Package ‘bnpMTP’

September 24, 2025

Title Bayesian Nonparametric Sensitivity Analysis of Multiple Testing Procedures for p Values

Version 1.0.0

Author George Karabatsos [aut, cre] (ORCID: <https://orcid.org/0000-0003-4243-2285>)

Maintainer George Karabatsos <gkarabatsos1@gmail.com>

Description Bayesian Nonparametric sensitivity analysis of multiple testing procedures for p values with arbitrary dependencies, based on the Dirichlet process prior distribution.

License GPL (>= 2)

Imports stats

Encoding UTF-8

Repository CRAN

RoxygenNote 7.3.3

NeedsCompilation no

Date/Publication 2025-09-24 08:30:07 UTC

Contents

---

bnpMTP *Bayesian Nonparametric Sensitivity Analyses of Multiple Testing Procedures for p-values with Arbitrary Dependence (Intercorrelations)*

Description

Given inputs of *p*-values p from m = length(p) hypothesis tests and their error rates alpha, this R package function bnpMTP() performs sensitivity analysis and uncertainty quantification for Multiple Testing Procedures (MTPs) based on a mixture of Dirichlet process (DP) prior distribution (Ferguson, 1973) supporting all MTPs providing Family-wise Error Rate (FWER) or False Discovery Rate (FDR) control for *p*-values with arbitrary dependencies, e.g., due to tests performed on shared data and/or correlated variables, etc. From such an analysis, bnpMTP() outputs the distribution of the number of significant *p*-values (discoveries); and a *p*-value from a global joint test of all m null hypotheses, based on the probability of significance (discovery) for each *p*-value.

The DP-MTP sensitivity analysis method can analyze a large number of *p*-values obtained from any mix of null hypothesis testing procedures, including one-sample and/or multi-sample tests of: location, scale, higher moments, distribution, or symmetry; correlation, association, regression coefficients, odds ratios; change-points; runs; networks; classification; clustering; posterior distributions; model fit; outlyingness; and/or continuous hypothesis tests (e.g., performed on a realization of a random field); among other tests. Also, this sensitivity analysis method handles *p*-values from traditional offline testing; and from *online testing* performed on a stream of null hypotheses arriving one-by-one (or in blocks) over time (or asynchronously), where each test is based only on previous fixed test decisions and evidence against the current hypothesis, with unknown future data and total number of hypotheses being tested (potentially infinite) (Robertson, 2023).

In any case, the DP-MTP sensitivity analysis method assumes that each *p*-value follows a super-uniform distribution under the null hypothesis (i.e., either a Uniform(0,1) distribution under a calibrated test; or a stochastically larger distribution under a conservative test). More **Details** about this method are below (run bnpMTP in R console to view code) and provided by Karabatsos (2025).

Usage

```
bnpMTP( p = NULL , alpha = 0.05 , N = 1000 , mu = 1 )
```

Arguments

p A vector of *p*-values. They can have arbitrary dependence.

alpha Scalar (default alpha = 0.05) or vector of Type I error probabilities specifying the error rate to be spent or invested (Foster & Stine, 2008) on each of the m = length(p) null hypothesis tests, with total error rate sum(alpha) such as sum(alpha) = 0.05 (which can be less than the desired total error rate if input p excludes *p*-values from future online tests to be done later), as:
"Once we have spent this (total) error rate, it is gone" (Tukey, 1991, p.104).
Input alpha helps define the random significance *thresholds*, by:
Delta_nu(r) = alpha * beta_nu(r),
for each ordered *p*-value p(r) from sort(p), based on a random probability measure nu on [0, m] from a mixture of Dirichlet process, and on a positive reshaping parameter beta_nu(.), used for MTP sensitivity analysis (see **Details**).

- bnpMTP() converts any scalar input alpha into the following vector input: alpha <- rep(alpha * (1 / m), m) = alpha * w, which defines the thresholds of the *Bonferroni* (1936) MTP for m = length(p) = length(w) tests, where each *p*-value p[i] is assigned a *standard weight* w[i] = 1 / m.
- Input alpha can be specified as a vector: alpha = alpha0 * w for some small positive number alpha0 (e.g., alpha0 = 0.05), which defines the

significance thresholds alpha of the *weighted Bonferroni* MTP, based on a prior distribution vector w representing the degree of belief for each of the m = length(p) null hypotheses (Genovese et al.2006), where sum(w) = 1 (Tamhane & Gou 2022), or sum(w) < 1 if p excludes *p*-values from future online tests to be done later (Tian & Ramdas, 2021, Section 2).

- Some alternatives for vector input alpha are defined by: the *Šidák* (1967), *Fallback* (Wiens & Dmitrienko, 2005), and *Adaptive Discarding* MTPs for offline or online FWER control (Tian & Ramdas, 2021); and *LORD* (Javanmard & Montanari 2018) and other *generalized alpha investing* methods (Aharoni & Rosset, 2014) for online FDR control (Robertson et al. 2023).

N, mu Number of random samples drawn from the mixture of Dirichlet process (DP(M, nu_0)) prior distribution for the random probability measure nu defined on [0, m], with mass parameter M assigned an Exponential(mu) hyper-prior distribution with rate mu, where m = length(p). Defaults: N = 1000 and mu = 1.

Details

The Dirichlet process (DP) based MTP sensitivity analysis method (Karabatsos, 2025) assigns a mixture of DP(M, nu_0) prior distribution that flexibly supports the entire space of random probability measures nu defined on the interval [0, m] for m = length(p) hypothesis tests, with Exponential(mu) hyper-prior distribution assigned to the DP mass parameter M, and with (mean) baseline probability measure (nu_0) defined by the Benjamini & Yekutieli (2001) MTP. In turn, this mixture DP prior also supports the space of all MTPs providing FWER or FDR control for *p* values with arbitrary dependencies, because each of these MTPs can be uniquely characterized by a random probability measure nu, based on the *shape function approach* to multiple hypothesis testing (Blanchard & Roquain, 2008, Sections 3.1-3.2; Lemma 3.2, Equation 6, pp.970–972, 976).

Specifically, the DP random probability measure, nu, drives the random number, r.hat_nu, of the smallest *p*-values (from input p with length(p) = m) that are significant discoveries, defined via the following DP random *step-up procedure* (using inequality <=):
r.hat_nu = max[r \in {0,1,...,m} | p_(r) <= alpha * beta_nu(r)],

where for r = 0,1,...,m, the p_(r) (with p_(0) := 0) are the ordered *p*-values (sort(p)) sorted in increasing order, with values of random significance *thresholds*:
Delta_nu(r) = alpha * beta_nu(r),
based on a random *shape function*:
beta_nu(r) = integral_0^r x d{nu(x)}
which *reshapes* (Ramdas et al. 2019, pp.2795-2796) or modifies alpha into new significance thresholds Delta_nu(r) to further account for arbitrary dependencies between *p*-values.

Further details are provided by Karabatsos (2025), who illustrated this DP-MTP sensitivity analysis method on over twenty-eight thousand *p*-values of different hypothesis tests performed on observations of 239 variables from a large dataset.

Value

Output of the DP-MTP sensitivity analysis results, as a list containing the following objects:

r.hat_nu A vector of N samples of the number, r.hat_nu, of the smallest *p*-values (from input p) that are significant discoveries, based on N samples of the random probability measure `nu` defined on `[0, m]` for `m` = `length(p)` hypothesis tests, with `nu` assigned the mixture of DP prior distribution.

`Delta_nu.r`, `beta_nu.r`

Two `N`-by-`(m + 1)` matrices of `N` mixture of DP samples of the threshold function and the shape function for the sorted *p*-values (`sort(p)`) in `colnames(Delta_nu.r)` and `colnames(beta_nu.r)`, respectively. Using `I` = `cbind(1:N,r.hat_nu+1)`, the `N` samples of threshold `Delta_nu(r.hat_nu)` and shape `beta_nu(r.hat_nu)` are obtained from `Delta_nu.r[I]` and `beta_nu.r[I]`.

`Table`

A `3`-by-`(m + 1)` matrix reporting the probability of significance (`PrSig.p`) for each of the `m` = `length(p)` total *p*-values in input `p` with respective error rate(s) `alpha`, based on the mixture DP prior.
For each ordered *p*-value `p_(r)` from `sort(p)`, the *probability of significance* is estimated by the proportion of `N` samples of `r.hat_nu` which satisfy inequality: `p_(r)` <= `Delta_nu(r.hat_nu)`, for `r` = `1`, `...`, `m = length(p)`.
The last column of the output `Table` shows the prior predictive *p*-value from the global joint test that all `m` = `length(p)` null hypotheses are true; and their total spent error rate `sum(alpha)` and their `max(PrSig.p)`. This *p*-value equals:
`min(1` - `PrSig.p)` = `1` - `max(PrSig.p)`
`=  mean(r.hat_nu == 0) = 1 - mean(r.hat_nu > 0)`
based on the idea that the joint null hypothesis should be rejected if at least one of the `m` null hypotheses is rejected (Simes, 1986).

References

Aharoni, E., and Rosset, S. (2014). Generalized alpha-investing: Definitions, optimality results and application to public databases. *Journal of the Royal Statistical Society Series B*, **76**, 771–794. https://www.jstor.org/stable/24774568

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188. https://www.jstor.org/stable/2674075

Blanchard, G., and Roquain, E. (2008). Two simple sufficient conditions for FDR control. *Electronic Journal of Statistics*, **2**, 963–992. DOI: 10.1214/08-EJS180

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, **8**, 3–62.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230. https://www.jstor.org/stable/2958008

Foster, D., and Stine, R. (2008). Alpha-investing: A procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society, Series B*, **70**, 429–444. https://www.jstor.org/stable/20203833

Genovese, C., Roeder, K., and Wasserman, L. (2006). False discovery control with *p*-value weighting, *Biometrika*, **93**, 509–524. https://www.jstor.org/stable/20441304

Javanmard, A., and Montanari, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. *Annals of Statistics*, **46**, 526–554. https://www.jstor.org/stable/26542797

Karabatsos, G. (2025). Bayesian nonparametric sensitivity analysis of multiple test procedures under dependence. *Biometrical Journal*. https://arxiv.org/abs/2410.08080. Paper presented in the *14th International Conference on Bayesian Nonparametrics* at UCLA on June 26, 2025.

Needleman, H., Gunnoe, C., Leviton, A., Reed, R., Presie, H., Maher, C., and Barret, P. (1979). Deficits in psychologic and classroom performance of children with elevated dentine lead levels. *New England Journal of Medicine*, **300**, 689–695. https://www.nejm.org/doi/10.1056/NEJM197903293001301

Ramdas, A., Barber, R., Wainwright, M., and Jordan, M. (2019). A unified treatment of multiple testing with prior knowledge using the *p*-filter. *Annals of Statistics*, **47**, 2790–2821. https://www.jstor.org/stable/26784046

Robertson, D., Wason, J., and Ramdas, A. (2023). Online multiple hypothesis testing. *Statistical Science*, **38**, 557–575. https://pmc.ncbi.nlm.nih.gov/articles/PMC7615519/

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, **62**, 626–633. https://www.jstor.org/stable/2283989

Simes, R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751–754. https://www.jstor.org/stable/2336545

Tamhane, A., and Gou, J. (2022). Multiple test procedures based on *p*-values. Chapter 2 of *Handbook of Multiple Comparisons*, by X. Cui, T. Dickhaus, Y. Ding, and J. Hsu (Eds.). CRC Press.

Tian, J., and Ramdas, A. (2021). Online control of the familywise error rate. *Statistical Methods in Medical Research*, **30**, 976–993. https://pubmed.ncbi.nlm.nih.gov/33413033/

Tukey, J. (1991). The philosophy of multiple comparisons. *Statistical Science*, **6**, 100–116. https://www.jstor.org/stable/2245714

Wiens, B., and Dmitrienko, A. (2005). The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics*, **15**, 929–942. https://www.tandfonline.com/doi/full/10.1080/10543400500265660

Examples

```
#---------------------------------------------------------------------------------------
# Consider a classic data set in the field of multiple hypothesis testing procedures.
# Needleman (1979,Table 3) from yes/no responses to 11 Teachers' Behavioral survey items
# compared 58 children exposed to high lead and 100 children exposed to low lead levels;
# by p-values from 11 chi-square null hypothesis tests of equal group % 'yes' responses;
# and a 2-tail p-value (0.02) from ANCOVA F-test of null hypothesis of equal group means
# in total sum score on the 11 items, while controlling for mother age at child's birth,
# number of pregnancies & educational level; father's socioeconomic status; parental IQ.
#---------------------------------------------------------------------------------------

# Below, enter the vector of twelve p-values (and then run this R code line, below):
p        = c(0.003, 0.05, 0.05, 0.14, 0.08, 0.01, 0.04, 0.01, 0.05, 0.003, 0.003, 0.02)

# Below, name these p-values (then run the three R code lines, below):
names(p) = c( "Distractible"  , "Impersistent"  , "Dependent"   , "Disorganized" ,
              "Hyperactive"   , "Impulsive"     , "Frustrated"  , "Daydreamer"   ,
              "MissEzDirect"  , "MissSeqDirect" , "LowFunction" , "SumScore"     )
```

```
# Get results of DP-MTP sensitivity analysis of the p-values: (Run 2 code lines, below):
set.seed(123) # for reproducibility of results of Monte Carlo sampling done by bnpMTP()
Result   = bnpMTP( p = p , alpha = 0.05 )

# Show probability of significance for each of m = length(p) = 12 p-values in input 'p'
# based on mixture of DP(M, nu_0) prior; and prior predictive p-value from global test
# of all 12 null hypotheses, and their total error sum(alpha) (run R code line below):
Result$Table

# Summarize mixture of DP(M, nu_0) prior distribution of number of significant p-values:
quantile( Result$r.hat_nu )

#-----------------------------------------------------------------------------------------
# Now suppose that the p-values were obtained from an online stream of hypothesis tests,
# with more hypothesis tests to be performed in the future (possibly infinite).
# Accordingly, we specify the alpha vector based on p-value weights (w) defined
# by the geometric distribution on {1,2,...} with 'success' probability 0.35,
# with sum(w) < 1 over the currently available twelve p-values in input p.
#-----------------------------------------------------------------------------------------

# Get results of DP-MTP sensitivity analysis of the p-values: (Run 5 code lines, below):
alpha0   = 0.05
w        = dgeom( ( 1 : length(p) ) - 1 , prob = 0.35 ) # specify p-value weights.
alpha    = alpha0 * w
set.seed(123) # for reproducibility of results of Monte Carlo sampling done by bnpMTP()
Online   = bnpMTP( p = p , alpha = alpha )

# Show probability of significance for each of m = length(p) = 12 p-values in input 'p'
# based on mixture of DP(M, nu_0) prior; and prior predictive p-value from global test
# of the 12 null hypotheses so far and their total error sum(alpha) (run line below):
Online$Table

# Summarize mixture of DP(M, nu_0) prior distribution of number of significant p-values:
quantile( Online$r.hat_nu )
```

Index