

Package: DescriptiveRepresentationCalculator (via r-universe)

October 8, 2026

Title Characterizing Observed and Expected Representation

Version 1.2.0

Description A system for analyzing descriptive representation, especially for comparing the composition of a political body to the population it represents. Users can compute the expected degree of representation for a body under a random sampling model, the expected degree of representation variability, as well as representation scores from observed political bodies. The package also provides district-level benchmarks for expected representation in single-member district systems, where district compositions follow a Dirichlet distribution and seats are filled by random sampling or largest-group-wins affinity voting, as well as group-level representation diagnostics and comparisons between bodies. The package is based on Gerring, Jerzak, and Oncel (2024) <doi:10.1017/S0003055423000680>.

URL <https://github.com/cjerzak/DescriptiveRepresentationCalculator-software/>

BugReports <https://github.com/cjerzak/DescriptiveRepresentationCalculator-software/issues>

Depends R (>= 3.3.3)

License GPL-3

Encoding UTF-8

LazyData false

Imports stats

Suggests knitr, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

Repository <https://cjerzak.r-universe.dev>

Date/Publication 2026-07-10 02:25:52 UTC
RemoteUrl <https://github.com/cjerzak/descriptiverepresentationcalculator-software>
RemoteRef HEAD
RemoteSha 50331945048488f308c318edc37b205e7e9d7de7
RemoteSubdir DescriptiveRepresentationCalculator

Contents

CompareRepresentation	2
ExpectedDistrictRepresentation	3
ExpectedRepresentation	6
GroupRepresentation	7
ObservedRepresentation	8
RelativeRepresentation	10
SDRepresentation	11
Index	13

CompareRepresentation	<i>Compare the observed representation of two political bodies</i>
-----------------------	--

Description

Calculates the difference in observed representation between two political bodies (or two tiers of the same legislature) measured against the same population shares. This is the tier-comparison statistic used in Gerrig, Hicken, Jerzak, Moser, and Oncel (book manuscript) to compare multi-member and single-member district tiers of mixed electoral systems.

Usage

```
CompareRepresentation(BodyMemberCharacteristics1, BodyMemberCharacteristics2,  
                      PopShares, BodyShares1, BodyShares2,  
                      a = -0.5, b = 1, metric = c("L1", "L2"))
```

Arguments

- BodyMemberCharacteristics1, BodyMemberCharacteristics2
Vectors specifying the characteristics for members of the first and second political body.
- PopShares
A numeric vector specifying population shares of identities specified in the body-member characteristics inputs. The names of the entries in PopShares should correspond to identities in those inputs.
- BodyShares1, BodyShares2
(optional) Numeric vectors with same structure as PopShares specifying group shares of the first and second body. If supplied, they take precedence over the corresponding body-member characteristics input.

a, b	Parameters controlling the affine transformation for how the representation measure is summarized, passed to ObservedRepresentation. By default, $a=-0.5$ and $b=1$ so that each body is scored with the Rose Index of Proportionality when <code>metric = "L1"</code> .
metric	A character string selecting the deviation metric underlying the representation index ("L1", the default, or "L2"), passed to ObservedRepresentation.

Value

A scalar giving the observed representation of the first body minus that of the second body. Positive values indicate that the first body more closely mirrors the population; negative values indicate that the second body does.

References

- John Gerring, Connor T. Jerzak, Erzen Oncel. (2024), The Composition of Descriptive Representation, *American Political Science Review*, 118(2): 784-801. doi:10.1017/S0003055423000680
- John Gerring, Allen Hicken, Connor T. Jerzak, Robert G. Moser, Erzen Oncel. Electoral Rules and Descriptive Representation: A Comprehensive View Across Multiple Identities. Book manuscript.

See Also

- [ObservedRepresentation](#) for the per-body representation index.
- [RelativeRepresentation](#) for comparing one body to its random-sampling expectation.

Examples

```
# Compare an MMD tier to an SMD tier of the same legislature
CompRep <- CompareRepresentation(
  BodyMemberCharacteristics1 = c("A", "B", "A", "B", "A", "A"),
  BodyMemberCharacteristics2 = c("A", "A", "A", "A", "A", "B"),
  PopShares = c("A" = 0.55, "B" = 0.45))

print( CompRep )
```

ExpectedDistrictRepresentation

Compute the expected degree of representation in a district-based electoral body

Description

Finds the expected degree of representation for a political body elected from single-member districts, where district-level group compositions are drawn from a Dirichlet distribution, as derived in Gerring, Hicken, Jerzak, Moser, and Oncel (book manuscript on electoral rules and descriptive representation).

Usage

```
ExpectedDistrictRepresentation(PopShares, concentration, nDistricts,
                              selectionRule = c("random", "affinity"),
                              a = -0.5, b = 1, nMonte = 10000)
```

Arguments

PopShares	A numeric vector containing the group-level population proportions (the mean district shares). Under <code>selectionRule = "affinity"</code> , all entries must be strictly positive; drop zero-share groups before calling.
concentration	A positive number giving the Dirichlet concentration parameter (<code>alpha_0</code>) governing the spatial concentration of groups across districts. District compositions are drawn from <code>Dirichlet(concentration * PopShares)</code> . Small values produce extreme district compositions (high geographic segregation); large values produce district compositions close to <code>PopShares</code> (geographic dispersion).
nDistricts	A positive integer denoting the number of single-member districts, or <code>Inf</code> for the large-district limit.
selectionRule	A character string specifying how each district seat is filled. <code>"random"</code> (the default): the probability that a group wins a district seat equals its population share in that district (random sampling). <code>"affinity"</code> : the largest group in each district deterministically captures the seat (largest-group-wins affinity voting).
a, b	Parameters controlling the affine transformation for how the representation measure is summarized. The expected sum of squared deviations between aggregate population shares and seat shares is multiplied by <code>a</code> ; <code>b</code> is an additive re-scaling term: $a \cdot E[\text{sum of squared deviations}] + b$. By default, $a = -0.5$ and $b = 1$ so that the expected squared-deviation representation index is returned.
nMonte	A positive integer denoting the number of Monte Carlo draws used to approximate the plurality-win probabilities when <code>selectionRule = "affinity"</code> and more than two groups are present. Ignored otherwise (those cases are computed in closed form).

Details

This function computes the expected value of the squared-deviation representation index (the metric = "L2" index; see [ObservedRepresentation](#)), not the L1-based Rose Index of Proportionality. The squared-deviation index upper-bounds the Rose Index and preserves the ordering of representation gaps when there are two groups.

Under `selectionRule = "random"`, the expectation has the closed form $1 - (1/(2 \cdot n\text{Districts})) \cdot \sum (\alpha \cdot (\alpha_0 - \alpha) / (\alpha_0 \cdot (\alpha_0 + 1)))$ (with default `a, b`), where $\alpha = \text{concentration} \cdot \text{PopShares}$ and $\alpha_0 = \text{concentration}$. As `nDistricts` grows, the expectation approaches 1: with many districts, random-sampling representation converges to the proportional benchmark. Setting `nDistricts = Inf` returns this limit.

Under `selectionRule = "affinity"`, the expectation depends on each group's plurality-win probability (the probability it is the largest group in a district) and, for finite `nDistricts`, on the upper-tail moment of the district-share distribution. With two groups, both quantities have closed forms in terms of the Beta distribution function and results are exact. With more than two groups, they are

approximated by Monte Carlo simulation of Dirichlet draws (set a seed for reproducibility). Setting `nDistricts = Inf` returns the large-district limit, $1 - (1/2) * \sum((\text{PopShares} - \text{PluralityProbs})^2)$ (with default `a, b`), which equals 1 only when each group's probability of being the district plurality matches its population share. The gap between these two quantities captures the representational consequences of geographic concentration under affinity voting.

Value

The expected degree of representation (a scalar).

References

- John Gerring, Connor T. Jerzak, Erzen Oncel. (2024), The Composition of Descriptive Representation, *American Political Science Review*, 118(2): 784-801. doi:[10.1017/S0003055423000680](https://doi.org/10.1017/S0003055423000680)
- John Gerring, Allen Hicken, Connor T. Jerzak, Robert G. Moser, Erzen Oncel. Electoral Rules and Descriptive Representation: A Comprehensive View Across Multiple Identities. Book manuscript.

See Also

- [ExpectedRepresentation](#) for expected representation of a single body under random sampling.
- [ObservedRepresentation](#) for calculating representation scores from observed data.

Examples

```
# Expected representation under random sampling with 100 districts
ExpectedDistrictRep <- ExpectedDistrictRepresentation(
  PopShares = c(1/4, 2/4, 1/4),
  concentration = 2,
  nDistricts = 100)
print( ExpectedDistrictRep )

# Large-district limit under largest-group-wins affinity voting (two groups;
# computed exactly via the Beta distribution function)
ExpectedAffinityRep <- ExpectedDistrictRepresentation(
  PopShares = c(0.3, 0.7),
  concentration = 4,
  nDistricts = Inf,
  selectionRule = "affinity")
print( ExpectedAffinityRep )
```

ExpectedRepresentation	<i>Compute the expected degree of representation for any group in a political body</i>
------------------------	--

Description

Finds the degree of expected representation for any group in a political body under a random sampling model as described in Gerring, Jerzak and Oncel (2024).

Usage

ExpectedRepresentation(PopShares, BodyN, a = -0.5, b = 1, metric = c("L1", "L2"))

Arguments

PopShares	A numeric vector containing the group-level population proportions.
BodyN	A positive integer denoting the size of the political body in question.
a, b	The a and b parameters control the affine transformation for how the representation measure is summarized. That is, a and b control how the expected L1 deviation of the population shares from the body shares is re-weighted. The expected L1 deviation is the average value of the absolute deviation of the population from body shares under a random sampling model. This expected L1 deviation is multiplied by a; b is as an additive re-scaling term: $a \cdot E[L1] + b$. By default, $a = -0.5$ and $b = 1$ so that the expected Rose Index of Proportionality is returned.
metric	A character string selecting the deviation metric underlying the representation index. "L1" (the default) uses the expected sum of absolute deviations between population and body shares, yielding the expected Rose Index of Proportionality under the default a and b. "L2" uses the expected sum of squared deviations, yielding the expected squared-deviation representation index analyzed in Gerring, Hicken, Jerzak, Moser, and Oncel (book manuscript). Under multinomial random sampling, the expected sum of squared deviations has the exact closed form $\text{sum}(\text{PopShares} \cdot (1 - \text{PopShares})) / \text{BodyN}$.

Value

The expected degree of representation (a scalar).

References

- John Gerring, Connor T. Jerzak, Erzen Oncel. (2024), The Composition of Descriptive Representation, *American Political Science Review*, 118(2): 784-801. doi:10.1017/S0003055423000680

See Also

- [ObservedRepresentation](#) for calculating representation scores from observed data.
- [SDRepresentation](#) for calculating representation unexplained under the random sampling model.
- [ExpectedDistrictRepresentation](#) for expected representation in district-based electoral systems.

Examples

```
ExpectedRep <- ExpectedRepresentation(PopShares = c(1/4, 2/4, 1/4),
                                     BodyN = 50)

print( ExpectedRep )
```

GroupRepresentation *Compute group-level representation diagnostics for a political body*

Description

Computes, for each group, the gap between its population share and its share of a political body, both as a raw difference ("shortfall") and as a fraction of the group's population share. These are the group-level "Shortfall" and "Share" measures used in Gerring, Hicken, Jerzak, Moser, and Oncel (book manuscript). The raw shortfall accentuates the under-representation of large groups; the ratio accentuates the under-representation of small groups.

Usage

```
GroupRepresentation(BodyMemberCharacteristics, PopShares, BodyShares)
```

Arguments

BodyMemberCharacteristics	A vector specifying the characteristics for members of a political body.
PopShares	A numeric vector specifying population shares of identities specified in the body-member characteristics input. The names of the entries in PopShares should correspond to identities in that body-member characteristics input (see Example).
BodyShares	(optional) A numeric vector with same structure as PopShares specifying group population shares of a given body. If supplied with names, they are matched to PopShares; otherwise, the order is assumed to correspond to that of PopShares.

Value

A data.frame with one row per group and columns:

- Group: the group label (names of PopShares, or the group index if unnamed).
- PopShare: the group's population share.
- BodyShare: the group's share of the body.
- Shortfall: PopShare - BodyShare. Positive values indicate under-representation; negative values indicate over-representation.
- ShortfallRatio: Shortfall / PopShare, the shortfall as a fraction of the group's population share (Inf or NaN when PopShare is zero). A value of 1 means the group is entirely absent from the body; 0 means proportional representation; negative values indicate over-representation.

Returns NA if the body is empty or shares cannot be resolved.

References

- John Gerring, Connor T. Jerzak, Erzen Oncel. (2024), The Composition of Descriptive Representation, *American Political Science Review*, 118(2): 784-801. doi:10.1017/S0003055423000680
- John Gerring, Allen Hicken, Connor T. Jerzak, Robert G. Moser, Erzen Oncel. Electoral Rules and Descriptive Representation: A Comprehensive View Across Multiple Identities. Book manuscript.

See Also

- [ObservedRepresentation](#) for summarizing these gaps into a single representation index.
- [CompareRepresentation](#) for comparing the representation indices of two bodies.

Examples

```
GroupRep <- GroupRepresentation(
  BodyMemberCharacteristics = c("A", "A", "C", "A", "C", "A"),
  PopShares = c("A"=1/4, "B"=2/4, "C"=1/4))

print( GroupRep )
```

ObservedRepresentation

Compute the observed degree of representation for any group in a political body

Description

Finds the degree of observed representation for any group in a political body.

Usage

```
ObservedRepresentation(BodyMemberCharacteristics, PopShares, BodyShares,
  a = -0.5, b = 1, metric = c("L1", "L2"))
```

Arguments

BodyMemberCharacteristics	A vector specifying the characteristics for members of a political body.
PopShares	A numeric vector specifying population shares of identities specified in the body-member characteristics input. The names of the entries in PopShares should correspond to identities in that body-member characteristics input (see Example).
BodyShares	(optional) A numeric vector with same structure as PopShares specifying group population shares of a given body. If supplied with names, they are matched to PopShares; otherwise, the order is assumed to correspond to that of PopShares.
a, b	Parameters controlling the affine transformation for how the representation measure is summarized. That is, a and b control how the deviation of the population shares from the body shares is re-weighted. This deviation is multiplied by a; b is as an additive re-scaling term: $a \cdot \text{deviation} + b$. By default, $a = -0.5$ and $b = 1$ so that the Rose Index of Proportionality is returned when <code>metric = "L1"</code> .
metric	A character string selecting the deviation metric underlying the representation index. "L1" (the default) uses the sum of absolute deviations between population and body shares, yielding the Rose Index of Proportionality under the default a and b. "L2" uses the sum of squared deviations, yielding the squared-deviation representation index analyzed in Gerring, Hicken, Jerzak, Moser, and Oncel (book manuscript). Because squared deviations of proportions are no larger than absolute deviations, the "L2" index is always greater than or equal to the "L1" index under the default a and b.

Value

The observed degree of representation (a scalar). By default, this quantity is the Rose Index of Proportionality.

See Also

- [ExpectedRepresentation](#) for calculating expected representation scores under random sampling.
- [SDRepresentation](#) for calculating representation unexplained under the random sampling model.

Examples

```
ObsRep <- ObservedRepresentation(
  BodyMemberCharacteristics = c("A", "A", "C", "A", "C", "A"),
  PopShares = c("A"=1/4, "B"=2/4, "C"=1/4))

print( ObsRep )
```

RelativeRepresentation

Compute relative representation compared to random sampling

Description

Calculates the difference between observed and expected representation. Optionally standardizes this difference using the standard deviation of representation under the random sampling model.

Usage

```
RelativeRepresentation(BodyMemberCharacteristics,
                      PopShares,
                      a = -0.5, b = 1,
                      standardize = FALSE,
                      nMonte = 10000,
                      metric = c("L1", "L2"))
```

Arguments

BodyMemberCharacteristics	A vector specifying characteristics for each member of a political body.
PopShares	A numeric vector of population group proportions. Names must correspond to identities in BodyMemberCharacteristics.
a, b	Parameters controlling the affine transformation of the representation index, passed to ObservedRepresentation and ExpectedRepresentation.
standardize	Logical. If TRUE, the difference between observed and expected representation is divided by the standard deviation of representation under random sampling.
nMonte	A positive integer denoting number of Monte Carlo iterations used for estimating the standard deviation when standardize = TRUE.
metric	A character string selecting the deviation metric underlying the representation index ("L1", the default, or "L2"), passed to ObservedRepresentation, ExpectedRepresentation, and SDRepresentation.

Value

A scalar giving the difference between observed and expected representation. If standardize = TRUE, the difference is divided by the standard deviation under the random sampling model.

See Also

[ObservedRepresentation](#), [ExpectedRepresentation](#), [SDRepresentation](#)

SDRepresentation	<i>Compute the amount of representation left unexplained by a random sampling model.</i>
------------------	--

Description

Finds the residual standard deviation when using the expected representation for any group in a political body to predict observed representation as described in Gerring, Jerzak and Oncel (2024).

Usage

```
SDRepresentation(PopShares, BodyN, a = -0.5, b = 1, nMonte = 10000,
                 metric = c("L1", "L2"))
```

Arguments

PopShares	A numeric vector containing the group-level population proportions.
BodyN	A positive integer denoting the size of the political body in question.
a, b	Parameters controlling the affine transformation for how the representation measure is summarized. That is, a and b control how the expected L1 deviation of the population shares from the body shares is re-weighted. The expected L1 deviation is the average value of the absolute deviation of the population from body shares under a random sampling model. This expected L1 deviation is multiplied by a; b is as an additive re-scaling term: $a \cdot E[L1] + b$. By default, $a = -0.5$ and $b = 1$ so that the expected Rose Index of Proportionality is used in the calculation.
nMonte	A positive integer denoting number of Monte Carlo iterations used to approximate the variance of representation under a random sampling model.
metric	A character string selecting the deviation metric underlying the representation index. "L1" (the default) uses the sum of absolute deviations between population and body shares, yielding the Rose Index of Proportionality under the default a and b. "L2" uses the sum of squared deviations, yielding the squared-deviation representation index analyzed in Gerring, Hicken, Jerzak, Moser, and Oncel (book manuscript).

Value

A scalar summary of the amount of representation not explained by a random sampling model. More precisely, this function returns the the residual standard deviation when using the expected degree of representation to predict observed representation under a random sampling model.

References

- John Gerring, Connor T. Jerzak, Erzen Oncel. (2024), The Composition of Descriptive Representation, *American Political Science Review*, 118(2): 784-801. doi:10.1017/S0003055423000680

See Also

- **ExpectedRepresentation** for calculating expected representation scores under random sampling.
- **ObservedRepresentation** for calculating representation scores from observed data.

Examples

```
SDRep <- SDRepresentation(PopShares = c(1/4, 2/4, 1/4),
                          BodyN = 50)

print( SDRep )
```

Index

CompareRepresentation, [2](#), [8](#)

ExpectedDistrictRepresentation, [3](#), [7](#)

ExpectedRepresentation, [5](#), [6](#), [9](#), [10](#), [12](#)

GroupRepresentation, [7](#)

ObservedRepresentation, [3–5](#), [7](#), [8](#), [8](#), [10](#), [12](#)

RelativeRepresentation, [3](#), [10](#)

SDRepresentation, [7](#), [9](#), [10](#), [11](#)