower bounds for proportions.
- `aggWmax` A three dimensional array of aggregate upper bounds for proportions.
- `Wmin` A three dimensional array of lower bounds for proportions.
- `Wmax` A three dimensional array of upper bounds for proportions.
- `Nmin` A three dimensional array of lower bounds for counts.
- `Nmax` A three dimensional array of upper bounds for counts.

The object can be printed through `print.ecoBD`.

**Author(s)**

Kosuke Imai, Department of Politics, Princeton University <kimai@princeton.Edu>, [http://imai.princeton.edu/](http://imai.princeton.edu/); Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

**References**


**See Also**

`eco`, `ecoNP`
ecoML

Examples

```r
## load the registration data
data(reg)

## calculate the bounds
res <- ecoBD(y ~ x, N = N, data = reg)

## print the results
print(res)
```

ecoML  

Fitting Parametric Models and Quantifying Missing Information for Ecological Inference in 2x2 Tables

Description

ecoML is used to fit parametric models for ecological inference in $2 \times 2$ tables via Expectation Maximization (EM) algorithms. The data is specified in proportions. At it’s most basic setting, the algorithm assumes that the individual-level proportions (i.e., $W_1$ and $W_2$) and distributed bivariate normally (after logit transformations). The function calculates point estimates of the parameters for models based on different assumptions. The standard errors of the point estimates are also computed via Supplemented EM algorithms. Moreover, ecoML quantifies the amount of missing information associated with each parameter and allows researcher to examine the impact of missing information on parameter estimation in ecological inference. The models and algorithms are described in Imai, Lu and Strauss (2008, 2011).

Usage

```r
ecoML(formula, data = parent.frame(), N = NULL, supplement = NULL,
theta.start = c(0, 0, 1, 1, 0), fix.rho = FALSE, context = FALSE,
sem = TRUE, epsilon = 10^(-6), maxit = 1000, loglik = TRUE,
hyptest = FALSE, verbose = FALSE)
```

Arguments

- `formula`: A symbolic description of the model to be fit, specifying the column and row margins of $2 \times 2$ ecological tables. $Y \sim X$ specifies $Y$ as the column margin (e.g., turnout) and $X$ (e.g., percent African-American) as the row margin. Details and specific examples are given below.
- `data`: An optional data frame in which to interpret the variables in `formula`. The default is the environment in which ecoML is called.
- `N`: An optional variable representing the size of the unit; e.g., the total number of voters. $N$ needs to be a vector of same length as $Y$ and $X$ or a scalar.
supplement
An optional matrix of supplemental data. The matrix has two columns, which contain additional individual-level data such as survey data for \( W_1 \) and \( W_2 \), respectively. If NULL, no additional individual-level data are included in the model. The default is NULL.

theta.start
A numeric vector that specifies the starting values for the mean, variance, and covariance. When \( \text{context} = \text{FALSE} \), the elements of theta.start correspond to \( (E(W_1), E(W_2), \text{var}(W_1), \text{var}(W_2), \text{cor}(W_1, W_2)) \). When \( \text{context} = \text{TRUE} \), the elements of theta.start correspond to \( (E(W_1), E(W_2), \text{var}(W_1), \text{var}(W_2), \text{corr}(W_1, X), \text{corr}(W_2, X), \text{corr}(W_1, W_2)) \). Moreover, when fix.rho=true, \( \text{corr}(W_1, W_2) \) is set to be the correlation between \( W_1 \) and \( W_2 \) when \( \text{context} = \text{FALSE} \), and the partial correlation between \( W_1 \) and \( W_2 \) given \( X \) when \( \text{context} = \text{FALSE} \). The default is \( c(0, 0, 1, 1, 0) \).

fix.rho
Logical. If TRUE, the correlation (when context=TRUE) or the partial correlation (when context=FALSE) between \( W_1 \) and \( W_2 \) is fixed through the estimation. For details, see Imai, Lu and Strauss(2006). The default is FALSE.

context
Logical. If TRUE, the contextual effect is also modeled. In this case, the row margin (i.e., \( X \)) and the individual-level rates (i.e., \( W_1 \) and \( W_2 \)) are assumed to be distributed tri-variate normally (after logit transformations). See Imai, Lu and Strauss (2006) for details. The default is FALSE.

sem
Logical. If TRUE, the standard errors of parameter estimates are estimated via SEM algorithm, as well as the fraction of missing data. The default is TRUE.

epsilon
A positive number that specifies the convergence criterion for EM algorithm. The square root of epsilon is the convergence criterion for SEM algorithm. The default is \( 10^{-6} \).

maxit
A positive integer specifies the maximum number of iterations before the convergence criterion is met. The default is \( 10^5 \).

loglik
Logical. If TRUE, the value of the log-likelihood function at each iteration of EM is saved. The default is TRUE.

hyptest
Logical. If TRUE, model is estimated under the null hypothesis that means of \( W_1 \) and \( W_2 \) are the same. The default is FALSE.

verbose
Logical. If TRUE, the progress of the EM and SEM algorithms is printed to the screen. The default is FALSE.

Details
When SEM is TRUE, ecoML computes the observed-data information matrix for the parameters of interest based on Supplemented-EM algorithm. The inverse of the observed-data information matrix can be used to estimate the variance-covariance matrix for the parameters estimated from EM algorithms. In addition, it also computes the expected complete-data information matrix. Based on these two measures, one can further calculate the fraction of missing information associated with each parameter. See Imai, Lu and Strauss (2006) for more details about fraction of missing information. Moreover, when hyptest=TRUE, ecoML allows to estimate the parametric model under the null hypothesis that \( \mu_1=\mu_2 \). One can then construct the likelihood ratio test to assess the hypothesis of equal means. The associated fraction of missing information for the test statistic can be also calculated. For details, see Imai, Lu and Strauss (2006) for details.
Value

An object of class ecoML containing the following elements:

call The matched call.
X The row margin, \(X\).
Y The column margin, \(Y\).
N The size of each table, \(N\).
context The assumption under which model is estimated. If context = FALSE, CAR assumption is adopted and no contextual effect is modeled. If context = TRUE, NCAR assumption is adopted, and contextual effect is modeled.
sem Whether SEM algorithm is used to estimate the standard errors and observed information matrix for the parameter estimates.
fix.rho Whether the correlation or the partial correlation between \(W_1\) an \(W_2\) is fixed in the estimation.
r12 If fix.rho = TRUE, the value that \(corr(W_1, W_2)\) is fixed to.
epsilon The precision criterion for EM convergence. \(\sqrt{\epsilon}\) is the precision criterion for SEM convergence.
theta.sem The ML estimates of \(E(W_1), E(W_2), var(W_1), var(W_2),\) and \(cov(W_1, W_2)\). If context = TRUE, \(E(X), cov(W_1, X), cov(W_2, X)\) are also reported.
W In-sample estimation of \(W_1\) and \(W_2\).
suff.stat The sufficient statistics for theta.em.
iters.em Number of EM iterations before convergence is achieved.
iters.sem Number of SEM iterations before convergence is achieved.
loglik The log-likelihood of the model when convergence is achieved.
loglik.log.em A vector saving the value of the log-likelihood function at each iteration of the EM algorithm.
mu.log.em A matrix saving the unweighted mean estimation of the logit-transformed individual-level proportions (i.e., \(W_1\) and \(W_2\)) at each iteration of the EM process.
Sigma.log.em A matrix saving the log of the variance estimation of the logit-transformed individual-level proportions (i.e., \(W_1\) and \(W_2\)) at each iteration of EM process. Note, non-transformed variances are displayed on the screen (when verbose = TRUE).
rho.fisher.em A matrix saving the fisher transformation of the estimation of the correlations between the logit-transformed individual-level proportions (i.e., \(W_1\) and \(W_2\)) at each iteration of EM process. Note, non-transformed correlations are displayed on the screen (when verbose = TRUE).

Moreover, when sem=TRUE, ecoML also output the following values:

DM The matrix characterizing the rates of convergence of the EM algorithms. Such information is also used to calculate the observed-data information matrix
Icom The (expected) complete data information matrix estimated via SEM algorithm. When context=FALSE, fix.rho=TRUE, Icom is 4 by 4. When context=FALSE, fix.rho=FALSE, Icom is 5 by 5. When context=TRUE, Icom is 9 by 9.
Iobs  The observed information matrix. The dimension of Iobs is same as Icom.
Imiss  The difference between Icom and Iobs. The dimension of Imiss is same as miss.
Vobs  The (symmetrized) variance-covariance matrix of the ML parameter estimates. The dimension of Vobs is same as Icom.
Iobs  The (expected) complete-data variance-covariance matrix. The dimension of Iobs is same as Icom.
Vobs.original  The estimated variance-covariance matrix of the ML parameter estimates. The dimension of Vobs is same as Icom.
Fmis  The fraction of missing information associated with each parameter estimation.
VFMis  The proportion of increased variance associated with each parameter estimation due to observed data.
Ieigen  The largest eigen value of Imiss.
Icom.trans  The complete data information matrix for the fisher transformed parameters.
Iobs.trans  The observed data information matrix for the fisher transformed parameters.
Fmis.trans  The fractions of missing information associated with the fisher transformed parameters.

Author(s)
Kosuke Imai, Department of Politics, Princeton University, <kimai@princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University, <ying.lu@nyu.edu>; Aaron Strauss, Department of Politics, Princeton University, <abstraus@princeton.Edu>.

References

See Also
eco, econp, summary.ecoML

Examples

## load the census data
data(census)

## NOTE: convergence has not been properly assessed for the following
ecoNP

Fitting the Nonparametric Bayesian Models of Ecological Inference in 2x2 Tables

Description

econoP is used to fit the nonparametric Bayesian model (based on a Dirichlet process prior) for ecological inference in $2 \times 2$ tables via Markov chain Monte Carlo. It gives the in-sample predictions as well as out-of-sample predictions for population inference. The models and algorithms are described in Imai, Lu and Strauss (2008, 2011).

Usage

econoP(formula, data = parent.frame(), N = NULL, supplement = NULL, context = FALSE, mu0 = 0, tau0 = 2, nu0 = 4, S0 = 10, alpha = NULL, a0 = 1, b0 = 0.1, parameter = FALSE, grid = FALSE, n.draws = 5000, burnin = 0, thin = 0, verbose = FALSE)
### Arguments

**formula**  
A symbolic description of the model to be fit, specifying the column and row margins of 2 × 2 ecological tables. \( Y \sim X \) specifies \( Y \) as the column margin (e.g., turnout) and \( X \) as the row margin (e.g., percent African-American). Details and specific examples are given below.

**data**  
An optional data frame in which to interpret the variables in `formula`. The default is the environment in which `ecoNP` is called.

**N**  
An optional variable representing the size of the unit; e.g., the total number of voters. \( N \) needs to be a vector of same length as \( Y \) and \( X \) or a scalar.

**supplement**  
An optional matrix of supplemental data. The matrix has two columns, which contain additional individual-level data such as survey data for \( W_1 \) and \( W_2 \), respectively. If `NULL`, no additional individual-level data are included in the model. The default is `NULL`.

**context**  
Logical. If `TRUE`, the contextual effect is also modeled, that is to assume the row margin \( X \) and the unknown \( W_1 \) and \( W_2 \) are correlated. See Imai, Lu and Strauss (2008, 2011) for details. The default is `FALSE`.

**mu0**  
A scalar or a numeric vector that specifies the prior mean for the mean parameter \( \mu \) of the base prior distribution \( G_0 \) (see Imai, Lu and Strauss (2008, 2011) for detailed descriptions of Dirichete prior and the normal base prior distribution). If it is a scalar, then its value will be repeated to yield a vector of the length of \( \mu \), otherwise, it needs to be a vector of same length as \( \mu \). When `context=TRUE`, the length of \( \mu \) is 3, otherwise it is 2. The default is `0`.

**tau0**  
A positive integer representing the scale parameter of the Normal-Inverse Wishart prior for the mean and variance parameter \( (\mu_i, \Sigma_i) \) of each observation. The default is 2.

**nu0**  
A positive integer representing the prior degrees of freedom of the variance matrix \( \Sigma_i \), the default is 4.

**S0**  
A positive scalar or a positive definite matrix that specifies the prior scale matrix for the variance matrix \( \Sigma_i \). If it is a scalar, then the prior scale matrix will be a diagonal matrix with the same dimensions as \( \Sigma_i \) and the diagonal elements all take value of \( S_0 \), otherwise \( S_0 \) needs to have same dimensions as \( \Sigma_i \). When `context=TRUE`, \( \Sigma \) is a 3 × 3 matrix, otherwise, it is 2 × 2. The default is `10`.

**alpha**  
A positive scalar representing a user-specified fixed value of the concentration parameter, \( \alpha \). If `NULL`, \( \alpha \) will be updated at each Gibbs draw, and its prior parameters \( a0 \) and \( b0 \) need to be specified. The default is `NULL`.

**a0**  
A positive integer representing the value of shape parameter of the gamma prior distribution for \( \alpha \). The default is 1.

**b0**  
A positive integer representing the value of scale parameter of the gamma prior distribution for \( \alpha \). The default is 0.1.

**parameter**  
Logical. If `TRUE`, the Gibbs draws of the population parameters, \( \mu \) and \( \Sigma \), are returned in addition to the in-sample predictions of the missing internal cells, \( W \). The default is `FALSE`. This needs to be set to `TRUE` if one wishes to make population inferences through `predict.eco`. See an example below.
grid Logical. If TRUE, the grid method is used to sample \( W \) in the Gibbs sampler. If FALSE, the Metropolis algorithm is used where candidate draws are sampled from the uniform distribution on the tomography line for each unit. Note that the grid method is significantly slower than the Metropolis algorithm.

n.draws A positive integer. The number of MCMC draws. The default is 5000.

burnin A positive integer. The burnin interval for the Markov chain; i.e. the number of initial draws that should not be stored. The default is 0.

thin A positive integer. The thinning interval for the Markov chain; i.e. the number of Gibbs draws between the recorded values that are skipped. The default is 0.

verbose Logical. If TRUE, the progress of the Gibbs sampler is printed to the screen. The default is FALSE.

Value

An object of class ecoNP containing the following elements:

call The matched call.
X The row margin, \( X \).
Y The column margin, \( Y \).
burnin The number of initial burnin draws.
thin The thinning interval.
u0 The prior degrees of freedom.
tau0 The prior scale parameter.
mu0 The prior mean.
S0 The prior scale matrix.
a0 The prior shape parameter.
b0 The prior scale parameter.
w A three dimensional array storing the posterior in-sample predictions of \( W \). The first dimension indexes the Monte Carlo draws, the second dimension indexes the columns of the table, and the third dimension represents the observations.
wmin A numeric matrix storing the lower bounds of \( W \).
wmax A numeric matrix storing the upper bounds of \( W \).

The following additional elements are included in the output when parameter = TRUE.

mu A three dimensional array storing the posterior draws of the population mean parameter, \( \mu \). The first dimension indexes the Monte Carlo draws, the second dimension indexes the columns of the table, and the third dimension represents the observations.

Sigma A three dimensional array storing the posterior draws of the population variance matrix, \( \Sigma \). The first dimension indexes the Monte Carlo draws, the second dimension indexes the parameters, and the third dimension represents the observations.

alpha The posterior draws of \( \alpha \).

nstar The number of clusters at each Gibbs draw.
Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.Lu@nyu.edu>

References


See Also

eco, ecoML, predict.eco, summary.ecoNP

Examples

## load the registration data
data(reg)

## NOTE: We set the number of MCMC draws to be a very small number in
## the following examples; i.e., convergence has not been properly
## assessed. See Imai, Lu and Strauss (2006) for more complete examples.

## fit the nonparametric model to give in-sample predictions
## store the parameters to make population inference later
## Not run: res <- econp(y ~ x, data = reg, n.draws = 50, param = TRUE, verbose = TRUE)

## summarize the results
summary(res)

## obtain out-of-sample prediction
out <- predict(res, verbose = TRUE)

## summarize the results
summary(out)

## density plots of the out-of-sample predictions
par(mfrow=c(2,1))
plot(density(out[,1]), main = "W1")
plot(density(out[,2]), main = "W2")

## load the Robinson's census data
data(census)
## fit the parametric model with contextual effects and N
## using the default prior specification

res1 <- econP(Y ~ X, N = N, context = TRUE, param = TRUE, data = census, 
n.drops = 25, verbose = TRUE)

## summarize the results
summary(res1)

## out-of-sample prediction
pres1 <- predict(res1)
summary(pres1)

## End (Not run)

---

### Foreign-born literacy in 1930

#### Description

This data set contains, on a state level, the proportion of white residents ten years and older who are foreign born, and the proportion of those residents who are literate. Data come from the 1930 census and were first analyzed by Robinson (1950).

#### Format

A data frame containing 5 variables and 48 observations

- **X** numeric: proportion of the white population at least 10 years of age that is foreign born
- **Y** numeric: proportion of the white population at least 10 years of age that is illiterate
- **W1** numeric: proportion of the foreign-born white population at least 10 years of age that is illiterate
- **W2** numeric: proportion of the native-born white population at least 10 years of age that is illiterate
- **ICPSR** numeric: the ICPSR state code

#### References


---

### Foreign-born literacy in 1930, County Level

#### Description

This data set contains, on a county level, the proportion of white residents ten years and older who are foreign born, and the proportion of those residents who are literate. Data come from the 1930 census and were first analyzed by Robinson (1950). Counties with fewer than 100 foreign born residents are dropped.
Format

A data frame containing 6 variables and 1976 observations

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>numeric</td>
<td>proportion of the white population at least 10 years of age that is foreign born</td>
</tr>
<tr>
<td>Y</td>
<td>numeric</td>
<td>proportion of the white population at least 10 years of age that is illiterate</td>
</tr>
<tr>
<td>W1</td>
<td>numeric</td>
<td>proportion of the foreign-born white population at least 10 years of age that is illiterate</td>
</tr>
<tr>
<td>W2</td>
<td>numeric</td>
<td>proportion of the native-born white population at least 10 years of age that is illiterate</td>
</tr>
<tr>
<td>state</td>
<td>numeric</td>
<td>the ICPSR state code</td>
</tr>
<tr>
<td>county</td>
<td>numeric</td>
<td>the ICPSR (within state) county code</td>
</tr>
</tbody>
</table>

References


**housep88**

Electoral Results for the House and Presidential Races in 1988

Description

This data set contains, on a House district level, the percentage of the vote for the Democratic House candidate, the percentage of the vote for the Democratic presidential candidate (Dukakis), the number of voters who voted for a major party candidate in the presidential race, and the ratio of voters in the House race versus the number who cast a ballot for President. Eleven (11) uncontested races are not included. Dataset compiled and analyzed by Burden and Kimball (1988). Complete dataset and documentation available at ICSPR study number 1140.

Format

A data frame containing 5 variables and 424 observations

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>numeric</td>
<td>proportion voting for the Democrat in the presidential race</td>
</tr>
<tr>
<td>Y</td>
<td>numeric</td>
<td>proportion voting for the Democrat in the House race</td>
</tr>
<tr>
<td>N</td>
<td>numeric</td>
<td>number of major party voters in the presidential contest</td>
</tr>
<tr>
<td>HPCT</td>
<td>numeric</td>
<td>House election turnout divided by presidential election turnout (set to 1 if House turnout exceeds presidential)</td>
</tr>
<tr>
<td>DIST</td>
<td>numeric</td>
<td>4-digit ICPSR state and district code: first 2 digits for the state code, last two digits for the district number</td>
</tr>
</tbody>
</table>

References


**predict.eco**

Out-of-Sample Posterior Prediction under the Parametric Bayesian Model for Ecological Inference in 2x2 Tables
Description

Obtains out-of-sample posterior predictions under the fitted parametric Bayesian model for ecological inference. `predict` method for class `eco` and `ecoX`.

Usage

```r
## S3 method for class 'eco'
predict(object, newdraw = NULL, subset = NULL, 
       verbose = FALSE, ...)
```

Arguments

- `object`: An output object from `eco` or `ecoNP`.
- `newdraw`: An optional list containing two matrices (or three dimensional arrays for the nonparametric model) of MCMC draws of $\mu$ and $\Sigma$. Those elements should be named as `mu` and `sigma`, respectively. The default is the original MCMC draws stored in `object`.
- `subset`: A scalar or numerical vector specifying the row number(s) of `mu` and `sigma` in the output object from `eco`. If specified, the posterior draws of parameters for those rows are used for posterior prediction. The default is `NULL` where all the posterior draws are used.
- `verbose`: logical. If `true`, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is `FALSE`.
- `...`: further arguments passed to or from other methods.

Details

The posterior predictive values are computed using the Monte Carlo sample stored in the `eco` output (or other sample if `newdraw` is specified). Given each Monte Carlo sample of the parameters, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, we apply the inverse logit transformation to obtain the predictive values of proportions, $W$. The computation may be slow (especially for the nonparametric model) if a large Monte Carlo sample of the model parameters is used. In either case, setting `verbose = TRUE` may be helpful in monitoring the progress of the code.

Value

`predict.eco` yields a matrix of class `predict.eco` containing the Monte Carlo sample from the posterior predictive distribution of inner cells of ecological tables. `summary.predict.eco` will summarize the output, and `print.summary.predict.eco` will print the summary.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.Edu>
See Also

eco, predict.ecoNP

---

**predict.ecoNP**

*Out-of-Sample Posterior Prediction under the Nonparametric Bayesian Model for Ecological Inference in 2x2 Tables*

**Description**

Obtains out-of-sample posterior predictions under the fitted nonparametric Bayesian model for ecological inference. `predict` method for class `ecoNP` and `ecoNPX`.

**Usage**

```r
## S3 method for class 'ecoNP'
predict(object, newdraw = NULL, subset = NULL, obs = NULL,
        verbose = FALSE, ...)
```

**Arguments**

- **object**: An output object from `ecoNP`.
- **newdraw**: An optional list containing two matrices (or three dimensional arrays for the nonparametric model) of MCMC draws of $\mu$ and $\Sigma$. Those elements should be named as `mu` and `sigma`, respectively. The default is the original MCMC draws stored in `object`.
- **subset**: A scalar or numerical vector specifying the row number(s) of `mu` and `sigma` in the output object from `eco`. If specified, the posterior draws of parameters for those rows are used for posterior prediction. The default is `NULL` where all the posterior draws are used.
- **obs**: An integer or vector of integers specifying the observation number(s) whose posterior draws will be used for predictions. The default is `NULL` where all the observations in the data set are selected.
- **verbose**: logical. If `TRUE`, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is `FALSE`.
- **...**: further arguments passed to or from other methods.

**Details**

The posterior predictive values are computed using the Monte Carlo sample stored in the `eco` or `ecoNP` output (or other sample if `newdraw` is specified). Given each Monte Carlo sample of the parameters, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, we apply the inverse logit transformation to obtain the predictive values of proportions, $W$. The computation may be slow (especially for the nonparametric model) if a large Monte Carlo sample of the model parameters is used. In either case, setting `verbose = TRUE` may be helpful in monitoring the progress of the code.
predict.ecoNPX

Value

predict.eco yields a matrix of class predict.eco containing the Monte Carlo sample from the posterior predictive distribution of inner cells of ecological tables. summary.predict.eco will summarize the output, and print.summary.predict.eco will print the summary.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

See Also

eco, ecoNP, summary.eco, summary.ecoNP

predict.ecoNPX  Out-of-Sample Posterior Prediction under the Nonparametric Bayesian Model for Ecological Inference in 2x2 Tables

Description

Obtains out-of-sample posterior predictions under the fitted nonparametric Bayesian model for ecological inference. predict method for class ecoNP and ecoNPX.

Usage

## S3 method for class 'ecoNPX'
predict(object, newdraw = NULL, subset = NULL, obs = NULL, cond = FALSE, verbose = FALSE, ...)

Arguments

- **object**: An output object from ecoNP.
- **newdraw**: An optional list containing two matrices (or three dimensional arrays for the nonparametric model) of MCMC draws of $\mu$ and $\Sigma$. Those elements should be named as mu and sigma, respectively. The default is the original MCMC draws stored in object.
- **subset**: A scalar or numerical vector specifying the row number(s) of mu and Sigma in the output object from eco. If specified, the posterior draws of parameters for those rows are used for posterior prediction. The default is NULL where all the posterior draws are used.
- **obs**: An integer or vector of integers specifying the observation number(s) whose posterior draws will be used for predictions. The default is NULL where all the observations in the data set are selected.
- **cond**: logical. If TRUE, then the conditional prediction will made for the parametric model with contextual effects. The default is FALSE.
verbose logical. If TRUE, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is FALSE.

... further arguments passed to or from other methods.

Details

The posterior predictive values are computed using the Monte Carlo sample stored in the eco or ecoNP output (or other sample if newdraw is specified). Given each Monte Carlo sample of the parameters, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, we apply the inverse logit transformation to obtain the predictive values of proportions, W. The computation may be slow (especially for the nonparametric model) if a large Monte Carlo sample of the model parameters is used. In either case, setting verbose = TRUE may be helpful in monitoring the progress of the code.

Value

predict.eco yields a matrix of class predict.eco containing the Monte Carlo sample from the posterior predictive distribution of inner cells of ecological tables. summary.predict.eco will summarize the output, and print.summary.predict.eco will print the summary.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@princeton.edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

See Also

eco, ecoNP, summary.eco, summary.ecoNP

predict.ecoX Out-of-Sample Posterior Prediction under the Parametric Bayesian Model for Ecological Inference in 2x2 Tables

Description

Obtains out-of-sample posterior predictions under the fitted parametric Bayesian model for ecological inference. predict method for class eco and ecoX.

Usage

```r
## S3 method for class 'ecoX'
predict(object, newdraw = NULL, subset = NULL,
    newdata = NULL, cond = FALSE, verbose = FALSE, ...)
```
predict.ecoX

Arguments

object An output object from eco or econp.
newdraw An optional list containing two matrices (or three dimensional arrays for the nonparametric model) of MCMC draws of $\mu$ and $\Sigma$. Those elements should be named as mu and Sigma, respectively. The default is the original MCMC draws stored in object.
subset A scalar or numerical vector specifying the row number(s) of mu and Sigma in the output object from eco. If specified, the posterior draws of parameters for those rows are used for posterior prediction. The default is NULL where all the posterior draws are used.
newdata An optional data frame containing a new data set for which posterior predictions will be made. The new data set must have the same variable names as those in the original data.
cond logical. If TRUE, then the conditional prediction will made for the parametric model with contextual effects. The default is FALSE.
verbose logical. If TRUE, helpful messages along with a progress report on the Monte Carlo sampling from the posterior predictive distributions are printed on the screen. The default is FALSE.

Details

The posterior predictive values are computed using the Monte Carlo sample stored in the eco output (or other sample if newdraw is specified). Given each Monte Carlo sample of the parameters, we sample the vector-valued latent variable from the appropriate multivariate Normal distribution. Then, we apply the inverse logit transformation to obtain the predictive values of proportions, $W$. The computation may be slow (especially for the nonparametric model) if a large Monte Carlo sample of the model parameters is used. In either case, setting verbose = TRUE may be helpful in monitoring the progress of the code.

Value

predict.eco yields a matrix of class predict.eco containing the Monte Carlo sample from the posterior predictive distribution of inner cells of ecological tables. summary.predict.eco will summarize the output, and print.summary.predict.eco will print the summary.

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

See Also
eco, predict.ecoNP
Description

summary method for class eco.

Usage

```
## S3 method for class 'summary.eco'
print(x, digits = max(3,getOption("digits") - 3), ...)
```

Arguments

- `x`: An object of class `summary.eco`.
- `digits`: the number of significant digits to use when printing.
- `...`: further arguments passed to or from other methods.

Value

`summary.eco` yields an object of class `summary.eco` containing the following elements:

- `call`: The call from `eco`.
- `n.obs`: The number of units.
- `n.draws`: The number of Monte Carlo samples.
- `agg.table`: Aggregate posterior estimates of the marginal means of $W_1$ and $W_2$ using $X$ and $N$ as weights.

If `param = TRUE`, the following elements are also included:

- `param.table`: Posterior estimates of model parameters: population mean estimates of $W_1$ and $W_2$ and their logit transformations.

If `units = TRUE`, the following elements are also included:

- `W1.table`: Unit-level posterior estimates for $W_1$.
- `W2.table`: Unit-level posterior estimates for $W_2$.

This object can be printed by `print.summary.eco`

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, [http://imai.princeton.edu](http://imai.princeton.edu); Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

See Also

eco, `predict.eco`
### Description

summary method for class eco.

### Usage

```r
## S3 method for class 'summary.ecoML'
print(x, digits = max(3,getOption("digits") - 3),
      ...)  
```

### Arguments

- `x`: An object of class `summary.ecoML`.
- `digits`: the number of significant digits to use when printing.
- `...`: further arguments passed to or from other methods.

### Value

`summary.eco` yields an object of class `summary.eco` containing the following elements:

- `call`: The call from `eco`.
- `sem`: Whether the SEM algorithm was executed, as specified by the user upon calling `ecoML`.
- `fix.rho`: Whether the correlation parameter was fixed or allowed to vary, as specified by the user upon calling `ecoML`.
- `epsilon`: The convergence threshold specified by the user upon calling `ecoML`.
- `n.obs`: The number of units.
- `iters.em`: The number iterations the EM algorithm cycled through before convergence or reaching the maximum number of iterations allowed.
- `iters.sem`: The number iterations the SEM algorithm cycled through before convergence or reaching the maximum number of iterations allowed.
- `loglik`: The final observed log-likelihood.
- `rho`: A matrix of `iters.em` rows specifying the correlation parameters at each iteration of the EM algorithm. The number of columns depends on how many correlation parameters exist in the model. Column order is the same as the order of the parameters in `param.table`.
- `param.table`: Final estimates of the parameter values for the model. Excludes parameters fixed by the user upon calling `ecoML`. See `ecoML` documentation for order of parameters.
- `agg.table`: Aggregate estimates of the marginal means of $W_1$ and $W_2$.
agg.wtable  Aggregate estimates of the marginal means of $W_1$ and $W_2$ using $X$ and $N$ as weights.

If units = TRUE, the following elements are also included:

w.table  Unit-level estimates for $W_1$ and $W_2$.

This object can be printed by print.summary.eco.

Author(s)
Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>; Aaron Strauss, Department of Politics, Princeton University, <abstraus@Princeton.Edu>

See Also
ecoML

print.summary.ecoNP  Print the Summary of the Results for the Bayesian Nonparametric Model for Ecological Inference in 2x2 Tables

Description
summary method for class ecoNP.

Usage
## S3 method for class 'summary.ecoNP'
print(x, digits = max(3, getOption("digits") - 3),
...)

Arguments
x  An object of class summary.ecoNP.
digits  the number of significant digits to use when printing.
...  further arguments passed to or from other methods.

Value
summary.ecoNP yields an object of class summary.ecoNP containing the following elements:
call  The call from ecoNP.
n.obs  The number of units.
n.draws  The number of Monte Carlo samples.
agg.table  Aggregate posterior estimates of the marginal means of \( W_1 \) and \( W_2 \) using \( X \) and \( N \) as weights.

If \( \text{param} = \text{TRUE} \), the following elements are also included:

param.table  Posterior estimates of model parameters: population mean estimates of \( W_1 \) and \( W_2 \). If subset is specified, only a subset of the population parameters are included.

If \( \text{unit} = \text{TRUE} \), the following elements are also included:

\( W_1 \).table  Unit-level posterior estimates for \( W_1 \).
\( W_2 \).table  Unit-level posterior estimates for \( W_2 \).

This object can be printed by \texttt{print.summary.ecoNP}

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@princeton.edu>, \url{http://imai.princeton.edu}; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

See Also

\texttt{ecoNP}, \texttt{predict.eco}

---

\textbf{Qfun}  \hspace{1cm} \textit{Fitting the Parametric Bayesian Model of Ecological Inference in 2x2 Tables}

Description

\texttt{Qfun} returns the complete log-likelihood that is used to calculate the fraction of missing information.

Usage

\texttt{Qfun(theta, suff.stat, n)}

Arguments

\texttt{theta}  A vector that contains the MLE \( E(W_1), E(W_2), \text{var}(W_1), \text{var}(W_2), \text{and} \text{cov}(W_1, W_2) \). Typically it is the element \texttt{theta.em} of an object of class \texttt{ecoML}.

\texttt{suff.stat}  A vector of sufficient statistics of \( E(W_1), E(W_2), \text{var}(W_1), \text{var}(W_2), \text{and} \text{cov}(W_1, W_2) \).

\texttt{n}  A integer representing the sample size.
**Author(s)**

Kosuke Imai, Department of Politics, Princeton University, <kimai@princeton.Edu>, [http://imai.princeton.edu](http://imai.princeton.edu); Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>; Aaron Strauss, Department of Politics, Princeton University, <abstraus@princeton.Edu>.

**References**


**See Also**

ecoML

---

**reg**

**Voter Registration in US Southern States**

**Description**

This data set contains the racial composition, the registration rate, the number of eligible voters as well as the actual observed racial registration rates for every county in four US southern states: Florida, Louisiana, North Carolina, and South Carolina.

**Format**

A data frame containing 5 variables and 275 observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>numeric</td>
<td>the fraction of Black voters</td>
</tr>
<tr>
<td>Y</td>
<td>numeric</td>
<td>the fraction of voters who registered themselves</td>
</tr>
<tr>
<td>N</td>
<td>numeric</td>
<td>the total number of voters in each county</td>
</tr>
<tr>
<td>W1</td>
<td>numeric</td>
<td>the actual fraction of Black voters who registered themselves</td>
</tr>
<tr>
<td>W2</td>
<td>numeric</td>
<td>the actual fraction of White voters who registered themselves</td>
</tr>
</tbody>
</table>

**References**

Summary

Summarizing the Results for the Bayesian Parametric Model for Ecological Inference in 2x2 Tables

Description

summary method for class eco.

Usage

```r
# S3 method for class 'eco'
summary(object, CI = c(2.5, 97.5), param = TRUE,
         units = FALSE, subset = NULL, ...)
```

Arguments

- `object`: An output object from eco.
- `CI`: A vector of lower and upper bounds for the Bayesian credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.
- `param`: Logical. If TRUE, the posterior estimates of the population parameters will be provided. The default value is TRUE.
- `units`: Logical. If TRUE, the in-sample predictions for each unit or for a subset of units will be provided. The default value is FALSE.
- `subset`: A numeric vector indicating the subset of the units whose in-sample predictions to be provided when units is TRUE. The default value is NULL where the in-sample predictions for each unit will be provided.
- `...`: further arguments passed to or from other methods.

Value

summary.eco yields an object of class summary.eco containing the following elements:

- `call`: The call from eco.
- `n.obs`: The number of units.
- `n.draws`: The number of Monte Carlo samples.
- `agg.table`: Aggregate posterior estimates of the marginal means of W1 and W2 using X and N as weights.

If param = TRUE, the following elements are also included:

- `param.table`: Posterior estimates of model parameters: population mean estimates of W1 and W2 and their logit transformations.

If units = TRUE, the following elements are also included:

- `W1.table`: Unit-level posterior estimates for W1.
- `W2.table`: Unit-level posterior estimates for W2.

This object can be printed by print.summary.eco
Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@princeton.edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>

See Also

eco, predict.eco

summary.ecoML

Summarizing the Results for the Maximum Likelihood Parametric Model for Ecological Inference in 2x2 Tables

Description

summary method for class eco.

Usage

## S3 method for class 'ecoML'
summary(object, CI = c(2.5, 97.5), param = TRUE,
         units = FALSE, subset = NULL, ...)

Arguments

object

An output object from eco.

CI

A vector of lower and upper bounds for the Bayesian credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.

param

Ignored.

units

Logical. If TRUE, the in-sample predictions for each unit or for a subset of units will be provided. The default value is FALSE.

subset

A numeric vector indicating the subset of the units whose in-sample predictions to be provided when units is TRUE. The default value is NULL where the in-sample predictions for each unit will be provided.

...

further arguments passed to or from other methods.

Value

summary.eco yields an object of class summary.eco containing the following elements:

call

The call from eco.

sem

Whether the SEM algorithm was executed, as specified by the user upon calling ecoML.

fix.rho

Whether the correlation parameter was fixed or allowed to vary, as specified by the user upon calling ecoML.
summary.ecoNP

epsilon The convergence threshold specified by the user upon calling ecoML.
n.obs The number of units.

iters.em The number iterations the EM algorithm cycled through before convergence or reaching the maximum number of iterations allowed.

iters.sem The number iterations the SEM algorithm cycled through before convergence or reaching the maximum number of iterations allowed.

loglik The final observed log-likelihood.

rho A matrix of iters.em rows specifying the correlation parameters at each iteration of the EM algorithm. The number of columns depends on how many correlation parameters exist in the model. Column order is the same as the order of the parameters in param.table.

param.table Final estimates of the parameter values for the model. Excludes parameters fixed by the user upon calling ecoML. See ecoML documentation for order of parameters.

agg.table Aggregate estimates of the marginal means of $W_1$ and $W_2$

agg.wtable Aggregate estimates of the marginal means of $W_1$ and $W_2$ using $X$ and $N$ as weights.

If units = TRUE, the following elements are also included:

w.table Unit-level estimates for $W_1$ and $W_2$.

This object can be printed by print.summary.eco

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@Princeton.Edu>, http://imai.princeton.edu; Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>; Aaron Strauss, Department of Politics, Princeton University, <abstraus@Princeton.Edu>

See Also

ecoML

---

**summary.ecoNP**

*Summarizing the Results for the Bayesian Nonparametric Model for Ecological Inference in 2x2 Tables*

**Description**

summary method for class ecoNP.
Usage

```r
## S3 method for class 'ecoNP'
summary(object, CI = c(2.5, 97.5), param = FALSE,
    units = FALSE, subset = NULL, ...)
```

Arguments

- `object`: An output object from `ecoNP`.
- `CI`: A vector of lower and upper bounds for the Bayesian credible intervals used to summarize the results. The default is the equal tail 95 percent credible interval.
- `param`: Logical. If TRUE, the posterior estimates of the population parameters will be provided. The default value is FALSE.
- `units`: Logical. If TRUE, the in-sample predictions for each unit or for a subset of units will be provided. The default value is FALSE.
- `subset`: A numeric vector indicating the subset of the units whose in-sample predictions to be provided when units is TRUE. The default value is NULL where the in-sample predictions for each unit will be provided.
- `...`: further arguments passed to or from other methods.

Value

`summary.ecoNP` yields an object of class `summary.ecoNP` containing the following elements:

- `call`: The call from `ecoNP`.
- `n.obs`: The number of units.
- `n.draws`: The number of Monte Carlo samples.
- `agg.table`: Aggregate posterior estimates of the marginal means of \( W_1 \) and \( W_2 \) using \( X \) and \( N \) as weights.

If `param = TRUE`, the following elements are also included:

- `param.table`: Posterior estimates of model parameters: population mean estimates of \( W_1 \) and \( W_2 \). If subset is specified, only a subset of the population parameters are included.

If `unit = TRUE`, the following elements are also included:

- `W1.table`: Unit-level posterior estimates for \( W_1 \).
- `W2.table`: Unit-level posterior estimates for \( W_2 \).

This object can be printed by `print.summary.ecoNP`

Author(s)

Kosuke Imai, Department of Politics, Princeton University, <kimai@princeton.edu>, [http://imai.princeton.edu](http://imai.princeton.edu); Ying Lu, Center for Promoting Research Involving Innovative Statistical Methodology (PRIISM), New York University <ying.lu@nyu.edu>
Description

This data set contains, on a county level, the proportion of county residents who are Black and the proportion of presidential votes cast for Wallace. Demographic data is based on the 1960 census. Presidential returns are from ICPSR study 13. County data from 10 southern states (Alabama, Arkansas, Georgia, Florida, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas) are included. (Virginia is excluded due to the difficulty of matching counties between the datasets.) This data is analyzed in Wallace and Segal (1973).

Format

A data frame containing 3 variables and 1009 observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>numeric</td>
<td>proportion of the population that is Black</td>
</tr>
<tr>
<td>Y</td>
<td>numeric</td>
<td>proportion presidential votes cast for Wallace</td>
</tr>
<tr>
<td>FIPS</td>
<td>numeric</td>
<td>the FIPS county code</td>
</tr>
</tbody>
</table>

References

Index

*Topic **datasets**
census, 2
fornlilt30, 17
fornlilt30c, 17
housep88, 18
reg, 28
wallace, 33

*Topic **methods**
predict.eco, 18
predict.ecoNP, 20
predict.ecoNPX, 21
predict.ecoX, 22
print.summary.eco, 24
print.summary.ecoML, 25
print.summary.ecoNP, 26
summary.eco, 29
summary.ecoML, 30
summary.ecoNP, 31

*Topic **models**
eco, 3
ecoBD, 6
ecoML, 9
ecoNP, 13
Qfun, 27
census, 2
eco, 3
ecoBD, 6
ecoML, 9
ecoNP, 13
fornlilt30, 17
fornlilt30c, 17

housep88, 18
predict.eco, 18
predict.ecoNP, 20
predict.ecoNPX, 21
predict.ecoX, 22
print.eco(summary.eco), 29
print.summary.eco, 24
print.summary.ecoML, 25
print.summary.ecoNP, 26
Qfun, 27
reg, 28
summary.eco, 29
summary.ecoML, 30
summary.ecoNP, 31
wallace, 33