Package 'mediaPlanR'

November 14, 2026

Type Package

Title Reach and Exposure-Distribution Models for Media Planning

Version 2.0.0

Date 2026-10-29

Author Manuel J. Sánchez-Franco [aut, cre] (ORCID:
<https://orcid.org/0000-0002-8042-3550>)

Maintainer Manuel J. Sánchez-Franco <majesus@us.es>

Description Reproducible implementations of reach and exposure-frequency models for media planning. Provides exact reach and exposure distributions under random duplication (Sainsbury and Binomial models); the Beta-Binomial accumulation model and its multi-vehicle version of Metheringham (1964) <doi:10.1080/00218499.1964.12519751>; the ad hoc duplication and accumulation formulas of Agostini (1961) <doi:10.1080/00218499.1961.12519620> and Hofmans (1966) <doi:10.1177/002224376600300307>; and the multivariate exposure distribution models of Danaher (1991) <doi:10.1177/002224379102800311> (canonical expansion), Leckenby and Rice (1986) <doi:10.1080/00913367.1986.10673014> (Morgensztern sequential aggregation), Hyo Gyoo Kim (2005, doctoral dissertation, University of Texas at Austin) (canonical sequential aggregation), Heejin Kim (1994, doctoral dissertation, University of Texas at Austin) (conditional Beta distribution), and Cheong (2007, doctoral dissertation, University of Texas at Austin) and Cheong, Leckenby and Eakin (2011) <doi:10.2753/JOA0091-3367400101> (multivariate Beta-Binomial). The three dissertations have no DOI; their complete references are in the help pages. Also includes an explicitly scoped Negative-Binomial count approximation, the fitting of the Beta-Binomial model to an observed reach, and the evaluation of predicted exposure distributions against observed ones.

License MIT + file LICENSE

URL https://github.com/majesus/mediaPlanR

BugReports https://github.com/majesus/mediaPlanR/issues

Depends R (>= 4.0)

Imports stats, utils

Suggests extraDistr, ggplot2 (>= 3.4.0), knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.3.2

NeedsCompilation no

Repository CRAN

Date/Publication 2026-11-13 23:20:02 UTC

Contents

bbd_reach_example . . . 3
beta_binomial_example . . . 3
calc_agostini_duplication . . . 4
calc_beta_binomial . . . 6
calc_binomial . . . 8
calc_canex . . . 10
calc_cbd . . . 14
calc_csd . . . 20
calc_hofmans_accumulation . . . 24
calc_hofmans_duplication . . . 26
calc_mbd . . . 28
calc_metheringham . . . 34
calc_msad . . . 36
calc_sainsbury . . . 40
calibrate_bbd . . . 42
canex_example . . . 44
csd_example . . . 45
duplication_example . . . 45
evaluate_exposure_model . . . 46
fit_bbd_to_reach . . . 49
fit_nbd_exposure . . . 53
hofmans_accumulation_example . . . 54
mbd_example . . . 55
mediaPlanR . . . 56
metheringham_example . . . 60
msad_example . . . 61
nbd_exposure_distribution . . . 61
print.reach_cbd . . . 63
ratings_example . . . 64

Index **66**

`bbd_reach_example` *Illustrative inputs for fitting a Beta-Binomial to an external reach*

Description

List of arguments ready for `do.call()` with `fit_bbd_to_reach()`: three vehicles and an external reach. Original illustrative data, not derived from any published source.

Usage

```
bbd_reach_example
```

Format

A list with the components:

insertions Number of insertions in each vehicle (three vehicles in this example; `fit_bbd_to_reach()` accepts any number).

audiences Audience of each vehicle, in people per insertion (one value per vehicle).

reach External schedule reach, in people.

population Population size, in people.

Examples

```
data(bbd_reach_example)
do.call(fit_bbd_to_reach, bbd_reach_example)
```

`beta_binomial_example` *Illustrative inputs for the Beta-Binomial model*

Description

List of arguments ready for `do.call()` with `calc_beta_binomial()`. Original illustrative data, not derived from any published source.

Usage

```
beta_binomial_example
```

Format

A list with the components:

R1 Reach after the first insertion, as a proportion.

R2 Cumulative reach after the second insertion, as a proportion.

insertions Total number of planned insertions.

population Population size, in people.

Examples

```
data(beta_binomial_example)
do.call(calc_beta_binomial, beta_binomial_example)
```

`calc_agostini_duplication` *Reach under Agostini's duplication model*

Description

Implements Agostini's (1961) ad hoc formula for the reach of a plan with one insertion in each of several vehicles when the audience of each vehicle and the audience duplicated between every pair of vehicles are known. It belongs to the "duplication" domain of Aldás Manzano's (1998) three-way split, together with `calc_hofmans_duplication()`, and estimates reach only, not the exposure distribution.

Usage

```
calc_agostini_duplication(audiences, population, duplication_matrix, k = 1.125)
```

Arguments

`audiences`	Numeric vector with the audience of each vehicle for one insertion, in people.
`population`	Population size, in people. No audience can exceed it; otherwise the function stops with an error.
`duplication_matrix`	Symmetric numeric matrix whose element `[i, j]` is the number of people who are in the audience of both vehicle `i` and vehicle `j` (one insertion in each), in people (not a proportion). Diagonal values are ignored.
`k`	Agostini's empirical duplication coefficient. The default, 1.125, is the value Agostini fitted to a 1957 French print readership survey and found to work for French and US magazines. Other authors fitted other values to other data (Aldás Manzano, 1998, p. 156), so `k` should be calibrated to the media type analyzed whenever data allow it.

Details

Let $A = \sum_{i=1}^{m} A_i$ be the gross audience (with m the number of vehicles), also called duplicated audience: the simple sum of the vehicle audiences, in which a person who belongs to the audiences of several vehicles is counted once for each of them (Aldás Manzano, 1998, "audiencia bruta"; Kim, 2005, "gross audience"). It is not the reach, which counts each person once. Let $D = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} A_{ij}$ be the sum of the duplicated audiences of every pair of vehicles, each pair counted once. Both A and D are in people, and neither counts distinct people: with one insertion per vehicle, A is the number of contacts of the plan, and D is a sum of coincidences between pairs of vehicles, so a person who belongs to three audiences adds three pairs to D (not one person). The reach R_m is the only quantity that counts people once. Agostini observed that the ratio of reach to

gross audience is a function of the ratio of duplication to gross audience, $R_m/A = 1/(1 + kD/A)$, which gives

$$R_m = \frac{A^2}{A + kD}.$$

The coefficient k is constant across vehicle pairs. `calc_hofmans_duplication()` replaces it by a pairwise coefficient computed from each pair's own audiences.

The formula is empirical and can return a reach outside its logical range (below the largest audience, or above the population or the gross audience). This can happen because the duplications or `k` are not consistent with the formula, but also because the formula is an approximation: five independent vehicles that each reach 50% of the population, with a duplication of 25% in every pair (a valid joint distribution, with a true reach of 96.875%), give 117.6% with the default `k`. The value is then returned unchanged, with a warning.

Value

A list of class `"reach_agostini_duplication"` with components:

- `reach`: list with `percent` and `people`.
- `k`: the duplication coefficient used.
- `gross_audience`: sum of the vehicle audiences (duplicated people counted once per vehicle), in people.
- `total_duplication`: sum of the pairwise duplicated audiences, in people.
- `n_vehicles`: number of vehicles.

References

Agostini, J.-M. (1961). How to estimate unduplicated audiences. Journal of Advertising Research, 1(3), 11-14. doi:10.1080/00218499.1961.12519620

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.2.1.1, equations 3.54-3.57.

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 44-45. Handle 2152/1590 (University of Texas at Austin repository).

See Also

`calc_hofmans_duplication()` for the pairwise refinement of this formula, and `calc_sainsbury()` and `calc_binomial()` for models that assume random duplication instead of observing it.

Examples

```
data(duplication_example)
result <- do.call(calc_agostini_duplication, duplication_example)
result

# A different duplication coefficient
```

```
calc_agostini_duplication(
  duplication_example$audiences, duplication_example$population,
  duplication_example$duplication_matrix, k = 1
)$reach$percent
```

`calc_beta_binomial` *Reach and exposure distribution under the Beta-Binomial model*

Description

Implements the Beta-Binomial model for one vehicle with several insertions. In Aldás Manzano's (1998) classification it is an *accumulation* model, like `calc_hofmans_accumulation()`: it describes how the audience of a single vehicle accumulates over its insertions (the function name omits the word, unlike the Hofmans one, because the Beta-Binomial name is the standard one in the literature and there is no second Beta-Binomial model that could be confused with it). Individuals differ in their exposure probability, which follows a Beta distribution with shape parameters alpha and beta; given that probability, the number of exposures in `insertions` insertions is Binomial. Mixing the two yields the Beta-Binomial exposure distribution. The model needs only two data points, the reach after one insertion (`R1`) and after two insertions (`R2`), to estimate both shape parameters.

Usage

```
calc_beta_binomial(R1, R2, insertions, population = 1)
```

Arguments

`R1`	Reach after the first insertion (the vehicle's audience), as a proportion of the population, a number in (0, 1] (not a percentage). If the audience is known in people, divide it by the population.
`R2`	Cumulative reach after the second insertion (the proportion of the population exposed at least once in two insertions), as a proportion (not a percentage), with `R1 <= R2 <= 2 * R1 - R1^2`, the range of the Beta-Binomial family (see Details). If `R2` is outside that range the function stops with an explanatory error.
`insertions`	Total number of planned insertions in the vehicle (a positive integer).
`population`	Population size, in people. It only scales the number of people reported; the default, 1, leaves `people` equal to the proportion.

Details

The method-of-moments estimators are

$$\hat{\alpha} = \frac{R_1(R_2 - R_1)}{2R_1 - R_1^2 - R_2}, \qquad \hat{\beta} = \frac{\hat{\alpha}(1 - R_1)}{R_1}.$$

The estimators and the mean and variance formulas of the Beta distribution apply to the interior, where $R_1 < R_2 < 2R_1 - R_1^2$ and both parameters are positive and finite. The two bounds are

degenerate cases, reached as limits keeping the mean R_1 fixed while the concentration $\alpha + \beta$ tends to zero or to infinity: at $R_2 = R_1$ all individuals are either always exposed or never exposed (the *polarized* limit: the probability mass is only at zero and at the number of insertions n), and at $R_2 = 2R_1 - R_1^2$ exposures are independent (the *binomial* limit, a Binomial(n, R_1)). The values `alpha = beta = 0` and `alpha = beta = Inf` that the function reports for them are indicators of the limit, not parameters of a proper Beta distribution, and the quotient $\alpha/(\alpha+\beta)$ must not be evaluated with them. Both limits are handled explicitly. $R_1 = 1$ is also degenerate: everyone is exposed.

The upper bound $R_2 \leq 2R_1 - R_1^2$ is the bound of the Beta-Binomial family, not a logical limit of two insertions. In a mixture of independent exposures whose individual probability θ follows a Beta distribution, the duplication between two insertions is $E[\theta^2] \geq E[\theta]^2 = R_1^2$, so that $R_2 = 2R_1 - E[\theta^2] \leq 2R_1 - R_1^2$: the upper bound corresponds to independent insertions, and the family excludes negative dependence between two insertions of the same vehicle. The logical bounds of the cumulative reach of two insertions with the same audience p are wider: $p \leq R_2 \leq \min(1, 2p)$. For example, $p = 0.30$ and $R_2 = 0.55$ are logically possible (the probabilities of being exposed to both insertions, only the first, only the second and neither can be 0.05, 0.25, 0.25 and 0.45), but they are outside the family, whose largest R_2 for $p = 0.30$ is 0.51, and the function stops. Data of that kind should not be replaced by the bound only to make the model run: review how the exposures were measured (see the section "Notation and units" of mediaPlanR) and whether this model suits the vehicle.

Value

A list of class `"reach_beta_binomial"` with components:

- `reach`: list with `percent` and `people`.
- `distribution`: list with `percent` and `people`, for one to `insertions` exposures.
- `cumulative`: list with `percent` and `people`, for individuals exposed at least once, at least twice, and so on.
- `population`: the population size used, in people.
- `parameters`: list with `alpha`, `beta`, `mean_probability` (the mean of the Beta distribution, `R1`), `zero_contact_probability` (a percentage, on a 0-100 scale: it is an exception to the rule that probabilities are proportions between 0 and 1, and its name is kept for compatibility; divide by 100 for a probability) and `type` (`"beta_binomial"`, `"binomial_limit"` or `"polarized_limit"`).

In the two limits, `alpha` and `beta` are `0` (polarized) or `Inf` (binomial). They are indicators of the limit, as `type` is, not parameters of a proper Beta distribution: the quotient `alpha / (alpha + beta)` is undefined there. The mean `R1` stays fixed while the concentration `alpha + beta` tends to zero or to infinity.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.1.2.5, equations 3.25-3.28, with the parameter estimators on page 134.

Metheringham, R. A. (1964). Measuring the net cumulative coverage of a print campaign. Journal of Advertising Research, 4(4), 23-28. doi:10.1080/00218499.1964.12519751

Chandon, J.-L. (1976). A comparative study of media exposure models. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Leckenby, J. D., & Boyd, M. M. (1984). An improved beta binomial reach/frequency model for magazines. Current Issues and Research in Advertising, 7(1), 1-24.

See Also

`calc_sainsbury()`, `calc_binomial()` and `calc_metheringham()` for plans with several vehicles, `calc_hofmans_accumulation()` for an ad hoc accumulation model, and `nbd_exposure_distribution()` for a Negative-Binomial count approximation, which applies to unbounded count processes rather than to a finite number of insertions.

Examples

```
result <- calc_beta_binomial(R1 = 0.50, R2 = 0.55, insertions = 5,
                             population = 1000000)
result$reach$percent
result$parameters$alpha
result$parameters$beta
```

`calc_binomial` *Reach and exposure distribution under the Binomial model*

Description

Implements the Binomial model of Lee and Burkart (1960), reviewed by Chandon, to calculate reach and the exposure distribution (and its cumulative counterpart) of a media plan with several vehicles and one or more insertions in each of them (argument `insertions`). The model assumes random duplication *and* random accumulation (repeat insertions in the same vehicle are also independent), homogeneous individuals and homogeneous vehicles. Under these assumptions every insertion is an independent Bernoulli trial with the same exposure probability, which is the insertion-weighted mean audience of the plan (each vehicle's audience counted once per insertion) divided by the population. Exposure probabilities are assumed stationary over time.

Usage

```
calc_binomial(audiences, population, insertions = rep(1L, length(audiences)))
```

Arguments

`audiences`	Numeric vector with the audience of each vehicle, in people per insertion.
`population`	Population size, in people. No audience can exceed it; otherwise the function stops with an error.
`insertions`	Non-negative integer vector with the number of insertions planned in each vehicle. It defaults to one insertion per vehicle, the model's original scope.

Details

Aldás Manzano (1998, Section 3.3.1.1) defines the exposure probability as $p = \bar{A}/P$, where $\bar{A}$ is the simple mean of the vehicle audiences and every vehicle receives the same number of insertions n, so a plan with m vehicles has $N = nm$ insertions. With unequal insertion counts, this function uses the insertion-weighted mean audience, $\sum_{i=1}^{m} n_i A_i / \sum_{i=1}^{m} n_i$. The two definitions coincide when all vehicles receive the same number of insertions, and the weighted mean preserves the plan's expected number of exposures in general.

The number of exposures then follows a Binomial distribution with N trials and success probability p, and reach equals $1 - (1 - p)^N$.

The Binomial calculation is exact for the homogenized model that it builds (every insertion with the same probability), not for any plan of independent vehicles. Replacing the different probabilities of the vehicles by their insertion-weighted mean preserves the mean number of contacts but, in general, changes the reach and the exposure distribution. For example, with audiences of 10 and 90 people in a universe of 100 and one insertion in each vehicle, both models have a mean of one contact per person, but `calc_sainsbury()`, which keeps the probability of each vehicle, gives a reach of 91% and the Binomial 75%. The difference is the effect of unequal audiences, not a numerical error. The Binomial reach can even be smaller than the audience of a single vehicle (here, 75 people against the 90 of the larger vehicle): it is the reach of the homogenized model, not of the plan, so it is not a figure that the original schedule can have.

Value

A list of class `"reach_binomial"` with components:

- `reach`: list with `percent` and `people`.
- `distribution`: list with `percent` and `people`, one value for each number of exposures from one to the total number of insertions.
- `cumulative`: list with `percent` and `people`, for individuals exposed at least once, at least twice, and so on.
- `mean_probability`: exposure probability used for every insertion.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.2.2.1 for one insertion per vehicle; Section 3.3.1.1 for several insertions.

Lee, A. M., & Burkart, A. J. (1960). Some optimization problems in advertising media planning. Operational Research Quarterly, 11(3), 113-122.

Chandon, J.-L. (1976). A comparative study of media exposure models. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

See Also

`calc_sainsbury()` for heterogeneous vehicles, `calc_beta_binomial()` and `calc_metheringham()` for models with heterogeneous individuals.

Examples

```
audiences <- c(300000, 400000, 200000)
population <- 1000000
result <- calc_binomial(audiences, population)
result$reach$percent
result$mean_probability

# The same three vehicles with several insertions each
calc_binomial(audiences, population, insertions = c(4, 6, 10))
```

`calc_canex` *Canonical Expansion model (CANEX)*

Description

Implements Danaher's (1991) canonical expansion model for the reach and exposure distribution of a multi-vehicle schedule. Each vehicle's exposure distribution is a Beta-Binomial, which accounts for the heterogeneity of individuals, and the observed duplication between vehicles enters through pairwise correlation terms (see "Order of the expansion").

Usage

```
calc_canex(vehicles_data, duplications, population = 1)
```

Arguments

`vehicles_data` Data frame with one row per vehicle and columns

- `insertions`: number of insertions in the vehicle (a positive integer). The column may also be named `k`, Danaher's notation, so that the same data frame can be used by `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`, which name it `insertions`;
- `R1`: reach after the first insertion, as a proportion between 0 and 1 (not a percentage), in (0, 1];
- `R2`: cumulative reach after the second insertion (the proportion exposed at least once in two insertions), a proportion between `R1` and the independence limit `2 * R1 - R1^2` (the range of the Beta-Binomial family; see `calc_beta_binomial()`). If `R2` is outside that range the function stops with an explanatory error. As in `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`, it may be `NA` for a vehicle with a single insertion, whose exposure is a Bernoulli variable of probability `R1` and does not use it: a value supplied for such a vehicle is ignored, not validated. It is required (and validated) for every vehicle with two or more insertions.

`duplications` Symmetric matrix whose element `[i, j]` is the proportion of the population that is exposed to both vehicle `i` and vehicle `j` (one insertion in each), a number between 0 and 1 (not a percentage). It is not the proportion exposed to each vehicle separately. The diagonal is ignored.

`population` Number of people in the population (a count, not a proportion). It only converts proportions into people: the `people` columns of the result are the probability times `population`, and with the default, 1, they equal the probability.

Details

The model follows these steps:

1. It computes each vehicle's Beta-Binomial parameters from `R1` and `R2` by the method of moments.
2. It turns the duplications into canonical correlations,

$$r_{ij} = \frac{p_{ij} - p_i p_j}{\sqrt{p_i(1-p_i)p_j(1-p_j)}},$$

   with p_i the one-insertion reach of vehicle i and p_{ij} the duplication of vehicles i and j: the proportion of the population exposed to both of them.
3. It evaluates, for every combination of exposure levels, the joint probability $\prod_{i=1}^{m} f_i(x_i)\{1 + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} r_{ij}(x_i-\mu_i)(x_j-\mu_j)/(\sigma_i\sigma_j)\}$, where m is the number of vehicles, the double sum runs over every pair of vehicles, and f_i, μ_i and σ_i^2 are the Beta-Binomial probability function, mean and variance of vehicle i. Only the pairwise terms, linear in the correlations, are kept (see "Order of the expansion" below).
4. It sets to zero the negative probabilities that the truncated expansion can produce, renormalizes the joint distribution to sum to one, and sums it over the total number of exposures.

The number of combinations, $\prod_{i=1}^{m}(n_i+1)$ with n_i the insertions of vehicle i, grows exponentially with the number of vehicles and insertions. The model is therefore intended for small and medium schedules, and the function stops with an informative error above 2,000,000 combinations instead of exhausting memory.

Value

A list of class `"reach_canex"` with components:

- `population`: the population size used.
- `reach`: list with `probability`, `percent` and `people`.
- `average_frequency`: average number of exposures among the people reached (`NA` if nobody is reached).
- `distribution`: data frame with `contacts` (total exposures, from zero), `probability`, `percent`, `people` and `cumulative_probability` (the probability of `contacts` or more exposures).
- `diagnostics`: list with `negative_mass_truncated` (probability mass that was negative before truncation), `mass_before_renormalization` and `correlation_min_eigenvalue` (the smallest eigenvalue of the correlation matrix), `mean_exposures_expected` (the mean number of exposures per person of the population: the sum over vehicles of insertions times `R1`, preserved by the expansion) and `mean_exposures_result` (the mean of the returned distribution). They show how much the truncated expansion had to be corrected.

Order of the expansion

Danaher (1991, pp. 362-364) speaks of "order" in more than one sense, and the help pages of this package use the following. The canonical expansion of the joint probability of two vehicles has one term for each canonical correlation. The model keeps the first, which is the ordinary correlation r_{ij} of the two exposures, and omits the canonical correlations of second and higher order. For three or more vehicles the joint probability is the product of the marginals times one plus the sum, over the pairs of vehicles, of the correlation of the pair times the product of their standardized exposures (step 3). It is linear in the correlations: the products of correlations (terms of second and higher order in them) and, with them, any interaction among three or more vehicles are omitted. The help pages of `calc_csd()` and `calc_cbd()`, which use the same expansion for the all-zero cell and for the (0,1) grid, call it the pairwise (or truncated) canonical expansion. The sources describe it by what it omits, the canonical terms "of the second order or higher" (Danaher, 1991; Cheong, 2007, p. 54), so it is not an expansion that keeps second-order terms.

Truncation of negative probabilities

Danaher (1991, pp. 363-364) shows that the probabilities of the expansion sum to one and reports, in his application, negative probabilities in a small fraction of the cells, with a mean magnitude of the order of 1e-7, which he sets to zero. That practice rests on the negative mass being negligible. This function applies it at any size and reports the truncated mass in `diagnostics$negative_mass_truncated`, together with the mean number of exposures per person that the expansion preserves (`mean_exposures_expected`, the sum over the vehicles of insertions times `R1`) and the one of the returned distribution (`mean_exposures_result`). Both are averages over the whole population, including the people never exposed, so they are not the average frequency among the people reached. When the truncated mass is not negligible, truncating and renormalizing can change the mean. The function then issues a warning of class `mediaPlanR_canex_truncation`; the threshold, a total negative mass of 1e-4, is a decision of this package and not of Danaher (1991). For example, two vehicles with two insertions each, `R1 = c(0.3, 0.2)`, `R2 = c(0.4, 0.3)` and zero duplication have an expected mean of 1 exposure per person (2 insertions x 0.3 plus 2 insertions x 0.2, counted over the whole population) and a negative mass of 0.0076, and the truncated distribution has a mean of 1.0227.

The ratio between contacts and reach holds only within one base. The `average_frequency` of the result is the mean of the returned (corrected) distribution divided by its reach, so `average_frequency * reach` equals `mean_exposures_result`, not the gross contacts per person of the input, which is what a GRP figure computed from the audiences measures (`mean_exposures_expected`). In the example above, the contacts of the input per person are 1 and those of the corrected distribution 1.0227. When `mean_exposures_result` and `mean_exposures_expected` differ, do not divide the input GRP by the corrected reach to obtain a frequency; report both diagnostics and say which base each figure uses.

Domain of validity

In plain words. These models take the overlaps between vehicles (the duplications) from your data and use them to build the whole distribution of exposures. For that to be possible the overlaps you supply must be possible, and plausible together:

- Two vehicles cannot overlap more than the smaller of their audiences, nor less than their audiences force them to (two audiences of 60% of the population must overlap by at least 20%). If a duplication breaks these limits, the function stops with an error naming the pair: the data cannot be right and should be reviewed.

- The overlaps of three vehicles must also fit together. If they cannot, the function stops with an error naming the three vehicles.
- The studies that evaluated these models mostly used vehicles that overlap at least as much as chance alone would produce, that is, at least the product of their audiences (two vehicles with audiences of 30% and 20% overlap by 6% if people chose vehicles independently). A smaller overlap is not impossible and is not necessarily a measurement error: it can be valid data. But when a pair overlaps clearly less than that, the approximations of the models can differ a lot from one model to another, and the function issues a warning that names the pair. Check how the data were defined and read the diagnostics; do not replace valid observations automatically.

An error has one of two causes. Either the inputs break a constraint of the data (impossible overlaps: correct them), or the inputs are possible but the model's approximation cannot reproduce them (the third limit below: a valid probabilistic construction can still be rejected by an approximate model). Read the message to tell them apart, and do not change valid data only to make a model run. A warning does not invalidate the calculation, but it asks for caution: check the duplication of the pair named, which is often a measurement problem (see below), and compare the result with another model. Every diagnostic must be inspected before the result is interpreted; none can be assumed to be harmless or, on the contrary, to invalidate the result. The rest of this section gives the technical detail.

The multivariate models (`calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`) take the observed one-insertion duplications as given and check them against the Fréchet bounds implied by the audiences and, for every triple of vehicles, against joint feasibility. The published methods have three limits that follow from this.

First, Hong (1998, pp. 238-239) formulates, for the models and the data of his study, the hypothesis that the duplication between vehicles is not below the random one, the product of the audiences, and treats null observed duplications as measurement problems. It is not a condition of the data or of the formulas: Danaher's canonical expansion itself admits negative correlations (Danaher, 1991, equation 3), and the logical limits of a duplication are the Fréchet bounds, $\max(0, p_i + p_j - 1) \leq p_{ij} \leq \min(p_i, p_j)$. For example, two vehicles with one insertion each and audiences of 0.30 and 0.20 can have a duplication of 0.03, below the random 0.06, with the valid joint distribution (neither, only the second, only the first, both) of (0.53, 0.17, 0.27, 0.03), and `calc_canex()` reproduces it without negative probabilities. With several insertions per vehicle, in contrast, a lower duplication can make the approximations diverge considerably. The functions warn (class `mediaPlanR_low_duplication`) when the canonical correlation of a pair is below -0.1. That threshold is a decision of this package: the published reference plans (Kim 2005, Appendix B, and Hong 1998, Appendix E; 80 plans) include pairs slightly below the random duplication, with canonical correlations down to -0.071, and they raise no warning. For two vehicles with two insertions each, `R1 = 0.30` and `R2 = 0.45`, a duplication of zero gives a reach of 0.90 under MSAD, 0.81 under MBD, 0.78 under CSD and CBD, and 0.79 under CANEX. Hong (1998, pp. 238-239) treats a null duplication as a measurement error and replaces it by $\hat{p}_{ij} = K' p_i p_j$, where K' is the ratio between observed and random duplication over the pairs with non-zero observed duplication. The package does not implement this imputation: `calc_agostini_duplication()` and `calc_hofmans_duplication()` estimate reach from a duplication matrix and do not impute its elements. If duplications are imputed beforehand, the result must be checked against the bounds described below. A null duplication is not always a measurement error: it can also be a genuine observation.

Second, pairwise duplications that satisfy the Fréchet bounds can still be impossible jointly for

three or more vehicles. For every triple of vehicles the functions check that the exposure to all three can have a non-negative probability, given the audiences and the three pairwise duplications, and stop with an error when it cannot. That condition is necessary but not sufficient for four or more vehicles, for which the published methods check feasibility at most to first order (Cheong 2007); `calc_mbd()` and `calc_cbd()` zero any negative cell that remains, and `calc_canex()` truncates it.

Third, a low duplication between vehicles with several insertions can make the target of a sequential model invalid. For two vehicles with two insertions each, `R1 = 0.5` and `R2 = 0.75`, `calc_csd()` and `calc_cbd()` stop with a duplication of zero, and `calc_msad()` with a duplication of zero or 0.5, instead of altering an invalid target.

References

Danaher, P. J. (1991). A canonical expansion model for multivariate media exposure distributions: A generalization of the "duplication of viewing law". Journal of Marketing Research, 28(3), 361-367. doi:10.1177/002224379102800311

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 56-58. Handle 2152/1590 (University of Texas at Austin repository).

See Also

`calc_beta_binomial()` for the univariate model (one vehicle), and `calc_csd()` and `calc_cbd()`, whose between-vehicle step reuses this canonical expansion.

Examples

```
# Two vehicles, two insertions each
data(canex_example)
result <- do.call(calc_canex, canex_example)
result
result$diagnostics

# Two vehicles with two insertions each, with a population of 500,000 people
vehicles <- data.frame(
  insertions = c(2, 2),
  R1 = c(0.40, 0.05),
  R2 = c(0.50, 0.08)
)
duplications <- matrix(c(NA, 0.02, 0.02, NA), nrow = 2)
calc_canex(vehicles, duplications, population = 500000)$reach
```

`calc_cbd` *Conditional Beta Distribution model*

Description

Implements the Conditional Beta Distribution (CBD) of Kim (1994), reviewed in Kim (2005), for several vehicles with several insertions each. It is a multivariate model of conditional expansion, not a sequential-aggregation model like `calc_csd()` and `calc_msad()`: the vehicles are not combined one at a time, and the result does not depend on their order. The "(0,1) grid" is the set of the 2^m combinations of being exposed (1) or not exposed (0) to the first insertion of each of the m vehicles; the notation "(0,1)" refers to those two states, not to an interval. Between-vehicle duplication is modeled first, at the one-insertion (0,1) level, by Danaher's (1991) canonical expansion, truncated to its pairwise terms – the mechanism `calc_canex()` and `calc_csd()` use (see the section "Order of the expansion" of `calc_canex()`). Conditionally on the (0,1) exposure state of every vehicle, the remaining insertions of each vehicle follow a Beta-Binomial distribution whose Beta parameters are updated by that state, and the vehicles are independent given the grid (Danaher 1992a). The zero cell of the grid is then conformed to the zero probability of the schedule with all the insertions, which the general formulation of CBD lets any existing exposure model supply and which this implementation takes from the canonical expansion, as in `calc_canex()` (see step 3 below).

Usage

```
calc_cbd(
  vehicles_data,
  duplications,
  aggregation_order = c("audience_desc", "given"),
  population = 1,
  tolerance = 1e-08
)
```

Arguments

`vehicles_data`	Data frame with columns `insertions`, `R1` and `R2`, with the same convention as `calc_csd()` (`R1` and `R2` are proportions between 0 and 1, not percentages, `R1` is strictly between zero and one, `R1 <= R2 <= 2 * R1 - R1^2`, the range of the Beta-Binomial family (see `calc_beta_binomial()`), and `R2` may be `NA` only when `insertions` is one; the function stops with an explanatory error if `R2` is outside that range). At most 12 vehicles are supported.
`duplications`	Symmetric matrix whose element `[i, j]` is the proportion of the population that is exposed to both vehicle `i` and vehicle `j` (one insertion in each), a number between 0 and 1 (not a percentage). It is not the proportion exposed to each vehicle separately. The diagonal is ignored.
`aggregation_order`	Either `"audience_desc"` (vehicles in decreasing order of one-insertion reach `R1`), `"given"` (the row order), or a permutation of the row indices. CBD does not use it: in Kim's specification the vehicles are conditionally independent given the (0,1) grid, so neither the reach nor the distribution depends on the order. The argument is accepted, validated and reported only so that the same call, and the same example data (for instance csd_example), work for `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`.

`population` Number of people in the population (a count, not a proportion). It only converts probabilities into people: the `people` columns of the result are the probability times `population`, and with the default, 1, they equal the probability.

`tolerance` Positive numerical tolerance for probability constraints. The Fréchet and triple-feasibility checks apply it relative to the smaller audience involved.

Details

The model follows Kim (1994, pp. 115-124):

1. The joint distribution of exposure to one insertion of each vehicle (the 2^m cells of the (0,1) grid) is obtained from the canonical expansion truncated to its pairwise terms, with Bernoulli marginals, from the pairwise correlations implied by the observed one-insertion duplications.
2. Given that vehicle i was ($x_i = 1$) or was not ($x_i = 0$) exposed to its first insertion, its remaining $n_i - 1$ insertions follow a Beta-Binomial distribution with parameters $(\alpha + x_i, \beta + 1 - x_i)$, where α and β are fitted from `R1` and `R2`. The vehicle's total exposure is that distribution shifted by x_i. The distributions of the vehicles are convolved within each cell of the grid and weighted by the cell probability. A vehicle with one insertion contributes only its state x_i.
3. In the all-zero cell of the grid, the probability of zero exposures is set to the zero probability of the schedule with all the insertions. In its general formulation CBD leaves the source of that probability open: Kim (1994, p. 124) lets any existing exposure model supply it, and Kim (2005, p. 65) gives CANEX as an example. The version implemented here, like Kim's worked example, takes it from the canonical expansion of Danaher (1991), as in `calc_canex()`. The remaining levels of that cell are rescaled so that the cell keeps its total probability (Kim 1994, pp. 124 and 131).

With a binomial-limit vehicle (`R2` equal to the reach under independence) the conditional distribution is the binomial one, and with a polarized one (`R2` equal to `R1`) it is the all-or-nothing exposure; both are the limits of the Beta-Binomial updating.

Unlike `calc_mbd()`, the between-vehicle step of CBD needs no imputation for three or more vehicles. The canonical expansion uses the mean R_{1i} and variance $R_{1i}(1 - R_{1i})$ of each vehicle's one-insertion exposure; this guarantees that, in the grid built in step 1 and before any correction, summing out all other vehicles returns each vehicle's own Bernoulli marginal exactly, and with zero correlation the model reduces to the exact convolution of the vehicles' Beta-Binomial marginals. That property belongs to the uncorrected grid, not to the returned distribution: the zero-cell adjustment of step 3 and, if it is engaged, the truncation of negative cells and renormalization change the marginals of each vehicle and the mean number of exposures. The departure of the mean is reported in `diagnostics$mean_error`. It can have either sign and it is not bounded by a universal constant: it is -0.0044 exposures per person for csd_example; in the maintainer's local runs it ranged from -0.011 to +0.002 over the 80 two-vehicle plans of Kim (2005) and Hong (1998) and was -0.009 in the three-vehicle example of Kim (1994). The marginals of the result should therefore be judged through those diagnostics, separating the binary marginal of each vehicle, its distribution within the vehicle and the final distribution of the plan. As with `calc_canex()`, the `average_frequency` is the mean of the returned distribution over its reach: it uses the corrected mean, not the gross contacts per person of the input, so the two bases should not be mixed.

Value

A reach_cbd object: a list with reach (probability, percent and people), average_frequency, the complete exposure distribution (contacts, probability, percent, people and cumulative_probability), the aggregation_order and aggregation_rule, the within-vehicle vehicle_expansion (the distribution used for each vehicle) and diagnostics, which include the canonical zero-exposure probability, the zero-cell rescaling factor and whether the negative-probability safety net of calc_mbd() had to be engaged (negative_mass_adjusted, cells_adjusted).

Validation against published results

The implementation reproduces the CBD columns of the two-vehicle comparisons of Kim (2005, Appendix B) and of Hong (1998, Appendix E) within the rounding of those sources (0.01 percentage points), taking the duplication that each source implies. The sources do not print those duplications: they were reconstructed from another model's published column (CANEX for Kim, the MSAD reach for Hong), so this agreement checks the consistency between columns of published tables, and algorithmic fidelity, not predictive accuracy against independent observations (the maintainer's local tests, whose input tables are not distributed, accept up to 0.02 percentage points, although the largest deviation measured is below 0.01). The function also reproduces the worked example of Kim (1994, pp. 125-139; three vehicles with 2, 1 and 3 insertions, SMRB 1979 data) except for its zero cell. In that example the zero of the canonical expansion is printed as 0.5066, whereas the formula applied to the printed inputs gives about 0.510 (and no choice of the inputs within their rounding goes below 0.507), so the printed value cannot be reproduced from the printed inputs. Page 139 of that source contains verifiable errata (a probability of 0.0055 printed as 0.06% instead of 0.55%, and percentages that add to 99.54% although labeled 100), but they do not by themselves establish the cause of the discrepancy in the zero cell, which is treated here as an unexplained difference with the printed value. This function evaluates the formula, so its zero and one-contact probabilities differ from the printed ones by 0.36 and 0.29 percentage points in that example. See the package tests for the exact cases.

Because the canonical expansion can assign small negative probabilities to some (0,1) cells – the limitation documented for calc_canex() – any cell that is still negative at the end is set to zero and that mass is redistributed proportionally, as in the final safety net of calc_mbd(). This step is an extension of the package, not part of Kim's specification, and is reported in diagnostics. If the canonical zero-exposure probability is not a probability, the function stops, as calc_csd() does.

The number of cells of the (0,1) grid grows as 2^m, so the number of vehicles is limited to 12.

Computational limits

The work of the schedule is estimated from the insertions before anything is allocated, and a schedule above the budget is refused with an explanation. The budget is the option mediaPlanR.max_work, 1e9 elementary operations by default (one to two minutes); see the section "Computational limits" of mediaPlanR.

Domain of validity

In plain words. These models take the overlaps between vehicles (the duplications) from your data and use them to build the whole distribution of exposures. For that to be possible the overlaps you supply must be possible, and plausible together:

- Two vehicles cannot overlap more than the smaller of their audiences, nor less than their audiences force them to (two audiences of 60% of the population must overlap by at least 20%). If a duplication breaks these limits, the function stops with an error naming the pair: the data cannot be right and should be reviewed.
- The overlaps of three vehicles must also fit together. If they cannot, the function stops with an error naming the three vehicles.
- The studies that evaluated these models mostly used vehicles that overlap at least as much as chance alone would produce, that is, at least the product of their audiences (two vehicles with audiences of 30% and 20% overlap by 6% if people chose vehicles independently). A smaller overlap is not impossible and is not necessarily a measurement error: it can be valid data. But when a pair overlaps clearly less than that, the approximations of the models can differ a lot from one model to another, and the function issues a warning that names the pair. Check how the data were defined and read the diagnostics; do not replace valid observations automatically.

An error has one of two causes. Either the inputs break a constraint of the data (impossible overlaps: correct them), or the inputs are possible but the model's approximation cannot reproduce them (the third limit below: a valid probabilistic construction can still be rejected by an approximate model). Read the message to tell them apart, and do not change valid data only to make a model run. A warning does not invalidate the calculation, but it asks for caution: check the duplication of the pair named, which is often a measurement problem (see below), and compare the result with another model. Every diagnostic must be inspected before the result is interpreted; none can be assumed to be harmless or, on the contrary, to invalidate the result. The rest of this section gives the technical detail.

The multivariate models (`calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`) take the observed one-insertion duplications as given and check them against the Fréchet bounds implied by the audiences and, for every triple of vehicles, against joint feasibility. The published methods have three limits that follow from this.

First, Hong (1998, pp. 238-239) formulates, for the models and the data of his study, the hypothesis that the duplication between vehicles is not below the random one, the product of the audiences, and treats null observed duplications as measurement problems. It is not a condition of the data or of the formulas: Danaher's canonical expansion itself admits negative correlations (Danaher, 1991, equation 3), and the logical limits of a duplication are the Fréchet bounds, $\max(0, p_i + p_j - 1) \leq p_{ij} \leq \min(p_i, p_j)$. For example, two vehicles with one insertion each and audiences of 0.30 and 0.20 can have a duplication of 0.03, below the random 0.06, with the valid joint distribution (neither, only the second, only the first, both) of (0.53, 0.17, 0.27, 0.03), and `calc_canex()` reproduces it without negative probabilities. With several insertions per vehicle, in contrast, a lower duplication can make the approximations diverge considerably. The functions warn (class `mediaPlanR_low_duplication`) when the canonical correlation of a pair is below -0.1. That threshold is a decision of this package: the published reference plans (Kim 2005, Appendix B, and Hong 1998, Appendix E; 80 plans) include pairs slightly below the random duplication, with canonical correlations down to -0.071, and they raise no warning. For two vehicles with two insertions each, `R1 = 0.30` and `R2 = 0.45`, a duplication of zero gives a reach of 0.90 under MSAD, 0.81 under MBD, 0.78 under CSD and CBD, and 0.79 under CANEX. Hong (1998, pp. 238-239) treats a null duplication as a measurement error and replaces it by $\hat{p}_{ij} = K' p_i p_j$, where K' is the ratio between observed and random duplication over the pairs with non-zero observed duplication. The package does not implement this imputation: `calc_agostini_duplication()` and `calc_hofmans_duplication()` estimate reach from a duplication matrix and do not impute its

elements. If duplications are imputed beforehand, the result must be checked against the bounds described below. A null duplication is not always a measurement error: it can also be a genuine observation.

Second, pairwise duplications that satisfy the Fréchet bounds can still be impossible jointly for three or more vehicles. For every triple of vehicles the functions check that the exposure to all three can have a non-negative probability, given the audiences and the three pairwise duplications, and stop with an error when it cannot. That condition is necessary but not sufficient for four or more vehicles, for which the published methods check feasibility at most to first order (Cheong 2007); `calc_mbd()` and `calc_cbd()` zero any negative cell that remains, and `calc_canex()` truncates it.

Third, a low duplication between vehicles with several insertions can make the target of a sequential model invalid. For two vehicles with two insertions each, `R1 = 0.5` and `R2 = 0.75`, `calc_csd()` and `calc_cbd()` stop with a duplication of zero, and `calc_msad()` with a duplication of zero or 0.5, instead of altering an invalid target.

References

Kim, H. (1994). A conditional beta distribution model for advertising media reach/frequency estimation. Unpublished doctoral dissertation, The University of Texas at Austin, pp. 115-139.

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 59-65. Handle 2152/1590 (University of Texas at Austin repository).

Hong, J. (1998). Advertising media models for Internet reach/frequency estimation. Unpublished doctoral dissertation, The University of Texas at Austin, Appendix E.

Danaher, P. J. (1991). A canonical expansion model for multivariate media exposure distributions: A generalization of the "duplication of viewing law". Journal of Marketing Research, 28(3), 361-367. doi:10.1177/002224379102800311

Danaher, P. J. (1992a). A Markov-chain model for multivariate magazine-exposure distributions. Journal of Business & Economic Statistics, 10(4), 401-407. doi:10.1080/07350015.1992.10509915

See Also

`calc_mbd()` for a model with an imputed between-vehicle step, and `calc_canex()` and `calc_csd()` for the canonical expansion that this model's first and last steps reuse.

Examples

```
# Same three-vehicle inputs as the CSD example, so that CBD can be compared
# with CSD and CANEX on identical data
data(csd_example)
result <- do.call(calc_cbd, csd_example)
result$reach
result$distribution
```

`calc_csd` *Canonical Sequential Aggregation Distribution model*

Description

Implements Kim's (2005) Canonical Sequential Aggregation Distribution (CSD): the Beta-Binomial marginal of each vehicle is combined sequentially with the vehicles already aggregated, while the reach at every step comes from Danaher's canonical expansion, truncated to its pairwise terms (see the section "Order of the expansion" of `calc_canex()`). The non-random convolution preserves both marginal distributions and sets the zero-exposure cell to one minus the canonical reach.

Usage

```
calc_csd(
  vehicles_data,
  duplications,
  aggregation_order = c("audience_desc", "given"),
  population = 1,
  tolerance = 1e-10
)
```

Arguments

`vehicles_data`	Data frame with one row per vehicle (at least two rows) and columns `insertions` (planned insertions, a positive integer), `R1` and `R2`. `R1` is the reach after one insertion, a proportion between 0 and 1 (not a percentage) and strictly between zero and one, and `R2` the cumulative reach after two insertions (the proportion exposed at least once in two insertions), a proportion with `R1 <= R2 <= 2 * R1 - R1^2` (the upper limit is the reach under independent insertions, the bound of the Beta-Binomial family; see `calc_beta_binomial()`). If `R2` is outside that range the function stops with an explanatory error. `R2` may be `NA` only when `insertions` is one.
`duplications`	Symmetric matrix whose element `[i, j]` is the proportion of the population that is exposed to both vehicle `i` and vehicle `j` (one insertion in each), a number between 0 and 1 (not a percentage). It is not the proportion exposed to each vehicle separately. The diagonal is ignored.
`aggregation_order`	Either `"audience_desc"` (vehicles in decreasing order of one-insertion reach `R1`), `"given"` (the row order), or a permutation of the row indices. The vehicles are combined one at a time in this order. The order does not change the reach of the whole schedule but can change the exposure distribution. Use an explicit permutation to reproduce a published aggregation order, such as the one of Kim's worked example (labeled "TD forward" in the dissertation: aggregation guided by the total duplication of each pair, between- plus within-vehicle, in forward order).

`population` Number of people in the population (a count, not a proportion). It only converts probabilities into people: the `people` columns of the result are the probability times `population`, and with the default, 1, they equal the probability.

`tolerance` Positive numerical tolerance for probability constraints. The Fréchet and triple-feasibility checks apply it relative to the smaller audience involved.

Details

For a subset of m vehicles, CSD estimates the all-zero probability as

$$P(0,\ldots,0) = \prod_{i=1}^{m} f_i(0)\left[1 + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \rho_{ij}\frac{(0-\mu_i)(0-\mu_j)}{\sigma_i\sigma_j}\right],$$

where f_i, μ_i and σ_i are the Beta-Binomial probability function, mean and standard deviation of vehicle i, and ρ_{ij} is the canonical correlation obtained from the observed one-insertion duplication. Reach is one minus this probability. It is evaluated as the reach of the vehicles already aggregated plus that of the vehicle added minus the reach of both, which is the same quantity without the cancellation by which one minus a product of probabilities close to one loses the reach of a small audience (the reach of the schedule is the sum of the positive cells of its distribution).

The canonical expansion is a truncated approximation (pairwise terms only). Unlike `calc_canex()`, which truncates negative probabilities and renormalizes the full joint grid, `calc_csd()` does not alter an invalid canonical target: it stops if the resulting reach is not a probability or is incompatible with the two margins being combined. This makes approximation failure visible to the analyst.

Kim's worked example rounds intermediate values to four decimals. Exact calculations from the same inputs therefore differ by a few hundredths of a percentage point from some displayed cells; the unrounded calculation is returned. The inputs of that example are not distributed with the package.

Value

A `reach_csd` object: a list with `reach` (`probability`, `percent` and `people`), `average_frequency`, the complete exposure `distribution` (`contacts`, `probability`, `percent`, `people` and `cumulative_probability`), the `vehicle_marginals`, the one-insertion `vehicle_reach`, the canonical `correlation_matrix`, the `aggregation_order` and `aggregation_rule`, the aggregation `steps` and numerical `diagnostics`. The diagnostics include `gross_mean_contacts`, the mean number of exposures per person that the inputs imply (the sum over vehicles of insertions times `R1`, averaged over the whole population), `distribution_mean_contacts`, the same mean computed from the returned distribution, and `mean_error`, their difference.

Computational limits

The cells of the largest joint table (the exposure levels of the vehicles already combined times those of the vehicle added) are counted from the insertions before anything is allocated, and a schedule above the budget is refused with an explanation. The budget is the option `mediaPlanR.max_table_cells`, `1e7` cells by default (80 MB for each matrix of that size); see the section "Computational limits" of mediaPlanR.

Domain of validity

In plain words. These models take the overlaps between vehicles (the duplications) from your data and use them to build the whole distribution of exposures. For that to be possible the overlaps you supply must be possible, and plausible together:

- Two vehicles cannot overlap more than the smaller of their audiences, nor less than their audiences force them to (two audiences of 60% of the population must overlap by at least 20%). If a duplication breaks these limits, the function stops with an error naming the pair: the data cannot be right and should be reviewed.
- The overlaps of three vehicles must also fit together. If they cannot, the function stops with an error naming the three vehicles.
- The studies that evaluated these models mostly used vehicles that overlap at least as much as chance alone would produce, that is, at least the product of their audiences (two vehicles with audiences of 30% and 20% overlap by 6% if people chose vehicles independently). A smaller overlap is not impossible and is not necessarily a measurement error: it can be valid data. But when a pair overlaps clearly less than that, the approximations of the models can differ a lot from one model to another, and the function issues a warning that names the pair. Check how the data were defined and read the diagnostics; do not replace valid observations automatically.

An error has one of two causes. Either the inputs break a constraint of the data (impossible overlaps: correct them), or the inputs are possible but the model's approximation cannot reproduce them (the third limit below: a valid probabilistic construction can still be rejected by an approximate model). Read the message to tell them apart, and do not change valid data only to make a model run. A warning does not invalidate the calculation, but it asks for caution: check the duplication of the pair named, which is often a measurement problem (see below), and compare the result with another model. Every diagnostic must be inspected before the result is interpreted; none can be assumed to be harmless or, on the contrary, to invalidate the result. The rest of this section gives the technical detail.

The multivariate models (`calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`) take the observed one-insertion duplications as given and check them against the Fréchet bounds implied by the audiences and, for every triple of vehicles, against joint feasibility. The published methods have three limits that follow from this.

First, Hong (1998, pp. 238-239) formulates, for the models and the data of his study, the hypothesis that the duplication between vehicles is not below the random one, the product of the audiences, and treats null observed duplications as measurement problems. It is not a condition of the data or of the formulas: Danaher's canonical expansion itself admits negative correlations (Danaher, 1991, equation 3), and the logical limits of a duplication are the Fréchet bounds, $\max(0, p_i + p_j - 1) \leq p_{ij} \leq \min(p_i, p_j)$. For example, two vehicles with one insertion each and audiences of 0.30 and 0.20 can have a duplication of 0.03, below the random 0.06, with the valid joint distribution (neither, only the second, only the first, both) of (0.53, 0.17, 0.27, 0.03), and `calc_canex()` reproduces it without negative probabilities. With several insertions per vehicle, in contrast, a lower duplication can make the approximations diverge considerably. The functions warn (class `mediaPlanR_low_duplication`) when the canonical correlation of a pair is below -0.1. That threshold is a decision of this package: the published reference plans (Kim 2005, Appendix B, and Hong 1998, Appendix E; 80 plans) include pairs slightly below the random duplication, with canonical correlations down to -0.071, and they raise no warning. For two vehicles with two

insertions each, `R1 = 0.30` and `R2 = 0.45`, a duplication of zero gives a reach of 0.90 under MSAD, 0.81 under MBD, 0.78 under CSD and CBD, and 0.79 under CANEX. Hong (1998, pp. 238-239) treats a null duplication as a measurement error and replaces it by $\hat{p}_{ij} = K' p_i p_j$, where K' is the ratio between observed and random duplication over the pairs with non-zero observed duplication. The package does not implement this imputation: `calc_agostini_duplication()` and `calc_hofmans_duplication()` estimate reach from a duplication matrix and do not impute its elements. If duplications are imputed beforehand, the result must be checked against the bounds described below. A null duplication is not always a measurement error: it can also be a genuine observation.

Second, pairwise duplications that satisfy the Fréchet bounds can still be impossible jointly for three or more vehicles. For every triple of vehicles the functions check that the exposure to all three can have a non-negative probability, given the audiences and the three pairwise duplications, and stop with an error when it cannot. That condition is necessary but not sufficient for four or more vehicles, for which the published methods check feasibility at most to first order (Cheong 2007); `calc_mbd()` and `calc_cbd()` zero any negative cell that remains, and `calc_canex()` truncates it.

Third, a low duplication between vehicles with several insertions can make the target of a sequential model invalid. For two vehicles with two insertions each, `R1 = 0.5` and `R2 = 0.75`, `calc_csd()` and `calc_cbd()` stop with a duplication of zero, and `calc_msad()` with a duplication of zero or 0.5, instead of altering an invalid target.

References

Danaher, P. J. (1991). A canonical expansion model for multivariate media exposure distributions: A generalization of the duplication of viewing law. Journal of Marketing Research, 28(3), 361-367. doi:10.1177/002224379102800311

Danaher, P. J. (1992). A Markov-chain model for multivariate magazine-exposure distributions. Journal of Business & Economic Statistics, 10(4), 401-407. doi:10.1080/07350015.1992.10509915

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 78-97. Handle 2152/1590 (University of Texas at Austin repository).

See Also

`calc_canex()` for the full-grid canonical expansion, `calc_msad()` for the Morgensztern sequential alternative and `calc_cbd()` for a multivariate model of conditional expansion (not a sequential-aggregation model) that uses the same canonical expansion at the (0,1) level.

Examples

```
data(csd_example)
result <- do.call(calc_csd, csd_example)
result$reach
result$distribution
```

`calc_hofmans_accumulation` *Cumulative reach under Hofmans' accumulation model*

Description

Implements Hofmans' (1966) accumulation model for the cumulative reach (the proportion of the population exposed at least once) of several insertions in one vehicle – the "accumulation" domain in Aldás Manzano's (1998) three-way split. It adapts Agostini's duplication formula by replacing the number of vehicles with the number of insertions, and it needs only the audience after one and two insertions. Hofmans (1966) derived it, together with `calc_hofmans_duplication()`, from the same basis, Agostini's formula, relying on the similarity between N insertions in one vehicle and one insertion in each of N vehicles (Aldás Manzano, 1998, p. 118). They are two applications of that basis to different plan shapes and different data, not interchangeable procedures: this function models the successive insertions of one vehicle, whereas `calc_hofmans_duplication()` models several vehicles with one insertion each, from their own audiences and duplications. In the homogeneous case the two coincide (see Details).

Usage

```
calc_hofmans_accumulation(R1, R2, insertions, R3 = NULL, show_steps = FALSE)
```

Arguments

`R1`	Reach after the first insertion, as a proportion between 0 and 1 (not a percentage), in (0, 1].
`R2`	Cumulative reach after the second insertion, as a proportion (not a percentage) with `R1 < R2 < 2 * R1` and `R2 <= 1`.
`insertions`	Number of insertions in the vehicle up to which the cumulative reach is computed (N in the formulas), an integer of at least 2 (at least 3 when `R3` is supplied).
`R3`	Optional cumulative reach after the third insertion, as a proportion (not a percentage) with `R2 < R3 < 3 * R1` and `R3 <= 1`. It is observed data, like `R1` and `R2`, not something the function estimates: when supplied, the exponent `alpha` of Hofmans' variable coefficient is estimated from it, so that the curve reproduces `R3` (see Details).
`show_steps`	Logical. If `TRUE`, prints the parameters `k`, `d` and `alpha` and the table of cumulative reach.

Details

Let $d = 2R_1 - R_2$ be the duplicated reach of two insertions (the proportion of the population exposed to both), assumed constant for every pair of insertions, and $k = 2R_1/R_2$ the coefficient that makes the formula reproduce R_2 exactly. Cumulative reach after N insertions is

$$R_N = \frac{(NR_1)^2}{NR_1 + k_N\, d\binom{N}{2}}.$$

With a constant coefficient, $k_N = k$ (equation 3.11 of Aldás Manzano, 1998). Hofmans showed that the coefficient is not constant and proposed $k_N = k(N-1)^{\alpha-1}$ (equation 3.12), where the exponent is estimated from a third observation,

$$\hat{\alpha} = \log\left[\frac{(3R_1 - R_3)R_2}{(2R_1 - R_2)R_3}\right] / \log 2.$$

`R3` therefore selects the model: without it, `alpha = 1` and the constant coefficient formula applies; with it, `alpha` is estimated and the curve reproduces `R3` exactly.

The model assumes a constant reach per insertion (`R1`) and a constant duplication between every pair of insertions.

Relation with `calc_hofmans_duplication()`. Take N vehicles with the same audience p and the same duplication d between every pair. The pairwise coefficient of the duplication model is $k_{ij} = 2p/(2p-d)$. With $p = R_1$ and $d = 2R_1 - R_2$, this is $2R_1/R_2 = k$, the constant coefficient above, so that the duplication formula and the constant-coefficient formula of this function give the same reach. For example, $p = 0.30$, $R_2 = 0.45$ ($d = 0.15$) and $N = 4$ give a reach of 0.60 with both functions. The equivalence holds only in that homogeneous case. The two functions take different data (the reach after two insertions of one vehicle, or the audiences and the duplications of several vehicles), and only this one has the variable coefficient adjusted with `R3`.

Valid `R1` and `R2` do not guarantee a valid curve: the extrapolation can decrease, exceed 1 or exceed NR_1. For example, `R1 = 0.60`, `R2 = 0.84` and `N = 4` give a cumulative reach of 1.05 at `N = 4`. The curve is then returned unchanged, with a warning.

Value

A list of class `"reach_hofmans_accumulation"` with components:

- `results`: data frame with `insertions` (number of insertions) and `RN` (cumulative reach, as a proportion).
- `parameters`: list with `k`, `d`, `alpha` and `R3` (`NA` when not supplied).
- `plot`: a ggplot2 object showing the evolution of cumulative reach, or `NULL` when the suggested package ggplot2 is not installed.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.1.1.4 (pages 118-121), equations 3.7-3.12.

Hofmans, P. (1966). Measuring the cumulative net coverage of any combination of media. Journal of Marketing Research, 3(3), 269-278. doi:10.1177/002224376600300307

See Also

`calc_beta_binomial()` for a stochastic accumulation model that also returns the exposure distribution, and `calc_hofmans_duplication()` for the same author's model for several vehicles, derived from the same basis.

Examples

```
# Constant-coefficient model (equation 3.11)
result <- calc_hofmans_accumulation(R1 = 0.06, R2 = 0.103, insertions = 5)
result$results
result$parameters

# With an observed third insertion, Hofmans' exponent is estimated
calc_hofmans_accumulation(R1 = 0.06, R2 = 0.103, insertions = 5,
                          R3 = 0.135)$parameters

# With equal vehicles and equal duplications, the duplication model gives the
# same reach: four insertions of a vehicle with R1 = 0.30 and R2 = 0.45 (in
# people, for a population of 1,000,000) ...
1000000 * calc_hofmans_accumulation(R1 = 0.30, R2 = 0.45,
                                    insertions = 4)$results$RN[4]
# ... and four vehicles of 300,000 people with 150,000 shared by each pair
calc_hofmans_duplication(audiences = rep(300000, 4), population = 1000000,
                         duplication_matrix = matrix(150000, 4, 4))$reach$people
```

`calc_hofmans_duplication`

Reach under Hofmans' duplication model

Description

Implements Hofmans' (1966) *duplication* model: an ad hoc correction of Agostini's (1961) formula for several vehicles with one insertion each – the same "duplication" domain as `calc_agostini_duplication()`. Where Agostini uses one empirical coefficient for every vehicle pair, Hofmans computes a pairwise coefficient directly from each pair's own audiences and duplication, so no coefficient has to be fitted to external data. Like Agostini's formula, it estimates total reach only, not the exposure distribution. Hofmans (1966) derived it, together with `calc_hofmans_accumulation()`, from the same basis, Agostini's formula, relying on the similarity between N insertions in one vehicle and one insertion in each of N vehicles (Aldás Manzano, 1998, p. 118). The other model, `calc_hofmans_accumulation()`, applies that basis to one vehicle with several insertions and needs different data; the two are not interchangeable, although they coincide when the vehicles have the same audience and the same duplication between every pair (see Details).

Usage

```
calc_hofmans_duplication(audiences, population, duplication_matrix)
```

Arguments

`audiences`	Numeric vector with the audience of each vehicle for one insertion, in people.
`population`	Population size, in people. No audience can exceed it; otherwise the function stops with an error.

`duplication_matrix`
: Symmetric numeric matrix whose element `[i, j]` is the number of people who are in the audience of both vehicle `i` and vehicle `j` (one insertion in each), in people (not a proportion). Diagonal values are ignored.

Details

For two vehicles, reach is exactly $A_1 + A_2 - A_{12}$. Equating that value to Agostini's formula gives the coefficient $k_{ij} = (A_i + A_j)/(A_i + A_j - A_{ij})$, which Hofmans applies to every pair of a plan with m vehicles:

$$R_m = \frac{(\sum_{i=1}^{m} A_i)^2}{\sum_{i=1}^{m} A_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} k_{ij} A_{ij}}.$$

The double sum runs over every pair of vehicles, each pair counted once. The gross audience $\sum_i A_i$ counts the contacts of the plan (one insertion per vehicle), and the duplications A_{ij} are coincidences between pairs, so a person present in three vehicles contributes three pairs, not three people: only the reach R_m counts each person once. Keep the units (people) of each term before forming any ratio.

For m vehicles with the same audience A and the same duplication D_0 between every pair, $k_{ij} = 2A/(2A - D_0)$ for every pair, and the formula reduces to the constant-coefficient formula of `calc_hofmans_accumulation()` with $R_1 = A/P$ and $R_2 = (2A - D_0)/P$, for $N = m$ insertions (see the Details of that function). Outside that homogeneous case the two functions are not equivalent.

The formula is empirical and can return a reach outside its logical range (below the largest audience, or above the population or the gross audience). This can happen because the duplications are not mutually consistent, but also because the formula is an approximation: five independent vehicles that each reach 50% of the population, with a duplication of 25% in every pair (data that come from a valid joint distribution, with a true reach of 96.875%), give a reach of 107.14%. The value is then returned unchanged, with a warning, so that an out-of-range result is not attributed automatically to the input data.

Value

A list of class `"reach_hofmans_duplication"` with components:

- `reach`: list with `percent` and `people`.
- `gross_audience`: sum of the vehicle audiences (duplicated people counted once per vehicle), in people.
- `weighted_duplication`: the sum $\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} k_{ij} A_{ij}$, in people.
- `n_vehicles`: number of vehicles.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.2.1.2 (pages 156-157), equations 3.59-3.62.

Hofmans, P. (1966). Measuring the cumulative net coverage of any combination of media. Journal of Marketing Research, 3(3), 269-278. doi:10.1177/002224376600300307

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 44-45. Handle 2152/1590 (University of Texas at Austin repository).

See Also

calc_agostini_duplication(), the single-coefficient formula this model refines pair by pair.

Examples

```
data(duplication_example)
do.call(calc_hofmans_duplication, duplication_example)
```

---

calc_mbd *Multivariate Beta Binomial Distribution model*

Description

Implements Cheong's (2007) Multivariate Beta Binomial Distribution (MBD). Vehicle co-exposure probabilities for pairs (observed) and for three or more vehicles (imputed from a Beta-Binomial fitted to the subset's own mean audience and mean pairwise duplication) are combined into the full 2^m joint zero/one exposure grid by Waring's (1792) inclusion-exclusion theorem. Vehicles are then peeled off one at a time, in reverse aggregation order: each remaining exposure pattern's row is expanded from a zero/one state into the peeled vehicle's own insertion-level Beta-Binomial distribution by a row-specific conditional allocation, and the two component distributions are convolved into a growing pseudo-vehicle.

Usage

```
calc_mbd(
  vehicles_data,
  duplications,
  aggregation_order = c("audience_desc", "given"),
  population = 1,
  tolerance = 1e-08
)
```

Arguments

vehicles_data	Data frame with columns insertions, R1 and R2, with the same convention as calc_csd() (R1 and R2 are proportions between 0 and 1, not percentages, R1 is strictly between zero and one, R1 <= R2 <= 2 * R1 - R1^2, the range of the Beta-Binomial family (see calc_beta_binomial()), and R2 may be NA only when insertions is one; the function stops with an explanatory error if R2 is outside that range). At most 12 vehicles are supported.

`duplications` Symmetric matrix whose element `[i, j]` is the proportion of the population that is exposed to both vehicle `i` and vehicle `j` (one insertion in each), a number between 0 and 1 (not a percentage). It is not the proportion exposed to each vehicle separately. The diagonal is ignored.

`aggregation_order` Either `"audience_desc"` (vehicles in decreasing order of one-insertion reach `R1`, ties in the row order; duplications are not used to sort), `"given"` (the row order), or a permutation of the row indices. Vehicles are peeled off starting from the *last* position of this order, as in Cheong's worked examples. In MBD the order can change both the reach and the exposure distribution (see Details).

`population` Number of people in the population (a count, not a proportion). It only converts probabilities into people: the `people` columns of the result are the probability times `population`, and with the default, 1, they equal the probability.

`tolerance` Positive numerical tolerance for probability constraints. The Fréchet and triple-feasibility checks apply it relative to the smaller audience involved.

Details

`calc_mbd()` (Cheong, 2007) is a different procedure from `fit_bbd_to_reach()`, which fits one Beta-Binomial to an externally given reach. MBD is the *Multivariate* Beta Binomial Distribution of several vehicles described here, whereas `fit_bbd_to_reach()` fits a single Beta-Binomial to a whole schedule.

Value

A `reach_mbd` object: a list with `reach` (`probability`, `percent` and `people`), `average_frequency`, the complete exposure `distribution` (`contacts`, `probability`, `percent`, `people` and `cumulative_probability`), the `aggregation_order` and `aggregation_rule`, the peeling `steps` and `diagnostics`, which report whether the final negative-probability safety net (this package's correction of the final distribution, not Cheong's "MBD-ADJ", which corrects the UD before it is expanded; see Details) had to be engaged and how much probability mass it redistributed (`negative_mass_adjusted`, `cells_adjusted`). They also report the mean number of exposures per person that the inputs imply (`gross_mean_contacts`, the sum over the vehicles of insertions times `R1`), that of the returned distribution (`distribution_mean_contacts`) and their difference (`mean_error`), which can differ from zero even when no cell was adjusted (see Details, "What the result conserves").

What is, and is not, guaranteed

Cheong's dissertation reports that the inclusion-exclusion grid can produce small negative cell probabilities once four or more vehicles are combined, because the co-exposure probability of three or more vehicles is *imputed* (no closed-form multivariate distribution stands behind it) rather than observed. Cheong's four-vehicle example is fully corrected by a first-order consistency check (every combined-exposure probability is shrunk to remain below each of its immediate subsets); the five-vehicle example is not, and the dissertation states that resolving this in general would need increasingly complex higher-order checks that were never implemented, proposing iterative proportional fitting as future work. `calc_mbd()` therefore applies exactly the checks Cheong specifies (first order only) and then, as a package safety net, sets any cell that is still negative in the final collapsed distribution to zero and redistributes that mass proportionally across the non-negative cells.

This is reported in `diagnostics`; a non-zero value means the result for that schedule relies on this fallback rather than on a value Cheong verified as internally consistent without it, and a zero value does not show that the result conserves the marginals or the mean of the inputs (see "What the result conserves" below). The function warns for four or more vehicles, the range where Cheong's own checks are known to be insufficient by themselves.

This safety net is not Cheong's MBD-ADJ. Cheong (2007, Ch. 4.4) applies the same zero-and-redistribute correction to the table that the dissertation calls the UD (univariate distribution; one cell for each pattern of exposure to the `m` vehicles, `2^m` cells), once it is formed and *before* it is expanded and collapsed, so that the later steps work with non-negative values; here the correction acts only on the final distribution, and only if it contains a negative cell. The two procedures can give different results even when the final distribution has no negative cell, so no result of this function should be described as "MBD-ADJ", and the empirical evidence for MBD-ADJ does not transfer to it. The package does not implement the variants MBD, MBD-ADJ and MBD-ADJ2 of the study as such (see the comparison of variants below).

Cheong also reports that the aggregation order can change the collapsed distribution, that this was not investigated systematically, and that the model was evaluated on schedules of 12 vehicles because the computing time of the exposure grid becomes prohibitive above 13 vehicles. `calc_mbd()` stops above 12 vehicles for this reason.

What the result conserves

Each vehicle enters the model with its own parameters (its Beta-Binomial fitted from `R1` and `R2`), but the distribution that the model returns does not necessarily reproduce the marginal distribution of each vehicle or the mean number of exposures of the inputs, and this can happen with no negative probability at all. The reason is the conditional allocation of the peeling step, which is part of the published procedure. When a vehicle is expanded, each row of the joint table is shared between the two expansions $BB(\alpha, \beta+1)$ and $BB(\alpha+1, \beta)$ of that vehicle, with weights $a = 1 - b$ and $b = C(\alpha+\beta+1)-\alpha$, where C is the conditional probability of exposure to the vehicle given the row, obtained from the joint and marginal masses of the table. Cheong (2007, pp. 65-66) does not use C as it comes: he bounds it to the interval $\alpha/(\alpha+\beta+1) \le C \le (\alpha+1)/(\alpha+\beta+1)$ that the Beta-Binomial marginal of the vehicle allows. When the value implied by the observed co-exposures falls outside the interval, replacing it by the nearest end changes the conditional probability that the masses of the table represent, and with it the marginals and the mean of the final distribution. Cheong (2007, p. 90) acknowledges, for a different reason (the imputed co-exposures of three or more vehicles), that the joint distributions of the UD do not always add up to the marginals of each vehicle, and leaves the correction, by iterative proportional fitting, for future research. The mechanism described here is not the final safety net, which acts only on negative cells, and it is not the expansion of `calc_cbd()`, which expands the remaining $n_i - 1$ insertions of each vehicle conditionally on its first exposure.

For example, two vehicles with two insertions each, `R1 = c(0.30, 0.20)`, `R2 = c(0.40, 0.30)` and a duplication of 0.15 between one insertion of each (the example below) are compatible data: a joint distribution of the numbers of exposures of the two vehicles exists with those audiences, reaches after two insertions and duplications, and the duplication (0.15) is above the random one (0.06). The inputs imply $2\times0.30+2\times0.20 = 1$ exposure per person. The distribution returned by `calc_mbd()` has no negative cell (`negative_mass_adjusted` is 0) and a mean of 1.0538, a `mean_error` of +0.0538, because the conditional probability of one of the expansions is bounded from 0.0714 to 0.125 and that of another from 0.75 to 0.6667. The result is a reach of 49.59% and an average frequency of 2.125, whereas the contacts of the input per person reached would be 2.017.

Therefore, always read `gross_mean_contacts`, `distribution_mean_contacts` and `mean_error`, whether or not `negative_mass_adjusted` is zero. The `average_frequency` of the result is the mean of the returned distribution divided by its reach, and it must not be combined with the contacts or the gross rating points of the inputs as if the mean had not changed: the distribution of the example implies about 105.4 rating points, not the 100 of the inputs, and the identity between rating points, reach and average frequency requires both quantities to be computed from the same distribution. With a binomial-limit vehicle the two expansions coincide and the allocation has no effect on that vehicle.

Validation against the published example

For the three-vehicle example of Cheong (2007, p. 75), which Cheong, Leckenby and Eakin (2011, Table 4) also publish, the largest absolute difference between this function and the printed seven-cell distribution is 0.0020348568 in probability (0.20348568 percentage points), in the cell of three exposures. The regression test accepts differences below 0.0025: that is a threshold above the real maximum, not an estimate of the error of the function. The rounding of one printed cell to three decimals (at most 0.0005) is smaller than that maximum. The difference of every cell is recorded in the installed file `DATA-PROVENANCE.md`. This is an approximate replication of one printed example, not an independent validation of the model's predictions. The printed intermediate tables contain arithmetic inconsistencies (a product printed as .024(.057) = .003, which is .0014; a sum of Beta-Binomial parameters of 7.294, where the printed formula gives 7.543; an exclusive-grid sum of 1.001). They are a plausible explanation for some of the differences, but their full causal attribution has not been established: the public tests do not include an independent implementation of the printed procedure. In the binomial limit of a vehicle's own distribution (`R2` equal to the reach under independence), the conditional distributions of the peeling step coincide, so the result of this model does not depend on the duplication between vehicles in that limit; this follows from Cheong's construction.

Published evidence belongs to the study's variants

Of the eleven models Cheong evaluated (comScore 2003 data, 440 schedules; Cheong, 2007, Table 6.2.1), the three versions of MBD were the most accurate for reach alone, but not for the complete exposure distribution. The figures below are the study's, for its own implementations of each version, not measurements of this package:

Model of the study	AER (percent)	APE (percent)
MBD	1.34	10.19
MBD-ADJ	1.40	12.10
MBD-ADJ2	1.18	11.83
CANEX	1.69	6.91
CBD	1.60	8.80

The best result for reach (1.18%) belongs to MBD-ADJ2, not to a plain MBD, and not to the safety net of this function, so it must not be read as the expected error of `calc_mbd()`. For the complete distribution, `calc_canex()` and the Conditional Beta Distribution model (`calc_cbd()`) were more accurate than all the MBD versions.

Computational limits

The work of the schedule is estimated from the insertions before anything is allocated, and a schedule above the budget is refused with an explanation. The budget is the option `mediaPlanR.max_work`, `1e9` elementary operations by default (one to two minutes); see the section "Computational limits" of mediaPlanR.

Domain of validity

In plain words. These models take the overlaps between vehicles (the duplications) from your data and use them to build the whole distribution of exposures. For that to be possible the overlaps you supply must be possible, and plausible together:

- Two vehicles cannot overlap more than the smaller of their audiences, nor less than their audiences force them to (two audiences of 60% of the population must overlap by at least 20%). If a duplication breaks these limits, the function stops with an error naming the pair: the data cannot be right and should be reviewed.
- The overlaps of three vehicles must also fit together. If they cannot, the function stops with an error naming the three vehicles.
- The studies that evaluated these models mostly used vehicles that overlap at least as much as chance alone would produce, that is, at least the product of their audiences (two vehicles with audiences of 30% and 20% overlap by 6% if people chose vehicles independently). A smaller overlap is not impossible and is not necessarily a measurement error: it can be valid data. But when a pair overlaps clearly less than that, the approximations of the models can differ a lot from one model to another, and the function issues a warning that names the pair. Check how the data were defined and read the diagnostics; do not replace valid observations automatically.

An error has one of two causes. Either the inputs break a constraint of the data (impossible overlaps: correct them), or the inputs are possible but the model's approximation cannot reproduce them (the third limit below: a valid probabilistic construction can still be rejected by an approximate model). Read the message to tell them apart, and do not change valid data only to make a model run. A warning does not invalidate the calculation, but it asks for caution: check the duplication of the pair named, which is often a measurement problem (see below), and compare the result with another model. Every diagnostic must be inspected before the result is interpreted; none can be assumed to be harmless or, on the contrary, to invalidate the result. The rest of this section gives the technical detail.

The multivariate models (`calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`) take the observed one-insertion duplications as given and check them against the Fréchet bounds implied by the audiences and, for every triple of vehicles, against joint feasibility. The published methods have three limits that follow from this.

First, Hong (1998, pp. 238-239) formulates, for the models and the data of his study, the hypothesis that the duplication between vehicles is not below the random one, the product of the audiences, and treats null observed duplications as measurement problems. It is not a condition of the data or of the formulas: Danaher's canonical expansion itself admits negative correlations (Danaher, 1991, equation 3), and the logical limits of a duplication are the Fréchet bounds, $\max(0, p_i + p_j - 1) \leq p_{ij} \leq \min(p_i, p_j)$. For example, two vehicles with one insertion each and audiences of 0.30 and 0.20 can have a duplication of 0.03, below the random 0.06, with the valid joint distribution (neither, only the second, only the first, both) of (0.53, 0.17, 0.27, 0.03), and

`calc_canex()` reproduces it without negative probabilities. With several insertions per vehicle, in contrast, a lower duplication can make the approximations diverge considerably. The functions warn (class `mediaPlanR_low_duplication`) when the canonical correlation of a pair is below -0.1. That threshold is a decision of this package: the published reference plans (Kim 2005, Appendix B, and Hong 1998, Appendix E; 80 plans) include pairs slightly below the random duplication, with canonical correlations down to -0.071, and they raise no warning. For two vehicles with two insertions each, `R1 = 0.30` and `R2 = 0.45`, a duplication of zero gives a reach of 0.90 under MSAD, 0.81 under MBD, 0.78 under CSD and CBD, and 0.79 under CANEX. Hong (1998, pp. 238-239) treats a null duplication as a measurement error and replaces it by $\hat{p}_{ij} = K' p_i p_j$, where K' is the ratio between observed and random duplication over the pairs with non-zero observed duplication. The package does not implement this imputation: `calc_agostini_duplication()` and `calc_hofmans_duplication()` estimate reach from a duplication matrix and do not impute its elements. If duplications are imputed beforehand, the result must be checked against the bounds described below. A null duplication is not always a measurement error: it can also be a genuine observation.

Second, pairwise duplications that satisfy the Fréchet bounds can still be impossible jointly for three or more vehicles. For every triple of vehicles the functions check that the exposure to all three can have a non-negative probability, given the audiences and the three pairwise duplications, and stop with an error when it cannot. That condition is necessary but not sufficient for four or more vehicles, for which the published methods check feasibility at most to first order (Cheong 2007); `calc_mbd()` and `calc_cbd()` zero any negative cell that remains, and `calc_canex()` truncates it.

Third, a low duplication between vehicles with several insertions can make the target of a sequential model invalid. For two vehicles with two insertions each, `R1 = 0.5` and `R2 = 0.75`, `calc_csd()` and `calc_cbd()` stop with a duplication of zero, and `calc_msad()` with a duplication of zero or 0.5, instead of altering an invalid target.

References

Cheong, Y. (2007). Multivariate Beta Binomial Distribution Model as a Web Media Exposure Model. Doctoral dissertation, The University of Texas at Austin. Handle 2152/3215 (University of Texas at Austin repository).

Cheong, Y., Leckenby, J. D., & Eakin, T. (2011). Evaluating the multivariate beta binomial distribution for estimating magazine and Internet exposure frequency distributions. Journal of Advertising, 40(1), 7-23. doi:10.2753/JOA00913367400101

Waring, E. (1792). On the principles of translating algebraic quantities into probable relations and annuities, &c. Cambridge.

See Also

`calc_csd()` and `calc_msad()` for sequential aggregation models, and `calc_cbd()` for the model that uses a canonical-expansion between-vehicle step instead of an imputed one.

Examples

```
data(mbd_example)
result <- do.call(calc_mbd, mbd_example)
result$reach
result$distribution
```

```
# Compatible data for which the mean of the result differs from the mean of
# the inputs, with no negative probability (see "What the result conserves")
vehicles <- data.frame(insertions = c(2, 2), R1 = c(0.30, 0.20),
                       R2 = c(0.40, 0.30))
duplications <- matrix(c(0, 0.15, 0.15, 0), nrow = 2)
mbd <- calc_mbd(vehicles, duplications, aggregation_order = "given")
mbd$diagnostics[c("negative_mass_adjusted", "gross_mean_contacts",
                  "distribution_mean_contacts", "mean_error")]
```

`calc_metheringham` *Reach and exposure distribution under the Metheringham model*

Description

Implements Metheringham's (1964) model for a plan with several vehicles and several insertions per vehicle, in the version of Aldás Manzano (1998). Individuals are heterogeneous, with Beta-distributed exposure probabilities; vehicles are treated as homogeneous, so that duplication between different vehicles is equivalent to accumulation within one hypothetical "average" vehicle. The audience and the duplication of that average vehicle are the averages, over all pairs of insertions, of the observed audiences and duplications. They fix the two parameters of a Beta-Binomial distribution for the plan's `N = sum(insertions)` insertions.

Usage

```
calc_metheringham(audiences, insertions, duplication_matrix, population)
```

Arguments

`audiences`	Numeric vector with the audience of each vehicle, in people per insertion.
`insertions`	Non-negative integer vector with the number of insertions planned in each vehicle. The plan as a whole (all vehicles together) must have at least two insertions, for example two vehicles with one insertion each, or one vehicle with two insertions, because the model works with duplication between pairs of insertions.
`duplication_matrix`	Symmetric numeric matrix, in people. Element `[i, j]` with `i != j` is the number of people who are exposed to both one insertion in vehicle `i` and one insertion in vehicle `j`. The diagonal element `[i, i]` is the audience duplicated between two insertions in the same vehicle `i`, that is, `2 * audiences[i]` minus the audience accumulated by two insertions in vehicle `i`. Diagonal elements are needed only for vehicles with at least two insertions, and elements involving a vehicle with no insertions are ignored.
`population`	Population size, in people. No audience can exceed it; otherwise the function stops with an error.

Details

There are `N = sum(insertions)` insertions and `choose(N, 2)` pairs of insertions: `insertions[i] * insertions[j]` pairs between vehicles `i` and `j`, and `choose(insertions[i], 2)` pairs within vehicle `i`. The model computes

1. the insertion-weighted mean audience $\bar{A}_1 = \sum_{i=1}^{m} n_i A_i / N$, with m the number of vehicles;
2. the pair-weighted mean duplication $\bar{D}$, over all pairs of insertions;
3. the mean audience accumulated by two insertions of the average vehicle, $\bar{A}_2 = 2\bar{A}_1 - \bar{D}$, which equals the average, over all pairs of insertions, of the audience reached by the pair;
4. alpha and beta by the Beta-Binomial estimators of `calc_beta_binomial()`, and the exposure distribution for `N` insertions.

The exposure distribution therefore ranges from zero to `N` exposures, and its mean equals the plan's gross number of exposures per person, $\sum_{i=1}^{m} n_i A_i / P$.

When the plan changes, the averaged inputs and hence the Beta-Binomial parameters are fitted again. Reach estimates across such plans need not increase when vehicles or insertions are added: this is the declining reach phenomenon described by Chandon (1976, p. 491). It is a limitation of the approximation. A decrease must not be interpreted as a real loss of people reached by adding an advertising opportunity. This differs from increasing the insertion count while holding the Beta-Binomial parameters fixed, which produces a non-decreasing reach curve.

Value

A list of class `"reach_metheringham"` with components:

- `reach`, `distribution`, `cumulative` and `parameters`, as in `calc_beta_binomial()` evaluated for `N` insertions;
- `mean_audience`: insertion-weighted mean audience (`A1`), in people;
- `mean_duplication`: pair-weighted mean duplication (`D`), in people;
- `second_audience`: mean audience accumulated by two insertions of the average vehicle (`A2`), in people;
- `opportunity_matrix`: number of pairs of insertions between and within vehicles;
- `opportunity_vector` and `duplication_vector`: the upper triangles of the opportunity and duplication matrices, linearized row by row;
- `total_insertions`: the number of insertions `N`.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.3.1.5 (several insertions per vehicle, pages 196-198) and Section 3.2.2.9 (one insertion per vehicle).

Metheringham, R. A. (1964). Measuring the net cumulative coverage of a print campaign. Journal of Advertising Research, 4(4), 23-28. doi:10.1080/00218499.1964.12519751

Chandon, J.-L. (1976). A comparative study of media exposure models. Unpublished doctoral dissertation, Northwestern University, Evanston, IL. Pages 485-491 describe the composite vehicle, its estimation and declining reach.

See Also

calc_beta_binomial(), called internally, for the univariate model; calc_sainsbury() and calc_binomial() for models that assume random duplication.

Examples

```
data(metheringham_example)
result <- do.call(calc_metheringham, metheringham_example)
result$reach$percent
result$parameters$alpha

# One insertion per vehicle reduces to Section 3.2.2.9 of Aldas Manzano
# (1998); the diagonal is not needed then
calc_metheringham(
  audiences = c(300000, 400000, 200000), insertions = c(1, 1, 1),
  duplication_matrix = matrix(c(NA, 150000, 90000,
                                150000, NA, 110000,
                                90000, 110000, NA), nrow = 3),
  population = 1000000
)$reach$percent
```

---

calc_msad *Morgensztern Sequential Aggregation Distribution model*

Description

Implements the MSAD structure described by Leckenby and Rice (1986) and Kim (2005): each vehicle is expanded with a Beta-Binomial marginal, the reach of every sub-schedule is estimated with the Morgensztern formula, and the marginals are combined sequentially with a non-random, margin-preserving convolution whose zero cell equals one minus that reach. Kim specifies the MSAD algorithm but refers readers to Lee (1988) for its complete worked numerical example; Kim's own three-vehicle example is CSD, not MSAD.

Usage

```
calc_msad(
  vehicles_data,
  duplications,
  aggregation_order = c("audience_desc", "given"),
  population = 1,
  tolerance = 1e-10
)
```

Arguments

`vehicles_data`	Data frame with one row per vehicle (at least two rows) and columns `insertions` (planned insertions, a positive integer), `R1` and `R2`. `R1` is the reach after one insertion, a proportion between 0 and 1 (not a percentage) and strictly between zero and one, and `R2` the cumulative reach after two insertions (the proportion exposed at least once in two insertions), a proportion with `R1 <= R2 <= 2 * R1 - R1^2` (the range of the Beta-Binomial family; see `calc_beta_binomial()`). If `R2` is outside that range the function stops with an explanatory error. `R2` may be `NA` only when `insertions` is one.
`duplications`	Symmetric matrix whose element `[i, j]` is the proportion of the population that is exposed to both vehicle `i` and vehicle `j` (one insertion in each), a number between 0 and 1 (not a percentage). It is not the proportion exposed to each vehicle separately. The diagonal is ignored.
`aggregation_order`	Either `"audience_desc"` (Kim's rule of combining the vehicle with the largest audience first: vehicles in decreasing order of one-insertion reach `R1`), `"given"` (the row order), or a permutation of the row indices. The vehicles are combined one at a time in this order. The order does not change the reach of the whole schedule but can change the exposure distribution.
`population`	Number of people in the population (a count, not a proportion). It only converts probabilities into people: the `people` columns of the result are the probability times `population`, and with the default, 1, they equal the probability.
`tolerance`	Positive numerical tolerance for probability constraints. The Fréchet and triple-feasibility checks apply it relative to the smaller audience involved.

Details

For a subset of m vehicles, the Morgensztern reach is

$$R_m = \frac{(\sum_{i=1}^{m} R_{n_i})^2}{\sum_{i=1}^{m} R_{n_i} + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} K_{ij} A_{ij} R_{n_i} R_{n_j} / (A_i A_j)},$$

with $K_{ij} = (A_i + A_j)/(A_i + A_j - A_{ij})$, where R_{n_i} is the reach of vehicle i for its own n_i insertions, A_i its one-insertion audience and A_{ij} the one-insertion duplication of vehicles i and j, all as proportions of the population. The double sum runs over every pair of vehicles. At every aggregation step the joint table is conformed to the two input marginal distributions and to the Morgensztern union reach. Consequently all probabilities remain non-negative, the margins are preserved, and the zero-exposure probability is exactly $1 - R_m$.

The function reproduces the three-vehicle numerical example of Lee (1988, pp. 81-93; two insertions per vehicle, aggregation in the order Lee obtains with his total-duplication criterion, `aggregation_order = "given"` with the vehicles listed in that order) to within 0.0011 in every cell of the exposure distribution, which Lee prints to three decimals.

MSAD is intended to reduce, but does not guarantee the elimination of, the declining reach phenomenon (Leckenby & Rice, 1986): the reach that some exposure-distribution models estimate for a schedule can fall when insertions or vehicles are added, which cannot happen in reality. The reach formula can also be incompatible with the supplied marginal distributions. In that case the function stops and reports the feasible interval instead of silently altering the target. The aggregation order can change the exposure distribution even when the final schedule reach is unchanged.

Value

A reach_msad object: a list with reach (probability, percent and people), average_frequency, the complete exposure distribution (contacts, probability, percent, people and cumulative_probability), the vehicle_marginals, the one-insertion vehicle_reach, the aggregation_order and aggregation_rule, the aggregation steps and numerical diagnostics (see calc_csd() for gross_mean_contacts, distribution_mean_contacts and mean_error).

Computational limits

The cells of the largest joint table (the exposure levels of the vehicles already combined times those of the vehicle added) are counted from the insertions before anything is allocated, and a schedule above the budget is refused with an explanation. The budget is the option mediaPlanR.max_table_cells, 1e7 cells by default (80 MB for each matrix of that size); see the section "Computational limits" of mediaPlanR.

Domain of validity

In plain words. These models take the overlaps between vehicles (the duplications) from your data and use them to build the whole distribution of exposures. For that to be possible the overlaps you supply must be possible, and plausible together:

- Two vehicles cannot overlap more than the smaller of their audiences, nor less than their audiences force them to (two audiences of 60% of the population must overlap by at least 20%). If a duplication breaks these limits, the function stops with an error naming the pair: the data cannot be right and should be reviewed.
- The overlaps of three vehicles must also fit together. If they cannot, the function stops with an error naming the three vehicles.
- The studies that evaluated these models mostly used vehicles that overlap at least as much as chance alone would produce, that is, at least the product of their audiences (two vehicles with audiences of 30% and 20% overlap by 6% if people chose vehicles independently). A smaller overlap is not impossible and is not necessarily a measurement error: it can be valid data. But when a pair overlaps clearly less than that, the approximations of the models can differ a lot from one model to another, and the function issues a warning that names the pair. Check how the data were defined and read the diagnostics; do not replace valid observations automatically.

An error has one of two causes. Either the inputs break a constraint of the data (impossible overlaps: correct them), or the inputs are possible but the model's approximation cannot reproduce them (the third limit below: a valid probabilistic construction can still be rejected by an approximate model). Read the message to tell them apart, and do not change valid data only to make a model run. A warning does not invalidate the calculation, but it asks for caution: check the duplication of the pair named, which is often a measurement problem (see below), and compare the result with another model. Every diagnostic must be inspected before the result is interpreted; none can be assumed to be harmless or, on the contrary, to invalidate the result. The rest of this section gives the technical detail.

The multivariate models (calc_canex(), calc_csd(), calc_msad(), calc_cbd() and calc_mbd()) take the observed one-insertion duplications as given and check them against the Fréchet bounds

implied by the audiences and, for every triple of vehicles, against joint feasibility. The published methods have three limits that follow from this.

First, Hong (1998, pp. 238-239) formulates, for the models and the data of his study, the hypothesis that the duplication between vehicles is not below the random one, the product of the audiences, and treats null observed duplications as measurement problems. It is not a condition of the data or of the formulas: Danaher's canonical expansion itself admits negative correlations (Danaher, 1991, equation 3), and the logical limits of a duplication are the Fréchet bounds, $\max(0, p_i + p_j - 1) \leq p_{ij} \leq \min(p_i, p_j)$. For example, two vehicles with one insertion each and audiences of 0.30 and 0.20 can have a duplication of 0.03, below the random 0.06, with the valid joint distribution (neither, only the second, only the first, both) of (0.53, 0.17, 0.27, 0.03), and `calc_canex()` reproduces it without negative probabilities. With several insertions per vehicle, in contrast, a lower duplication can make the approximations diverge considerably. The functions warn (class `mediaPlanR_low_duplication`) when the canonical correlation of a pair is below -0.1. That threshold is a decision of this package: the published reference plans (Kim 2005, Appendix B, and Hong 1998, Appendix E; 80 plans) include pairs slightly below the random duplication, with canonical correlations down to -0.071, and they raise no warning. For two vehicles with two insertions each, `R1 = 0.30` and `R2 = 0.45`, a duplication of zero gives a reach of 0.90 under MSAD, 0.81 under MBD, 0.78 under CSD and CBD, and 0.79 under CANEX. Hong (1998, pp. 238-239) treats a null duplication as a measurement error and replaces it by $\hat{p}_{ij} = K' p_i p_j$, where K' is the ratio between observed and random duplication over the pairs with non-zero observed duplication. The package does not implement this imputation: `calc_agostini_duplication()` and `calc_hofmans_duplication()` estimate reach from a duplication matrix and do not impute its elements. If duplications are imputed beforehand, the result must be checked against the bounds described below. A null duplication is not always a measurement error: it can also be a genuine observation.

Second, pairwise duplications that satisfy the Fréchet bounds can still be impossible jointly for three or more vehicles. For every triple of vehicles the functions check that the exposure to all three can have a non-negative probability, given the audiences and the three pairwise duplications, and stop with an error when it cannot. That condition is necessary but not sufficient for four or more vehicles, for which the published methods check feasibility at most to first order (Cheong 2007); `calc_mbd()` and `calc_cbd()` zero any negative cell that remains, and `calc_canex()` truncates it.

Third, a low duplication between vehicles with several insertions can make the target of a sequential model invalid. For two vehicles with two insertions each, `R1 = 0.5` and `R2 = 0.75`, `calc_csd()` and `calc_cbd()` stop with a duplication of zero, and `calc_msad()` with a duplication of zero or 0.5, instead of altering an invalid target.

References

Leckenby, J. D., & Rice, M. D. (1986). The declining reach phenomenon in exposure distribution models. Journal of Advertising, 15(3), 13-20. doi:10.1080/00913367.1986.10673014

Morgensztern, A. (1970). Fréquentation de la presse, généralisation, effet publicitaire. Étude No. 19, Institut de Recherches et d'Études Publicitaires (IREP), Paris. Listed and attributed to the reach formula by Kim (2005); the original study was not consulted for this package.

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 65-71. Handle 2152/1590 (University of Texas at Austin repository). Kim identifies the detailed MSAD numerical example as Lee, H.-K. (1988), Sequential

aggregation advertising media models, unpublished doctoral dissertation, The University of Texas at Austin, pp. 81-101; the numerical example itself occupies pp. 81-93.

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.3.2.2, equation 3.106, for the Morgensztern reach formula.

See Also

`calc_csd()` for the canonical alternative, and `calc_cbd()` and `calc_mbd()` for multivariate models that expand each vehicle into insertion-level distributions instead of aggregating the vehicles sequentially.

Examples

```
data(msad_example)
result <- do.call(calc_msad, msad_example)
result$reach
```

`calc_sainsbury` *Reach and exposure distribution under the Sainsbury model*

Description

Implements the unmodified Sainsbury model (random duplication) to calculate reach and the exposure distribution (and its cumulative counterpart) of a set of advertising vehicles. Aldás Manzano (1998, Section 3.2.2.2) attributes its development to the London Press Exchange and its formalization to Caffyn and Sagovsky (1963). The function accepts one or more insertions per vehicle (argument `insertions`). The model assumes random duplication *and* random accumulation, homogeneous individual exposure probabilities and heterogeneous vehicle exposure probabilities. The probability that an individual is exposed to one insertion in vehicle `i` is therefore the ratio between the vehicle's audience and the population. Because duplication and accumulation are both random, every insertion – whether in a different vehicle or a repeat insertion in the same one – is an independent Bernoulli trial with that vehicle's own exposure probability, and the number of exposures follows a Poisson-binomial distribution (the distribution of the number of successes in independent trials whose success probabilities differ).

Usage

```
calc_sainsbury(audiences, population, insertions = rep(1L, length(audiences)))
```

Arguments

`audiences`	Numeric vector with the audience of each vehicle, in people per insertion.
`population`	Population size, in people. No audience can exceed it; otherwise the function stops with an error.

`insertions` Non-negative integer vector with the number of insertions planned in each vehicle. It defaults to one insertion per vehicle, the model's original scope; values above one apply the same independence hypothesis to repeat insertions in the same vehicle.

Details

The exposure distribution is computed exactly, by combining one insertion at a time (a dynamic convolution of the per-insertion Bernoulli distributions). The number of calculations grows with the square of the number of insertions N, so doubling the insertions roughly quadruples the work. In a test on a typical computer, a plan with 1,000 insertions took about 0.01 seconds, one with 5,000 about 0.2 seconds and one with 20,000 about 3 seconds; much larger numbers of opportunities, such as digital impressions counted one by one, need another formulation and are outside the scope of this finite contract, in which each insertion offers at most one opportunity per person. Reach equals $1 - \prod_{i=1}^{m}(1 - A_i/P)^{n_i}$, where m is the number of vehicles, A_i is the audience of vehicle i, n_i its number of insertions and P the population.

Value

A list of class `"reach_sainsbury"` with components:

- `reach`: list with `percent` and `people`.
- `distribution`: list with `percent` and `people`, one value for each number of exposures from one to the total number of insertions.
- `cumulative`: list with `percent` and `people`, for individuals exposed at least once, at least twice, and so on.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.2.2.2 for one insertion per vehicle; Section 3.3.1.2, equations 3.86-3.87, for several insertions (the multivariable binomial model with independent vehicles that Aldás Manzano attributes to Chandon).

Caffyn, J. M., & Sagovsky, M. (1963). Net audiences of British newspapers: A comparison of the Agostini and Sainsbury methods. Journal of Advertising Research, 3(1), 21-25. doi:10.1080/00218499.1963.12527134

Chandon, J.-L. (1976). A comparative study of media exposure models. Unpublished doctoral dissertation, Northwestern University, Evanston, IL. Pages 350-352 describe the random-duplication Sainsbury model; the modified Sainsbury formula on pages 375-380 uses observed duplication and is a different model.

See Also

`calc_binomial()` for the homogeneous-vehicle counterpart, `calc_beta_binomial()` and `calc_metheringham()` for models with heterogeneous individuals.

Examples

```
audiences <- c(300000, 400000, 200000)
population <- 1000000
result <- calc_sainsbury(audiences, population)
result$reach$percent
result$distribution$people

# The same three vehicles with several insertions each
calc_sainsbury(audiences, population, insertions = c(4, 6, 10))
```

`calibrate_bbd` *Calibrate a Beta-Binomial effective-reach model*

Description

Searches for a Beta-Binomial exposure distribution whose effective reach is as close as possible to a target. The reach after one insertion fixes the mean of the Beta distribution, and only its concentration is searched, by a one-dimensional search instead of a two-dimensional grid over both shape parameters. The search returns the closest candidate within the stated bounds; it does not guarantee that the target is reached. Always check `converged`, `predicted_reach` and `error` before using `alpha` and `beta`. The concentration is searched in a finite interval, and the two limits of the family, the binomial and the polarized ones, are evaluated explicitly (see Details).

Usage

```
calibrate_bbd(
  R1,
  target_reach,
  frequency,
  max_insertions,
  type = c("at_least", "exact"),
  tolerance = 1e-06
)
```

Arguments

`R1`	Reach after one insertion (the vehicle's audience), as a proportion (not a percentage) strictly between zero and one.
`target_reach`	Target proportion of the population (not a percentage), strictly between zero and one. With `type = "at_least"` it is the effective reach: the proportion exposed `frequency` times or more. With `type = "exact"` it is the proportion exposed exactly `frequency` times (which is then not a reach in the usual sense).
`frequency`	Positive integer: the number of exposures that defines the target (`frequency` or more exposures if `type = "at_least"`, exactly `frequency` exposures if `type = "exact"`).
`max_insertions`	Maximum number of insertions considered, an integer of at least `frequency`.

type	What target_reach refers to. "at_least" (default): the probability of frequency or more exposures (for frequency = 2, being exposed twice, three times, and so on; this is the effective reach). "exact": the probability of exactly frequency exposures (for frequency = 2, exactly twice).
tolerance	Desired absolute probability error; it only determines the converged flag.

Details

The effective frequency frequency is a threshold chosen by the analyst for the objective, the creative, the medium and the period; the function does not estimate a relation between that number of exposures and recall, persuasion, sales or profit. Reaching the threshold is a delivery measure, not an outcome. It is advisable to examine several thresholds.

For every number of insertions n from frequency to max_insertions, the function searches for the concentration of the Beta-Binomial, with mean R1, whose probability equals target_reach, first on a grid of the log concentration and then by stats::optimize() around the best grid point. It returns the number of insertions with the smallest error, the smaller n in case of ties: the error is the criterion, and neither cost nor the minimum number of insertions is prioritized except to break ties. When the target cannot be reached, converged is FALSE and the closest candidate is returned.

The search interval is finite: the natural logarithm of the concentration alpha + beta is searched between -15 and 15, that is, concentrations from about 3e-7 to 3.3e6. The two limits of the family that calc_beta_binomial() also represents, the polarized limit (concentration tending to zero: either all of the insertions or none) and the binomial limit (concentration tending to infinity: independent insertions), are not points of that interval, so they are evaluated explicitly as one more candidate for every n of at least two. A limit is returned only when its error is smaller, by more than 1e-12, than that of the best candidate of the interval, so that the results that the interval already resolves do not change. In that case alpha, beta and concentration are 0 (polarized) or Inf (binomial), as in calc_beta_binomial(), where they indicate the limit and are not the parameters of a proper Beta distribution, and fit_type says which one it is. A target is outside the family only when it is farther than the limits allow (for example, an effective reach above the binomial limit with that number of insertions); a target merely beyond the ends of the interval receives the nearest of the limit and the interval. For example, with R1 = 0.3, target_reach = 0.99, frequency = 1 and max_insertions = 1 the only candidate has a probability of 0.3, so predicted_reach is 0.3, error is 0.69 and converged is FALSE.

Value

A bbd_calibration object: a list with alpha, beta, concentration, insertions (the selected n), R1, frequency, type, target_reach, predicted_reach, error, converged, fit_type ("beta_binomial", "binomial_limit" or "polarized_limit"), the full distribution (contacts, probability and cumulative_probability) and the candidates evaluated for each n (the selected one for that n, with its log_concentration, which is Inf or -Inf for a limit, its predicted probability, its error and its fit_type).

References

Metheringham, R. A. (1964). Measuring the net cumulative coverage of a print campaign. Journal of Advertising Research, 4(4), 23-28. doi:10.1080/00218499.1964.12519751

Chandon, J.-L. (1976). A comparative study of media exposure models. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

See Also

`calc_beta_binomial()` for the model itself and `fit_bbd_to_reach()` to fit a Beta-Binomial to a whole schedule's reach.

Examples

```
# A first insertion reaches 30% and the share exposed two or more times
# should be 20%
calibrate_bbd(R1 = 0.30, target_reach = 0.20, frequency = 2,
              max_insertions = 6)

# A target that cannot be reached: the closest candidate is returned and
# `converged` is FALSE
unreachable <- calibrate_bbd(R1 = 0.30, target_reach = 0.99, frequency = 1,
                             max_insertions = 1)
unreachable$converged
unreachable$error
```

---

canex_example *Illustrative inputs for the CANEX model*

Description

List of arguments ready for `do.call()` with `calc_canex()`: two vehicles with three and two insertions. Original illustrative data, not derived from any published source.

Usage

```
canex_example
```

Format

A list with the components:

vehicles_data Data frame with columns `insertions`, `R1` and `R2` (reach after one and two insertions, as proportions between 0 and 1).

duplications Symmetric matrix of proportions of the population exposed to both vehicles of each pair (one insertion in each).

population Population size, in people.

Examples

```
data(canex_example)
do.call(calc_canex, canex_example)
```

csd_example *Illustrative inputs for the CSD model*

Description

List of arguments ready for `do.call(calc_csd, csd_example)`: a small three-vehicle scenario. Original illustrative data, not derived from any published source.

Usage

```
csd_example
```

Format

A list with the components:

vehicles_data Three rows with columns `insertions`, `R1` and `R2`.

duplications Symmetric matrix of proportions of the population exposed to both vehicles of each pair (one insertion in each).

aggregation_order The order of aggregation, `1:3`.

See Also

`calc_csd()`, msad_example

Examples

```
data(csd_example)
result <- do.call(calc_csd, csd_example)
result$distribution
```

duplication_example *Illustrative inputs for the Agostini and Hofmans duplication models*

Description

List of arguments ready for `do.call()` with `calc_agostini_duplication()` and `calc_hofmans_duplication()`, the two ad hoc duplication models for several vehicles with one insertion each. Original illustrative data, not derived from any published source.

Usage

```
duplication_example
```

Format

A list with the components:

audiences Numeric vector with the audience of each of three vehicles, in people per insertion.

population Population size, in people.

duplication_matrix Symmetric matrix whose element `[i, j]` is the number of people in the audience of both vehicle `i` and vehicle `j`, in people. The diagonal is not used.

See Also

`calc_agostini_duplication()`, `calc_hofmans_duplication()`, hofmans_accumulation_example

Examples

```
data(duplication_example)
do.call(calc_agostini_duplication, duplication_example)
do.call(calc_hofmans_duplication, duplication_example)
```

---

evaluate_exposure_model

Evaluate an observed exposure distribution against a model prediction

---

Description

Computes Kim's relative reach error and exposure-distribution error for one or more schedules, together with model-agnostic diagnostics. Observed data are always supplied by the analyst; the function never treats model inputs or fitted values as observations.

Usage

```
evaluate_exposure_model(
  observed,
  predicted,
  observed_scale,
  predicted_scale = c("auto", "count", "probability", "percent"),
  schedule_col = NULL,
  sum_tolerance = 0.005,
  censored_last_level = FALSE
)
```

Arguments

`observed`	Data frame with columns `contacts` and `observed`. It must include an explicit zero-exposure row and every integer exposure level up to its maximum. For several schedules, also include the column named by `schedule_col`.

`predicted` Either a supported model result whose `distribution` is a data frame with `contacts` and `probability` or `percent` columns (for example the results of `calc_csd()`, `calc_msad()`, `calc_cbd()`, `calc_mbd()` and `calc_canex()`), or a data frame with columns `contacts` and `predicted`. Data-frame predictions must use the same schedule column when `schedule_col` is supplied.

`observed_scale` Required declaration of the `observed` column: `"count"`, `"probability"`, or `"percent"`. Counts may be weighted and are normalized within schedule.

`predicted_scale`
Scale of a predicted data frame. The default `"auto"` is allowed only for supported model objects, whose scale is known.

`schedule_col` Optional name of a schedule identifier column present in both data frames. Model objects represent one schedule and therefore require `schedule_col = NULL`.

`sum_tolerance` Maximum absolute departure from probability mass one accepted for the total of each probability/percent distribution (in probability units, so 0.005 is half a percentage point). The default tolerates small deviations compatible with publication rounding and rejects total-mass discrepancies larger than the stated threshold. It cannot distinguish rounding from a small omitted tail: the analyst must establish that both tables cover the same complete support, or identically collapsed tails, before interpreting the comparison. The tolerance applies to the total only: no individual probability may exceed one (or percentage exceed 100), whatever the tolerance. Masses within the tolerance are used as supplied and are not renormalized (see Details); counts, in contrast, are always normalized.

`censored_last_level`
`FALSE` (default) when the last exposure level of both distributions is an exact count. Set it to `TRUE` when it is a collapsed tail ("k or more exposures", as when a Negative-Binomial table has been truncated at `k` identically in observed and predicted data). The exact mean number of exposures is then not available: the observed and predicted means and `mean_contact_bias` are `NA`, and the means of the censored codes (the last level counted as exactly `k`) are reported in separate columns instead.

Details

For schedule `i`, the function calculates

$$e_{R,i} = |R_i^{obs} - R_i^{pred}|/R_i^{obs}$$

and

$$e_{P,i} = \sum_{j \geq 1} |p_{ij}^{obs} - p_{ij}^{pred}|/R_i^{obs}.$$

where R_i is the reach (one minus the share with zero exposures) and p_{ij} the share exposed exactly j times. `kim_aer` (Kim's average percentage error in reach, AER) and `kim_ape` (Kim's average percentage error in the exposure distribution, APE) are the means of $e_{R,i}$ and $e_{P,i}$ across schedules, reported here as proportions (multiply by 100 for percentages). They are descriptive predictive-error measures, not inferential tests.

Kim (2005, p. 118) writes the denominator of $e_{P,i}$ as the sum of the observed probabilities of one or more exposures, and this function divides by the observed reach R_i^{obs}, one minus the zero cell.

The two are identical when the observed distribution has total mass one, which is always the case for counts (they are normalized) and for probability or percent tables that add up to one. They can differ for a rounded table that is used without renormalizing (see below): with observed cells 0.500 and 0.496 (mass 0.996) and predicted cells 0.800 and 0.200, the error is 0.296 / 0.500 = 0.592 here, whereas dividing by the sum of the observed positive cells would give 0.296 / 0.496 = 0.597. The function keeps the definition by the reach, and `input_mass` reports the supplied totals.

The other elements of `summary` are, averaged over schedules:

- `mean_total_variation`: half the sum over all exposure levels (including zero) of the absolute difference between observed and predicted probabilities;
- `mean_cell_mae`: the mean absolute difference per exposure level;
- `mean_reach_absolute_error`: $|R_i^{obs} - R_i^{pred}|$;
- `mean_contact_bias`: predicted minus observed mean number of exposures per person (positive when the model overestimates exposures).

Exact support equality is required. Open-tail NBD output is rejected because a cell such as `10+` is not equivalent to an exact ten-exposure cell. To evaluate a censored distribution, the analyst must first collapse observed and predicted tails identically, provide explicit data frames and set `censored_last_level = TRUE`. The mean number of exposures of such a table treats the last cell as exactly `k` exposures, so it understates the mean of the process (a Negative-Binomial with mean 2, size 1 and a last level of `3+` has a table mean of about 1.41); that is why the exact mean and the bias are then reported as `NA`. The AER and APE of the collapsed cells remain valid comparisons of the collapsed distributions, not of the exact ones.

The observed reach must be positive: the relative errors divide by it, so a schedule whose observed mass is entirely at zero exposures stops with an error, even though the absolute distances could be computed. Because the reach is one minus the zero cell, it is resolved only down to the double-precision epsilon (about 2.2e-16): any observed reach above that is accepted, including one person reached in a universe of a billion.

Masses are used as supplied. A probability or percent distribution whose total is within `sum_tolerance` of one is not renormalized, so its reach is one minus its zero cell: a predicted table with cells 0.5 and 0.496 (total 0.996) has a predicted reach of 0.5, and the other metrics are those of the rounded tables, not exactly of two normalized distributions; `input_mass` reports the supplied totals. This keeps the replication of published tables, which carry rounding, unaltered. Counts are always normalized. A total within `sum_tolerance` of one does not show that a table is complete: a mass of 0.996 can be rounding or a small omitted tail, and having every level up to the declared maximum does not show that no posterior tail was omitted. The total only rejects larger deficits. Whether both tables cover the same complete support, or identically collapsed tails, is a property of how the data were prepared and is the responsibility of the analyst.

Value

An `exposure_model_evaluation` object, a list with:

- `summary`: `schedules`, `kim_aer`, `kim_ape` (proportions), `mean_total_variation`, `mean_cell_mae`, `mean_reach_absolute_error` (a difference between proportions: multiply by 100 for percentage points) and `mean_contact_bias` (exposures per person). With `censored_last_level = TRUE`, `mean_contact_bias` is `NA` and `mean_censored_contact_bias` is added.

- by_schedule: one row per schedule. The reaches are observed_reach and predicted_reach. The errors in reach are reach_signed_error (predicted minus observed), reach_absolute_error and kim_relative_reach_error. The errors in the distribution are kim_distribution_error, total_variation and cell_mae. The mean numbers of exposures are observed_mean_contacts and predicted_mean_contacts. When the last level is censored, the columns observed_censored_mean and predicted_censored_mean are added. All are proportions or exposures per person.
- aligned_distribution: the observed and predicted probabilities by schedule and exposure level, after normalizing counts and converting percentages to probabilities.
- input_mass: the total mass of each supplied distribution before any normalization, by schedule (observed_input_mass, predicted_input_mass).
- inputs: the scales, the source of the prediction, sum_tolerance and censored_last_level.

References

Kim, H. G. (2005). A Canonical Sequential Aggregation Media Model. Doctoral dissertation, The University of Texas at Austin, pp. 117-118. Handle 2152/1590 (University of Texas at Austin repository).

Examples

```
# Observations must be supplied explicitly. These counts are invented for
# the example (they are not real data); there is one row per exposure level,
# from zero to the total number of insertions, and zero exposures included.
data(csd_example)
csd <- do.call(calc_csd, csd_example)
observed <- data.frame(
  contacts = 0:9,
  observed = c(520, 150, 130, 80, 50, 40, 20, 5, 3, 2)
)
predicted <- data.frame(
  contacts = 0:9,
  predicted = 100 * csd$distribution$probability
)
evaluation <- evaluate_exposure_model(
  observed, predicted,
  observed_scale = "count", predicted_scale = "percent"
)
evaluation$summary[c("kim_aer", "kim_ape")]

# A fitted model object can be passed directly as the prediction.
evaluate_exposure_model(observed, csd, observed_scale = "count")
```

---

fit_bbd_to_reach *Fit a Beta-Binomial distribution to an external reach estimate*

Description

Fits one Beta-Binomial exposure distribution to a whole schedule so that it preserves the insertion-weighted mean exposure probability and reproduces a schedule reach supplied by the analyst. The reach can come from an ad hoc formula such as `calc_hofmans_duplication()` or the Morgensztern formula, or from any other independently justified estimator. This is the "estimation by mean and zeros" step of the Hofmans Beta-Binomial (HBBD) procedure, which starts from the reach of `calc_hofmans_duplication()` (see the section "Hofmans Beta-Binomial (HBBD)" and the example); it is not the Morgensztern Sequential Aggregation Distribution (`calc_msad()`).

Usage

```
fit_bbd_to_reach(
  insertions,
  audiences,
  reach,
  population,
  precision = NULL,
  max_iter = 100
)
```

Arguments

`insertions`	Numeric vector with the number of insertions of each vehicle (positive integers).
`audiences`	Numeric vector with the audience of each vehicle, in people per insertion.
`reach`	External schedule reach, in people.
`population`	Population size, in people. No audience can exceed it; otherwise the function stops with an error.
`precision`	Convergence criterion, in people: the fit is `converged` when the fitted and the external reach differ by at most this many people. The default, `NULL`, is `1e-4 * population` (100 people for a universe of one million). It is expressed in people, so it must be adapted to the scale of `population`; with `population = 1`, for example, a value of 100 would accept any reach, which is why the default is relative. It measures the fit to the target, and it is also the margin by which a reach may lie outside the interval of the Beta-Binomial family with this mean (see below). It never widens the bounds of the schedule: those are checked with the rounding error of double arithmetic only, so a reach one person below the largest audience, or above the population or the total number of contacts, is rejected whatever `precision` is.
`max_iter`	Maximum number of iterations of the root finder.

Details

When it is useful: several models (`calc_agostini_duplication()`, `calc_hofmans_duplication()`, the Morgensztern formula) and many real sources (a panel, a syndicated survey, planning software) give only the reach of a schedule, not how many times people are exposed. This function completes the picture: it finds the Beta-Binomial exposure distribution with that reach and the mean number of exposures implied by the audiences, so that the frequency distribution, the average frequency and

the effective reach can be read from it. It is a bridge from a reach-only figure to a distribution, not a model of duplication: the result is only as good as the reach supplied and as the assumption that the Beta-Binomial shape suits the schedule.

With n_i insertions of audience A_i in vehicle i, the insertion-weighted mean exposure probability is $\bar{p} = \sum_{i=1}^{m} n_i A_i / (P \sum_{i=1}^{m} n_i)$, where m is the number of vehicles and P is the population (with the audiences A_i in people). The Beta-Binomial has $N = \sum_{i=1}^{m} n_i$ trials and shape parameters $\alpha = c\bar{p}$ and $\beta = c(1 - \bar{p})$, so its mean is preserved for every concentration $c > 0$. The function solves, with `stats::uniroot()` on the log scale of c, for the concentration whose reach, $1 - \Pr(K = 0)$, equals the external reach. The initial log-concentration interval, `[-30, 20]`, is expanded as needed to bracket an interior solution, up to `[-700, 700]`. If double precision cannot bracket the target even there, the function stops with an explanatory numerical-resolution error.

Two different conditions are checked, in this order. First, the external reach must be logically possible for the schedule: at least the largest audience of a vehicle (one insertion of that vehicle alone reaches those people) and at most the smaller of the population and the total number of contacts, $\sum_i n_i A_i$. A reach outside these bounds cannot occur whatever the distribution, and the function stops. Second, the reach that a Beta-Binomial with this mean can take lies between the polarized limit ($c \to 0$), $\bar{p}P$, and the binomial limit ($c \to \infty$), $(1 - (1 - \bar{p})^N)P$. A reach that is possible for the schedule but outside that interval, beyond `precision`, is not representable by this family and is also rejected, with a different message. That interval, returned as `feasible_reach`, belongs to the family with this mean and ignores the audiences of the vehicles: the interval that the schedule can actually use is its intersection with the bounds of the first condition, which are returned as `schedule_bounds`.

The fitted reach is also checked against the bounds of the schedule before it is returned. A limit is chosen only when the target lies at or beyond the reach that the family attains at that end of the concentration range, and only if that limit is itself inside the bounds of the schedule; it is not chosen merely because it is within `precision` of the target. When no Beta-Binomial with this mean is compatible with the schedule, the function stops instead of returning an impossible fit. For two insertions of audiences of 50 and 10 people in a universe of 100, for example, the family covers 30 to 51 people, the schedule 50 to 60, and a reach of 50 is fitted with a reach of 50, not with the polarized limit of 30. At either limit the distribution is computed directly (all the mass at zero and at `N` in the polarized limit; a Binomial(`N`, `mean_probability`) in the binomial limit) and `alpha` and `beta` are reported as `0` or `Inf`. These values only flag the limit, as `fit_type` does: they are not parameters of a proper Beta distribution (the quotient `alpha / (alpha + beta)` is undefined there), and the mean stays fixed at `mean_probability`.

Fitting one distribution to the aggregate reach does not recover the exposure marginals of each vehicle or the duplication between vehicles, and `converged = TRUE` shows only that the numerical fit is within `precision` of the reach supplied, not that the reach is coherent with the individual vehicles beyond the bounds above.

Aldás Manzano (1998) starts the original iterative procedure from an arbitrary initial value of `alpha`; root finding needs no starting value.

Value

A list of class `"bbd_reach_fit"` with components:

- `parameters`: list with `alpha`, `beta`, `N` (total insertions), `m` (number of vehicles), `population`, `iterations`, `converged`, `fit_type` (`"beta_binomial"`, `"polarized_limit"` or `"binomial_limit"`), `mean_probability`, `feasible_reach` (the interval of reaches that a Beta-Binomial with this

mean can take, in people) and `schedule_bounds` (the logical bounds of the schedule, in people; the reach that can be used lies in the intersection of both).

- `reach`: list with the `external` reach, the `fitted` Beta-Binomial reach and their `difference`, in people.
- `distribution`: data frame with `contacts` (0 to `N`), `probability` and `cumulative_probability`.
- `iteration_history`: data frame with the evaluations of the root finder.

Hofmans Beta-Binomial (HBBD)

Applied to the reach of `calc_hofmans_duplication()`, this function completes the Hofmans Beta-Binomial (HBBD) procedure of Leckenby and Boyd, described in Aldás Manzano (1998, Section 3.3.1.10) and Kim (2005, pp. 52-53). It has two steps:

1. the reach of the schedule comes from Hofmans' formula, which needs the audience of each vehicle and the audience duplicated between each pair of vehicles (`calc_hofmans_duplication()`);
2. a Beta-Binomial with the mean exposure implied by the audiences and that reach gives the whole frequency distribution (`fit_bbd_to_reach()`).

The package does not wrap the two steps in one function; the example below composes them. Three points to keep in mind. (1) Hofmans' duplication formula takes one insertion in each vehicle, so `insertions` is `rep(1, m)`. (2) The reach is passed in people, as it comes out of step 1. (3) Hofmans' formula is empirical and can return a reach outside the logical range of the schedule, with a warning; this function then stops instead of fitting an impossible reach. The composition follows the published description of the procedure but has not been validated against a published numeric example.

References

Aldás Manzano, J. (1998). Modelos de determinación de la cobertura y la distribución de contactos en la planificación de medios publicitarios impresos (Models for determining reach and exposure distribution in print media planning). Doctoral dissertation, Universidad de Valencia, Spain. Section 3.3.1.10 (the Hofmans Beta-Binomial method of Leckenby and Boyd, pages 208-209) and Section 3.3.2.2 (the Morgensztern reach formula).

Leckenby, J. D., & Boyd, M. M. (1984). An improved beta binomial reach/frequency model for magazines. Current Issues and Research in Advertising, 7(1), 1-24.

See Also

`calc_hofmans_duplication()` and `calc_agostini_duplication()`, whose reach this function can complete with a distribution. For models that estimate the exposure distribution without an external reach, see `calc_sainsbury()` and `calc_binomial()` (they need only audiences, population and insertions), `calc_beta_binomial()` (also needs the reach after two insertions) and `calc_metheringham()` (also needs the observed duplications).

Examples

```
data(bbd_reach_example)
fit <- do.call(fit_bbd_to_reach, bbd_reach_example)
fit
```

```
# Hofmans Beta-Binomial (HBBD): from a reach-only formula to an exposure
# distribution, in two steps (three vehicles, one insertion each)
data(duplication_example)

# Step 1: reach of the plan from the audiences and the pairwise duplications
hofmans <- do.call(calc_hofmans_duplication, duplication_example)
hofmans$reach$people

# Step 2: Beta-Binomial with that reach and the mean implied by the audiences
hbbd <- fit_bbd_to_reach(
  insertions = rep(1, length(duplication_example$audiences)),
  audiences = duplication_example$audiences,
  reach = hofmans$reach$people,
  population = duplication_example$population
)
hbbd$parameters$converged  # TRUE: the fitted reach equals the Hofmans reach
hbbd$distribution          # probability of 0, 1, 2 and 3 contacts
```

`fit_nbd_exposure` *Fit a Negative-Binomial exposure-count model*

Description

Estimates the mean and Gamma heterogeneity parameter from observed person- level exposure counts by maximum likelihood. When the data are not overdispersed relative to Poisson, the fit returns the Poisson boundary.

Usage

```
fit_nbd_exposure(counts, report_max = NULL, conf_level = 0.95)
```

Arguments

`counts`	Vector of observed non-negative integer exposure counts, one per person (not the frequencies of a histogram), for a sample of the whole universe and period of interest that includes the people with zero exposures. A sample made only of exposed people, such as the records of an ad server, is truncated at zero, and fitting the ordinary Negative-Binomial to it estimates another distribution and does not recover the reach of the universe; the zero-truncated model is not implemented. At least two counts, at least one of them positive, are required. Observations should be comparable in exposure period and unweighted: the likelihood treats the people as an independent sample with equal weight, without survey weights or dependence between observations.
`report_max`	Optional open-tail threshold passed to `nbd_exposure_distribution()`.
`conf_level`	Confidence level of the Wald interval for `size`, computed on the log scale of `size`, when an interior Negative-Binomial solution exists.

Details

It is meant for data in which each person has a count of exposures that is not limited by a number of insertions, such as page views or ad-server impressions per person in a period. From those counts it estimates the mean and the heterogeneity between people, and from them the reach and the frequency distribution, which can be extended to scenarios with `nbd_exposure_distribution()`. It is not used for plans with a fixed number of insertions.

Value

An `nbd_exposure_fit` object, a list with:

- `estimates`: `mean_contacts` (the sample mean), `size` (`Inf` at the Poisson boundary), `size_interval` (a Wald interval on the log scale of `size`, `NA` at the boundary or where the curvature is not usable; it describes the uncertainty of `size` only, not that of the reach, and is not a prediction interval) and `variance` (of the fitted distribution).
- `diagnostics`: `n`, `observed_variance`, `dispersion_index` (variance over mean; one for a Poisson), `poisson_boundary` (`TRUE` when the data show no overdispersion and the Poisson solution is returned, so that the implied `size` is infinite), `converged` (the likelihood was maximized inside the search range; it is not a test of goodness of fit), `log_likelihood` and `AIC`.
- `observed`: data frame with `contacts`, `count` (number of people) and `proportion`, for 0 up to the maximum observed count.
- `fitted_distribution`: the `nbd_exposure_distribution()` object of the fitted parameters.

References

Danaher, P. J. (2007). Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction. Marketing Science, 26(3), 422-437. doi:10.1287/mksc.1060.0226

Ehrenberg, A. S. C. (1959). The Pattern of Consumer Purchases. Applied Statistics, 8(1), 26-41. doi:10.2307/2985810

Examples

```
counts <- c(rep(0, 40), rep(1, 25), rep(2, 15), rep(3, 8), 5, 7)
fit <- fit_nbd_exposure(counts)
fit$estimates
```

---

hofmans_accumulation_example

Illustrative inputs for the Hofmans accumulation model

Description

List of arguments ready for `do.call()` with `calc_hofmans_accumulation()`. One vehicle with several insertions: the "accumulation" domain. For the other Hofmans model (several vehicles, one insertion each), see duplication_example. Original illustrative data, not derived from any published source.

Usage

```
hofmans_accumulation_example
```

Format

A list with the components:

R1 Reach after the first insertion, as a proportion.

R2 Cumulative reach after the second insertion, as a proportion.

insertions Number of insertions up to which the cumulative reach is calculated.

See Also

`calc_hofmans_accumulation()`, duplication_example

Examples

```
data(hofmans_accumulation_example)
do.call(calc_hofmans_accumulation, hofmans_accumulation_example)
```

---

mbd_example *Illustrative inputs for the MBD model*

Description

List of arguments ready for `do.call(calc_mbd, mbd_example)`: a small two-vehicle scenario. Original illustrative data, not derived from any published source. Two vehicles avoid the Beta-Binomial imputation of co-exposure that three or more vehicles require.

Usage

```
mbd_example
```

Format

A list with the components:

vehicles_data Two rows with columns `insertions`, `R1` and `R2`.

duplications Symmetric matrix of proportions of the population exposed to both vehicles of each pair (one insertion in each).

aggregation_order The order of aggregation, `1:2`.

See Also

`calc_mbd()`

Examples

```
data(mbd_example)
result <- do.call(calc_mbd, mbd_example)
result$distribution
```

`mediaPlanR` *mediaPlanR: Reach and Exposure-Distribution Models for Media Planning*

Description

Reproducible implementations of the classical reach and exposure-distribution models of the media-planning literature, with tools to fit them to observed reach and to evaluate their predictions against observed distributions.

Scope of the validation

The models can be applied to any mix of media, but the published evidence on their accuracy comes mainly from magazine and Internet data. Their accuracy for a particular combination of media and population should be checked against observed data, for example with `evaluate_exposure_model()`. The published evidence for Internet data (Cheong, Leckenby and Eakin, 2011) takes a website as the vehicle and a week as the measurement interval, and counts several visits within the week as one exposure: see "Unit of exposure" in the section "Notation and units".

Three kinds of evidence must not be confused. *Replication of a published example or table*: in the maintainer's local tests, whose input tables are not distributed, the package reproduces, within 0.01 percentage points, the CBD columns of the 80 two-vehicle plans of Kim (2005) and Hong (1998), but those sources do not print the duplications, which were reconstructed from another model's published column; this is a consistency check between columns, not an independent test. *Fidelity to the printed algorithm*: the functions follow the procedures described in the sources, and where the printed numbers cannot be reproduced the difference is recorded in the tests. *Predictive validation*: agreement with independent observations of reach and frequency. The package does not provide it and the replicated tables do not supply it; use `evaluate_exposure_model()` with your own data. The provenance of each dataset is recorded in the installed file `DATA-PROVENANCE.md` (see `system.file("DATA-PROVENANCE.md", package = "mediaPlanR")`).

Notation and units

The same quantity has the same meaning in every function. Names follow the sources of each model only where the literature fixes them.

- *Population* (number of people in the planning universe): `population` everywhere.
- *Audience* (people exposed by one insertion of a vehicle): `audiences` (one value per vehicle) in the `calc_*()` functions, always in people. As a proportion of the population it is `R1`.
- *Reach after the first and second insertion of the same vehicle*: `R1` and `R2` (Kim, 2005, and the sources of each model) are proportions of the population in `calc_beta_binomial()`, `calc_hofmans_accumulation()`, `calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()`

and `calc_mbd()`. `calibrate_bbd()` uses `R1` but does not take `R2` as an input. `R2` is cumulative: the proportion exposed at least once in two insertions. If the audience is known in people, divide it by the population.

- *Number of insertions*: `insertions` (one value per vehicle) in all functions; the formulas use n_i for a vehicle and N for the total. The column of `vehicles_data` in `calc_canex()` may also be named `k`, Danaher's notation. The argument `k` of `calc_agostini_duplication()` is something else: Agostini's duplication coefficient.
- *Duplication between two vehicles*: the argument is `duplication_matrix`, in people, for `calc_agostini_duplication()`, `calc_hofmans_duplication()` and `calc_metheringham()`, and `duplications`, as proportions of the population, for `calc_canex()` and the sequential models. The different names signal the different units.
- *Reach* (also net audience) counts each person once. *Contacts* (also impressions, or gross audience in the duplication models) count a person once per opportunity of exposure, that is, once per insertion (see "Unit of exposure" below): with several insertions they are the sum of each audience times its insertions, $I = \sum n_i A_i$; with one insertion per vehicle they are the sum of the audiences, the *gross audience* of `calc_agostini_duplication()` and `calc_hofmans_duplication()`. *Duplications* between vehicles are coincidences between pairs, so a person present in three vehicles contributes three pairs, not three people. Write the unit (people, contacts, pairs or proportion) before forming a ratio. With these symbols, the average frequency is $F = I/R$ (contacts per person reached, defined only if $R > 0$) and, per person of the universe, the mean number of contacts is $\mu = I/P$, so that $GRP = 100\mu$.
- *Unit of exposure*: the models count opportunities of exposure, and the data define what an opportunity is. An audience is the number of people exposed by one insertion (an issue, a spot, a post, the visits to a site in a week), with at most one opportunity per person and insertion, and the audiences, `R1`, `R2` and the duplications must measure the same population, the same unit of exposure and the same period. Contacts are those opportunities, and they can be read as modeled advertising impressions only when the audiences and duplications measure that same opportunity. The audience of a vehicle does not show by itself that an advertisement was served, rendered, viewable or noticed; if the data count weekly visitors of a website, the reach is a reach of those visitors, not of viewable advertisements. An effective-frequency threshold is a delivery criterion of the analyst, not a measure of attention, conversion or causal effect.
- *Effective frequency*: `frequency` in `calibrate_bbd()`; effective reach is the proportion exposed that many times or more.
- *Units*: every argument documents its unit. A *proportion* is a number between 0 and 1 (0.35 for 35%, not 35); percentages appear in result columns named `percent` (and `reach_percent`); counts of people appear in arguments described as "in people" and in result columns named `people`. There is one exception, listed in the help page of its function: `zero_contact_probability` in `calc_beta_binomial()` is a percentage (0 to 100) despite its name.
- *Errors of evaluation*: the average percentage errors of `evaluate_exposure_model()` (`kim_aer`, `kim_ape`) are relative errors returned as proportions (multiply by 100 for percent); absolute errors of reach are differences between proportions (multiply by 100 for percentage points, so 0.002 of probability is 0.2 percentage points).

Assumptions of the models

Each model rests on assumptions about four things. They are the first thing to check when choosing a model.

- *Individuals*: homogeneous (every person has the same probability of being exposed to an insertion of a vehicle) or heterogeneous (probabilities differ between people, as between loyal and occasional readers). Heterogeneity makes repeated exposures concentrate in the same people, so reach grows more slowly with the number of insertions. `calc_binomial()` and `calc_sainsbury()` assume homogeneous individuals; the Beta-Binomial family (`calc_beta_binomial()`, `calc_metheringham()`, `calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()`, `calc_mbd()`) lets the probability vary between people.
- *Vehicles*: homogeneous (every vehicle has the plan's average audience) or heterogeneous (each has its own audience). `calc_binomial()` and `calc_metheringham()` treat vehicles as homogeneous; `calc_sainsbury()` and the multivariate models take the audience of each vehicle as an input. Whether the distribution that a multivariate model returns reproduces those audiences and the plan's mean number of contacts exactly depends on the model: `calc_csd()` and `calc_msad()` conform their joint tables to the marginal distributions that they combine and therefore conserve the mean, whereas `calc_canex()`, `calc_cbd()` and `calc_mbd()` can alter both (by the truncation of negative cells, by the adjustment of the zero cell or by the bounded conditional allocation of MBD), in the last case even when no probability is negative. Read their `diagnostics` before interpreting the average frequency.
- *Duplication between vehicles*: random (people choose vehicles independently, so the audience shared by two vehicles is the product of their audiences) or observed (taken from data). `calc_binomial()` and `calc_sainsbury()` assume random duplication; the other models with several vehicles use observed duplication. The reference studies of the multivariate models mostly use observed duplications at or above the random one (Hong, 1998, p. 238); a lower duplication can be valid data, and its effect on each model is described in the section "Domain of validity" of `calc_canex()`.
- *Accumulation in the same vehicle*: random (a repeat insertion in a vehicle is as independent as an insertion in another vehicle) or observed through the reach after two insertions (`R2`), which captures that those who saw the first insertion tend to see the next. `calc_binomial()` and `calc_sainsbury()` assume random accumulation; the models for several insertions in a vehicle use the observed accumulation: `R2` or, in `calc_metheringham()`, the audience duplicated between two insertions of the same vehicle. The Beta-Binomial family requires `R2 <= 2 * R1 - R1^2`, that is, a duplication between two insertions of the vehicle not below the random one (it excludes negative dependence between insertions). That bound belongs to the family and is not the logical limit of two insertions, which is `R2 <= min(1, 2 * R1)` (see `calc_beta_binomial()`).

All the models also assume that exposure probabilities are stationary (they do not change over the period of the plan). Each help page states the assumptions of its model.

Computational limits

The schedule-level models build tables whose size grows with the number of insertions, and their cost grows exponentially with the number of vehicles. Before allocating anything, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()` estimate the size or the work that a schedule needs from its insertions, and stop with an explanatory error when it exceeds a budget. `calc_canex()` has its own limit of two million combinations of exposure levels, and `calc_cbd()` and `calc_mbd()` accept at most 12 vehicles. The two budgets can be raised, if the memory or the computing time allow it, with options:

- `mediaPlanR.max_table_cells` (default `1e7`): cells of the largest joint table of `calc_csd()` and `calc_msad()`, which combine two distributions at each step; it is the number of exposure levels of the vehicles already combined times that of the vehicle added, and each matrix of that size takes 8 bytes per cell (80 MB at the default).
- `mediaPlanR.max_work` (default `1e9`): estimated elementary operations of `calc_cbd()` and `calc_mbd()`, roughly one to two minutes of computation at the default. For `calc_cbd()` it is $2^m \sum_i (1 + n_1 + \cdots + n_{i-1})(n_i + 1)$, with m vehicles of n_i insertions; for `calc_mbd()`, the work of peeling the vehicles one at a time, which does not include the inclusion-exclusion grid that depends only on the number of vehicles.

For example, `options(mediaPlanR.max_table_cells = 5e7)`. An option that is not a single positive number is ignored.

Suggested study path

The help pages of a reference manual are in alphabetical order, which assumes prior knowledge. For sequential study, read in this order: (1) the sections "Notation and units" and "Assumptions of the models" above, to fix what is counted (people, contacts, pairs) and what each model assumes; (2) the models of independence, `calc_sainsbury()` and `calc_binomial()`; (3) the ad hoc duplication formulas, `calc_agostini_duplication()` and `calc_hofmans_duplication()`; (4) the accumulation of one vehicle, `calc_beta_binomial()` and `calc_hofmans_accumulation()`, with `fit_bbd_to_reach()` and `calibrate_bbd()`; (5) the multivariate models, `calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`, and the section "Domain of validity" of `calc_canex()`; (6) the count approximation `nbd_exposure_distribution()`; and (7) evaluation, `evaluate_exposure_model()`.

The vignette `vignette("mediaPlanR-intro")` gives worked examples of the models; it is the place for a longer teaching development, which the reference pages do not attempt.

Reach and exposure-distribution models

Plans with one or more insertions in each of several vehicles, with random duplication and accumulation: `calc_sainsbury()`, `calc_binomial()`. One vehicle with several insertions (accumulation models): `calc_beta_binomial()`, `calc_hofmans_accumulation()`. Several vehicles with one insertion each: `calc_agostini_duplication()`, `calc_hofmans_duplication()`. Several vehicles with several insertions: `calc_metheringham()`, `calc_canex()`, `calc_csd()`, `calc_msad()`, `calc_cbd()` and `calc_mbd()`. Fitting and validation: `fit_bbd_to_reach()`, `calibrate_bbd()`, `fit_nbd_exposure()`, `nbd_exposure_distribution()` and `evaluate_exposure_model()`.

Author(s)

Manuel J. Sánchez-Franco <majesus@us.es>

References

References for individual models are listed in their help pages. Development version and issue tracker: https://github.com/majesus/mediaPlanR.

See Also

Useful links:

- https://github.com/majesus/mediaPlanR
- Report bugs at https://github.com/majesus/mediaPlanR/issues

---

`metheringham_example` *Illustrative inputs for the Metheringham model*

Description

List of arguments ready for `do.call()` with `calc_metheringham()`: three vehicles with four, three and five insertions. Original illustrative data, not derived from any published source.

Usage

```
metheringham_example
```

Format

A list with the components:

audiences Numeric vector with the audience of each vehicle, in people per insertion.

insertions Numeric vector with the number of insertions in each vehicle.

duplication_matrix Symmetric matrix, in people: the people exposed to both one insertion in vehicle `i` and one in vehicle `j` (off-diagonal), and the people exposed to two insertions of the same vehicle (diagonal, as defined in `calc_metheringham()`).

population Population size, in people.

Examples

```
data(metheringham_example)
do.call(calc_metheringham, metheringham_example)
```

msad_example *Illustrative inputs for the MSAD model*

Description

List of arguments ready for `do.call(calc_msad, msad_example)`: a small two-vehicle scenario. Original illustrative data, not derived from any published source.

Usage

```
msad_example
```

Format

A list with the components:

vehicles_data Data frame with columns `insertions`, `R1` and `R2`.

duplications Symmetric matrix of proportions of the population exposed to both vehicles of each pair (one insertion in each).

aggregation_order The order of aggregation, `1:2`.

See Also

`calc_msad()`

Examples

```
data(msad_example)
do.call(calc_msad, msad_example)
```

nbd_exposure_distribution *Negative-Binomial exposure-count approximation*

Description

Builds a Poisson-Gamma (Negative-Binomial) distribution for exposure counts. This is an unbounded count-process approximation intended for page views, ad-server exposures, or other continuous opportunity processes. It is not a finite-insertion reach model and does not represent cross-vehicle dependence.

Usage

```
nbd_exposure_distribution(
  mean_contacts,
  size,
  report_max = NULL,
  opportunities = NULL,
  tail_tolerance = 1e-06
)
```

Arguments

`mean_contacts`	Positive expected exposures per person.
`size`	Positive Gamma heterogeneity parameter. Smaller values imply stronger heterogeneity. `Inf` gives the Poisson limit.
`report_max`	Positive integer at which the reported table becomes an open tail (`report_max` or more), not an exact final count.
`opportunities`	Optional finite number of exposure opportunities. It is used only to diagnose the probability that the unbounded model assigns to impossible counts above that number.
`tail_tolerance`	Non-negative diagnostic tolerance.

Details

The model assumes an individual Poisson exposure process with a Gamma- distributed rate across people. Its variance is $\mu + \mu^2/size$. Because its support is unbounded, the final table row is explicitly labeled as an open tail. No tail mass is reassigned to an impossible exact exposure count.

Danaher (2007) provides direct media-planning support for Negative-Binomial page-view models and develops the multivariate extension needed to represent dependence across websites. This function is deliberately univariate and therefore must not be interpreted as Danaher's multivariate model.

Value

An `nbd_exposure` object, a list with:

- `reach`: `probability` and `percent` of people with at least one exposure.
- `mean_contacts`: the mean number of exposures per person of the whole process, as supplied. It is not censored: the mean of the reported table is lower, because its last row stands for `report_max` or more exposures but is counted as exactly `report_max`.
- `average_frequency`: conditional mean among the people reached, `mean_contacts / reach` (exposures per person reached).
- `size` and `variance`: the heterogeneity parameter and the variance of the number of exposures per person, `mean_contacts + mean_contacts^2 / size`.
- `distribution`: data frame with `contacts` (an integer code; the last row carries `report_max`), `label` (`"0"`, `"1"`, ..., and `"report_max+"` for the last row), `open_tail` (`TRUE` only in that last row, whose probability is that of `report_max` or more exposures), `probability` and `cumulative_probability`.

- diagnostics: process and scope (text), report_max, opportunities (as supplied or NULL), probability_above_opportunities (the probability that the unbounded model assigns to more than opportunities exposures; NA if opportunities is NULL), finite_opportunity_compatible and tail_tolerance.

finite_opportunity_compatible is TRUE when the mass above opportunities is within tail_tolerance, or when opportunities was not supplied (then compatibility with a finite plan has not been checked). It is an approximate diagnostic only: it does not turn this unbounded model into a finite one.

References

Danaher, P. J. (2007). Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction. Marketing Science, 26(3), 422-437. doi:10.1287/mksc.1060.0226

Ehrenberg, A. S. C. (1959). The Pattern of Consumer Purchases. Applied Statistics, 8(1), 26-41. doi:10.2307/2985810

Examples

```
model <- nbd_exposure_distribution(
  mean_contacts = 2.5, size = 1.7, report_max = 10
)
model$reach
```

---

print.reach_cbd *Print the result of CSD, MSAD, CBD or MBD*

Description

By default, the full report used by the classical models: headline metrics, model parameters, the vehicles and aggregation steps, the exposure distribution, the cumulative distribution (the share of the population exposed N times or more, for each N), and numerical diagnostics. With full = FALSE, a compact summary: reach, average frequency, aggregation order and numerical diagnostics.

Usage

```
## S3 method for class 'reach_cbd'
print(x, full = TRUE, max_rows = 30L, ...)

## S3 method for class 'reach_csd'
print(x, full = TRUE, max_rows = 30L, ...)

## S3 method for class 'reach_mbd'
print(x, full = TRUE, max_rows = 30L, ...)

## S3 method for class 'reach_msad'
print(x, full = TRUE, max_rows = 30L, ...)
```

Arguments

x	A result of calc_csd(), calc_msad(), calc_cbd() or calc_mbd().
full	Logical. TRUE (default) prints the full report; FALSE prints the compact summary.
max_rows	Maximum number of exposure levels listed in the full report; longer distributions are cut with a note. Use Inf to list all of them.
...	Unused.

Details

Every result class has its own print method, because the objects are long lists and each model reports different intermediate results: CSD and MSAD list the vehicles and the aggregation steps (CSD also the canonical correlations between vehicles), CBD lists how each vehicle is modeled and the canonical zero probability, and MBD lists the peeling steps. The four methods are documented together on this page.

Value

x, invisibly.

Examples

```
data(csd_example)
csd <- do.call(calc_csd, csd_example)
csd
print(csd, full = FALSE)
```

---

ratings_example *Illustrative inputs for the Sainsbury and Binomial models*

Description

List of arguments ready for do.call() with calc_sainsbury() and calc_binomial(), the two models that assume random duplication and random accumulation. Original illustrative data, not derived from any published source.

Usage

```
ratings_example
```

Format

A list with the components:

audiences Numeric vector with the audience of each of three vehicles, in people per insertion.

population Population size, in people.

Examples

```
data(ratings_example)
do.call(calc_sainsbury, ratings_example)
do.call(calc_binomial, ratings_example)
```

Index

∗ **datasets**
bbd_reach_example, 3
beta_binomial_example, 3
canex_example, 44
csd_example, 45
duplication_example, 45
hofmans_accumulation_example, 54
mbd_example, 55
metheringham_example, 60
msad_example, 61
ratings_example, 64

bbd_reach_example, 3
beta_binomial_example, 3

calc_agostini_duplication, 4
calc_agostini_duplication(), *13*, *18*, *23*, *26*, *28*, *33*, *39*, *45, 46*, *50*, *52*, *57*, *59*
calc_beta_binomial, 6
calc_beta_binomial(), *3*, *9, 10*, *14, 15*, *20*, *25*, *28*, *35–37*, *41*, *43, 44*, *52*, *56–59*
calc_binomial, 8
calc_binomial(), *5*, *8*, *36*, *41*, *52*, *58, 59*, *64*
calc_canex, 10
calc_canex(), *13–23*, *31–33*, *38, 39*, *44*, *47*, *56–59*
calc_cbd, 14
calc_cbd(), *10*, *12–15*, *18, 19*, *22, 23*, *30–33*, *38–40*, *47*, *56*, *58, 59*, *64*
calc_csd, 20
calc_csd(), *10*, *12–15*, *17–19*, *22, 23*, *28*, *32, 33*, *38–40*, *45*, *47*, *56*, *58, 59*, *64*
calc_hofmans_accumulation, 24
calc_hofmans_accumulation(), *6*, *8*, *26, 27*, *54–56*, *59*
calc_hofmans_duplication, 26
calc_hofmans_duplication(), *4, 5*, *13*, *18*, *23–25*, *33*, *39*, *45, 46*, *50*, *52*, *57*, *59*
calc_mbd, 28
calc_mbd(), *10*, *13–19*, *22, 23*, *31–33*, *38–40*, *47*, *55*, *57–59*, *64*
calc_metheringham, 34
calc_metheringham(), *8, 9*, *41*, *52*, *57–60*
calc_msad, 36
calc_msad(), *10*, *13–15*, *18, 19*, *22, 23*, *32, 33*, *38, 39*, *47*, *50*, *56*, *58, 59*, *61*, *64*
calc_sainsbury, 40
calc_sainsbury(), *5*, *8, 9*, *36*, *52*, *58, 59*, *64*
calibrate_bbd, 42
calibrate_bbd(), *57*, *59*
canex_example, 44
csd_example, *15, 16*, 45

duplication_example, 45, *54, 55*

evaluate_exposure_model, 46
evaluate_exposure_model(), *56, 57*, *59*

fit_bbd_to_reach, 49
fit_bbd_to_reach(), *3*, *29*, *44*, *59*
fit_nbd_exposure, 53
fit_nbd_exposure(), *59*

hofmans_accumulation_example, *46*, 54

mbd_example, 55
mediaPlanR, *7*, *17*, *21*, *32*, *38*, 56
mediaPlanR-package *(*mediaPlanR*)*, 56
metheringham_example, 60
msad_example, *45*, 61

nbd_exposure_distribution, 61
nbd_exposure_distribution(), *8*, *53, 54*, *59*

print.reach_cbd, 63
print.reach_csd *(*print.reach_cbd*)*, 63
print.reach_mbd *(*print.reach_cbd*)*, 63
print.reach_msad *(*print.reach_cbd*)*, 63
print_sequential *(*print.reach_cbd*)*, 63

ratings_example, 64

stats::optimize(), *43*
stats::uniroot(), *51*