# Package 'EDE'

November 18, 2026

**Title** Extinction Date Estimation from Sighting Records

**Version** 0.4.0

**Maintainer** Rodrigo Fonseca Villa <rodrigo03.villa@gmail.com>

**Description** Infers the extinction of a species from a time-ordered record of sightings, collecting methods from the sighting-record and stratigraphic-range literature under one interface. Includes endpoint estimators (Robson & Whitlock, 1964; Strauss & Sadler, 1989; Roberts & Solow, 2003), frequentist persistence tests (Solow, 1993a, 1993b; Burgman, Grimson & Ferson, 1995; Solow & Roberts, 2003; Solow, 2005; McInerny, Roberts, Davy & Cribb, 2006; Jarić & Ebenhard, 2010), Bayesian models (Rout, Salomon & McCarthy, 2009; Caley & Barry, 2014; Alroy, 2014), methods for sightings of uncertain reliability (Thompson et al., 2013; Jarić & Roberts, 2014; Lee et al., 2014; Brook, Buettel & Jarić, 2019; Kodikara et al., 2020, 2021), models combining records and surveys (Thompson et al., 2017, 2019), and confidence intervals on the end-points of fossil ranges (Marshall, 1994, 1997; Weiss & Marshall, 1999), a test of a common end-point of several taxa (Solow, 1996), and an estimator of the extinction time from radiometrically dated remains that allows for dating error (Solow, 1997; Solow, Roberts & Robbirt, 2006; Bradshaw et al., 2012).
Also provides tools from extreme value theory, applied to the latest sightings as upper order statistics: end-point estimators (Hall & Wang, 1999; Dekkers, Einmahl & de Haan, 1989; Fraga Alves & Neves, 2014; Smith, 1985), tests of whether a finite end-point is supported (Neves, Picek & Fraga Alves, 2006; Hasofer & Wang, 1992), and an automatic choice of the number of recent sightings. Where the source paper gives a worked example the method is tested against it. All methods return a common result class; persistence tests are inverted to give one-sided upper confidence bounds.

**Language** en-US

**License** GPL (>= 3)

**URL** https://github.com/rodrigosqrt3/EDE

**BugReports** https://github.com/rodrigosqrt3/EDE/issues

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 4.1.0)

**Imports** graphics, stats, utils

**Suggests** testthat (>= 3.0.0), knitr, rmarkdown

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**Config/roxygen2/version** 8.1.1

**NeedsCompilation** no

**Author** Rodrigo Fonseca Villa [aut, cre] (ORCID:
<https://orcid.org/0009-0005-2938-2270>)

**Repository** CRAN

**Date/Publication** 2026-11-18 10:00:39 UTC

# Contents

alaska_megafauna . . . 3
alroy2014 . . . 4
as.data.frame.ede_estimate . . . 6
bradshaw2012 . . . 7
brook2019 . . . 9
burgman1995 . . . 11
caley_barry2014 . . . 13
dekkers1989 . . . 15
dodo . . . 17
ede_compare . . . 18
ede_endpoint_test . . . 19
ede_estimate . . . 21
ede_evi . . . 22
ede_field_test . . . 23
ede_select_k . . . 24
extinct_birds . . . 26
ferret . . . 27
fraga_alves_neves2014 . . . 27
hall_wang1999 . . . 29
jaric2010 . . . 31
jaric_roberts2014 . . . 32
kodikara2020 . . . 33
kodikara2021 . . . 35
lee2014 . . . 37
marshall1994 . . . 39
marshall1994_trend . . . 41
marshall1997 . . . 42
mccarthy1998 . . . 43

mcinerny2006 . . . 45
monk_seal . . . 46
nz_shells . . . 46
ole . . . 47
orchid . . . 48
plot.ede_compare . . . 48
plot.ede_estimate . . . 49
plot.sighting_data . . . 50
robson1964 . . . 50
rout2009 . . . 51
rout2009_absences . . . 52
seymour_ammonites . . . 53
sighting_data . . . 54
smith1985 . . . 54
solow1993 . . . 56
solow1993b . . . 57
solow1993_bayes . . . 58
solow1996 . . . 59
solow2005 . . . 60
solow2006 . . . 61
solow2006_check . . . 63
solow2006_region . . . 64
solow_roberts2003 . . . 66
strauss1989 . . . 67
strauss1989_bayes . . . 67
strauss1989_curve . . . 69
thompson2013 . . . 69
thompson2017 . . . 71
thompson2019 . . . 73
vogel2009 . . . 75
weiss_marshall1999 . . . 77
weiss_marshall1999_prior . . . 78

**Index** **80**

---

`alaska_megafauna` *Radiocarbon ages of Alaskan mammoths and horses*

---

## Description

The 25 most recent radiocarbon ages of remains of woolly mammoths (*Mammuthus primigenius*) and caballoid horses (*Equus ferus*) in Alaska, with the standard deviations of their dating errors, as listed in Table 1 of Solow, Roberts & Robbirt (2006), who took them from Guthrie (2003, 2004). The authors omit the oldest horse age (17.65 ka BP) from their analysis. Used as the example of `solow2006()` and `solow2006_region()`.

## Usage

```
alaska_megafauna
```

## Format

A data frame with 50 rows: `taxon` (`"mammoth"` or `"horse"`), `age_ka_bp` (radiocarbon age in thousand years before present) and `sd_ka` (standard deviation of the dating error, in thousand years).

## Source

Solow, A. R., Roberts, D. L., & Robbirt, K. M. (2006). On the Pleistocene extinctions of Alaskan mammoths and horses. Proceedings of the National Academy of Sciences, 103(19), 7351-7353, Table 1.

## Examples

```
head(alaska_megafauna)
```

---

# alroy2014 *Alroy (2014) "creeping-shadow-of-a-doubt" posterior probability of extinction*

---

## Description

Posterior probability that a species is extinct, interval by interval after its last sighting, by the iterative Bayesian method of Alroy (2014). The record is a series of time units (years, say) in each of which the species was either sighted (a presence) or not (an absence). The prior probability of going extinct in one unit, `E`, comes from the length of the observed range, by assuming that extinction is an exponential decay process with a 50% chance that the species has gone extinct by the end of the range; the sampling probability `ps` is the frequency of sightings within the range. Bayes' theorem then gives the posterior `e[1]` for the first unit after the last sighting, which is carried to the next unit as the prior, where extinction might instead have happened: `e[t] = (e[t-1] + (1 - e[t-1]) E) / ((e[t-1] + (1 - e[t-1]) E) + (1 - e[t-1]) (1 - E) (1 - ps))` (equations 3 and 4 of the paper), starting from `e[0] = 0` at the last sighting. The values are Alroy's "shadow values".

## Usage

```
alroy2014(
  sd,
  test_year,
  alpha = 0.05,
  interval = c("one-sided", "two-sided"),
  base = NULL,
  adjust_range = TRUE,
  truncated = FALSE,
  run_based = FALSE,
  data_out = FALSE
)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object whose times are whole numbers (time units); a unit is a presence if its count is positive. |
| test_year | Latest time unit for which a shadow value is computed, a whole number later than the last sighting. |
| alpha | Significance level, in (0, 1), for the credible limits. |
| interval | "one-sided" for an upper credible limit, or "two-sided" for an equal-tailed credible interval. |
| base | First time unit at which extinction was plausible; defaults to the first sighting and cannot be later than it. |
| adjust_range | Correct the range for sighting failures (r (np + 1) / (np - 1)). |
| truncated | If TRUE, the first sighting is not the first appearance of the species, which changes the sampling probability (see above). |
| run_based | Use the sampling probabilities conditioned on the preceding unit. |
| data_out | If TRUE, return the shadow values as a data frame instead of the estimate. |

## Details

The range r is the number of time units from base (by default the first sighting) to the last sighting, both included, and the sighting probability uses the np units with sightings and the na units without them between the first and the last sighting, ps = (np - 2) / (np - 2 + na) (the first and last presences are required to exist). With adjust_range = TRUE the range is replaced by r (np + 1) / (np - 1), the correction of Strauss & Sadler (1989) used in the simulations of the paper. The paper's remarks on how to choose the range apply: the species may be older than the record (truncated = TRUE uses ps = (np - 1) / (np - 1 + na)), and for a recent extinction the range should start at the first date at which extinction was plausible rather than at the first sighting, which base allows; results sensitive to that choice should be treated with caution. With run_based = TRUE the sighting probability is conditioned on the state of the preceding unit, ps given a presence for the first unit and given an absence afterwards (the "run-based" version of the paper, which it recommends when the sample is large enough).

Following the paper, the best estimate of the extinction time is the first unit at which the shadow value reaches 0.5, and credible limits are the first units at which it exceeds 1 - alpha (one-sided) or alpha / 2 and 1 - alpha / 2 (two-sided). A species with fewer than three units with sightings cannot be analysed. The versions of the paper that use sampling controls (several sampling events per unit) and the maximum-likelihood alternative of its Alternative Method section are not implemented.

## Value

An ede_estimate object with estimate the first time unit at which the shadow value is at least 0.5 and the credible limits in lower and upper (NA, with a warning, when not reached by test_year), and the components shadow (a data frame with time and e), E, ps, r, n_presence and n_absence. With data_out = TRUE the data frame of shadow values.

## References

Alroy, J. (2014). A simple Bayesian method of inferring extinction. Paleobiology, 40(4), 584-607. doi:10.1666/13074

Strauss, D., & Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology, 21(4), 411-427.

## See Also

`weiss_marshall1999()` and `strauss1989_bayes()` for other Bayesian estimates of the end of a range.

## Examples

```
# a species seen in 14 of 50 units (about once every four), then not seen
times <- c(0, 4, 7, 11, 15, 18, 22, 26, 29, 33, 37, 41, 45, 49)
sd <- sighting_data(data.frame(time = times, count = 1))
alroy2014(sd, test_year = 80)
head(alroy2014(sd, test_year = 80, data_out = TRUE))
alroy2014(sd, test_year = 80, run_based = TRUE, interval = "two-sided")
```

---

# as.data.frame.ede_estimate

*Coerce an extinction date estimate to a data frame*

---

## Description

Coerce an extinction date estimate to a data frame

## Usage

```
## S3 method for class 'ede_estimate'
as.data.frame(x, row.names = NULL, optional = FALSE, ...)
```

## Arguments

| | |
|---|---|
| `x` | An ede_estimate object. |
| `row.names, optional` | Not used; present for consistency with `base::as.data.frame()`. |
| `...` | Not used. |

## Value

A one-row data frame with columns `method`, `estimate`, `lower`, `upper`, `interval_type` and `alpha`.

## Examples

```
as.data.frame(ole(sighting_data(dodo)))
```

`bradshaw2012` *Gaussian-resampled, inverse-weighted McInerny estimate (GRIWM) (Bradshaw et al. 2012)*

## Description

Estimate of the extinction time from a series of dated records that allows for dating error, by the GRIWM method of Bradshaw et al. (2012). It extends the sighting-rate model of McInerny et al. (2006), see `mcinerny2006()`, in two ways. First, the estimate is computed for every number `k` of the most recent records, from 2 to `n`, and the results are averaged with weights that decrease with the age of the `k`-th record, because the most recent records are taken to be more informative about the sighting rate as extinction is approached. Second, each date is resampled from a normal distribution with its dating error and the weighted estimate is recomputed for each draw.

## Usage

```
bradshaw2012(
  sd,
  error = NULL,
  alpha = 0.05,
  level = 0.95,
  nsim = 10000,
  k = NULL
)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object whose positive-count rows are the dated records (a count above one repeats the date), or, with `error`, a data frame with columns `time` and `error`, one row per record. |
| `error` | Dating standard deviation of each record: one value, or one per row of `sd` in the order of its rows; or `NULL` for no dating error. For a data frame with an `error` column it is taken from the data frame. |
| `alpha` | Threshold probability of the sighting-rate model, in (0, 1); the paper uses 0.05. |
| `level` | Coverage of the interval from the resampled estimates, in (0, 1); the paper uses 0.95. |
| `nsim` | Number of resampled series; the paper uses 10,000. |
| `k` | Number of most recent records to use (at least 2); by default all. The paper uses the 25 most recent. |

## Details

For the `k` most recent records, with `x[1]` the most recent and `s[k] = x[1] - x[k]` the span of the series, the probability of no further record up to a time `x[1] + d` after the last one is `(1 - k / s[k])^d` (equation 3 of the paper, with `k` records over a period `s[k]`). The estimate `theta[k]` for level `alpha` is the first time at which this probability is `alpha`, `x[1] + log(alpha) / log(1 - k / s[k])`. The

weighted estimate is `sum(w[k] theta[k]) / sum(w[k])` over `k = 2, ..., n`, with weights `w[k] = 1 / s[k]` (equation 4; the paper expresses them as a ratio to a common constant, which cancels). The time unit is the one in which the sighting probability per unit is meant, as in `mcinerny2006()`: for radiocarbon ages, years, not thousands of years.

With dating errors, the estimate is the median of `nsim` draws of the weighted estimate and the interval is formed by the quantiles of the draws, as in the paper. Without errors, the result is the inverse-weighted McInerny estimate of the paper (its Figure 2) and no interval is given. For a draw, `theta[k]` is undefined if the series has `k` or more records in a span of at most `k` time units (the probability above is then not defined); such values are left out of the weighted mean, and a draw with none left is dropped and counted in `failed`.

The paper compares the method with others on simulated series and reports that it has the lowest average bias, with wider intervals, when the sighting rate declines towards extinction; the weights are an heuristic of the authors, not a derived estimator.

## Value

An ede_estimate object with the median weighted estimate in `estimate` and, if there are dating errors, the interval in `lower` and `upper` (`interval_type "two-sided"`), and the components `theta_k` (a data frame with `k`, `span` and `theta` for the dates as recorded), `theta_w` (the weighted estimate for the dates as recorded), `draws` and `failed`.

## References

Bradshaw, C. J. A., Cooper, A., Turney, C. S. M., & Brook, B. W. (2012). Robust estimates of extinction time in the geological record. Quaternary Science Reviews, 33, 14-19. doi:10.1016/j.quascirev.2011.11.021

McInerny, G. J., Roberts, D. L., Davy, A. J., & Cribb, P. J. (2006). Significance of sighting rate in inferring extinction and threat. Conservation Biology, 20(2), 562-567.

## See Also

`mcinerny2006()`, `solow2006()` for the likelihood approach to dated records, alaska_megafauna.

## Examples

```
mam <- alaska_megafauna[alaska_megafauna$taxon == "mammoth", ]
# years, increasing towards the present
dates <- data.frame(time = -mam$age_ka_bp * 1000, error = mam$sd_ka * 1000)
set.seed(1)
bradshaw2012(dates, nsim = 1000)
# without dating error: the inverse-weighted McInerny estimate
bradshaw2012(sighting_data(data.frame(time = dates$time, count = 1)))
```

brook2019 *Reliability-weighted resampling of a sighting record (Brook et al., 2019)*

## Description

Applies any extinction-date estimator to many random sub-records drawn from a sighting record of mixed reliability, and summarises the resulting empirical distribution (the "BBJ" weighted-resampling approach of Brook, Buettel & Jaric, 2019).

## Usage

```
brook2019(
  sd,
  reliability,
  method = strauss1989,
  n_resamples = 1000,
  statistic = c("estimate", "upper"),
  alpha = 0.05,
  persist_year = NULL,
  min_sightings = 3,
  seed = NULL,
  ...
)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| reliability | Sighting reliabilities in [0, 1]. Either a numeric vector with one value per row of sd or per positive-count row (taken as the combined reliability of that time unit), or a list with one numeric vector per positive-count row giving the reliabilities of the individual sightings in that time unit (its length must equal the count); these are combined as in step 1. Values for zero-count rows are ignored. |
| method | An estimator function whose first argument is a sighting_data object, such as ole(), strauss1989(), robson1964() or solow1993(). It may return an ede_estimate object or a single number. The "S93" estimator of the authors' code, t[n] + (t[n] - t[1]) / (n - 1) with the first sighting as origin, is the point estimate of strauss1989() (the default). |
| n_resamples | Number of random sub-records (the paper suggests about 1,000 or more; the authors' code uses 10,000). |
| statistic | Component of the ede_estimate returned by method to collect: "estimate" (default) or "upper". Ignored if method returns a number. |
| alpha | Level for the upper quantile, in (0, 1). |
| persist_year | Optional time at which to report the proportion of resampled values later than it. |
| min_sightings | Minimum number of kept records for a sub-record to be analysed. |

| | |
|---|---|
| seed | Optional random seed. The global random number state is restored on exit. |
| ... | Further arguments passed to method (e.g. test_year). |

## Details

The algorithm follows the paper and the authors' R code (Data S1, Zenodo doi:10.5281/zenodo.3247189, bbj.2018()):

1. Sightings in the same time unit are combined into one record with reliability p[j] = 1 - prod(1 - p[ij]) (the unnumbered equation in the Methods of Brook et al., 2019).
2. Each record is kept independently with probability equal to its reliability (resampling "without replacement", using reliability as an inclusion probability). Records with reliability 1 are therefore always kept; there is no other rule that forces the most recent certain sighting to be retained.
3. method is applied to each sub-record and the chosen statistic is stored. Sub-records with fewer than min_sightings records (the authors' code requires at least 3), and sub-records for which method fails or returns NA, are discarded; their number and proportion are reported.
4. The kept values form the empirical distribution of the statistic.

Summaries: estimate is the median of the distribution, the point estimate named in the paper's Methods and in its Fig. 1 (the authors' code instead reports the mean, which is returned as mean). upper is the 1 - alpha quantile of the distribution, the "upper 95% confidence bound" of the paper (its text calls it the "2.5th percentile", apparently in error; the authors' code takes the last year at which the proportion of values not yet reached exceeds alpha, i.e. the 1 - alpha quantile). The extinction years in Table 1(b) of the paper are these upper bounds. If persist_year is given, persistence is the proportion of values later than persist_year, the "probability of persistence" in Table 1. Quantiles use type = 1 (inverse empirical distribution), so infinite values (e.g. a non-existent upper bound) are handled.

Kept records enter the sub-record with count 1. Zero-count rows of sd (for example a leading row that fixes the observation origin) are kept in every sub-record; otherwise the earliest kept sighting becomes the origin for estimators that use one, as in the authors' code.

Status of the method: Solow (2023) criticised the statistical properties of the extinction date and interval produced this way, noting in particular that the estimated extinction time exceeds the true one when reliabilities are high. Jaric et al. (2023) replied that the procedure is not itself an extinction-date estimator but a pre-processing step for the input data, whose result is passed to any estimator (McCrea et al., 2024). The summaries returned here are descriptive statistics of that resampling distribution, not confidence intervals with established coverage.

## Value

An ede_estimate object (interval_type = "one-sided") with estimate (median), upper (1 - alpha quantile) and additional components mean, values (the kept resampled values), n_resamples, n_failed, prop_failed, statistic, combined_reliability and, if persist_year is given, persistence and persist_year.

## References

Brook, B. W., Buettel, J. C., & Jaric, I. (2019). A fast re-sampling method for using reliability ratings of sightings with extinction-date estimators. Ecology, 100(9), e02787.

Jaric, I., Buettel, J. C., & Brook, B. W. (2023). A fast re-sampling method for using reliability ratings of sightings with extinction-date estimators: reply. Ecology, 104(9), e4124.

Solow, A. R. (2023). A fast re-sampling method for using reliability ratings of sightings with extinction-date estimators: comment. Ecology, 104(9), e4123.

McCrea, R. S., Cheale, T., Campillo-Funollet, E., & Roberts, D. L. (2024). Inferring species extinction from sighting data. Cambridge Prisms: Extinction, 2, e19.

---

`burgman1995` *Burgman, Grimson & Ferson (1995) tests for frequency sighting data*

---

## Description

Exact tests for sighting records divided into `CT` equal periods, where `N` is the total number of sightings (all records counted, not only occupied periods).

## Usage

```
burgman1995(sd, alpha = 0.05, test_year, data_out = FALSE, equation = 2)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). The returned upper bound is the earliest candidate time at which persistence is rejected at this level. |
| `test_year` | Latest time to test. Must be later than the last sighting. A bound beyond `test_year` is reported as `NA` with a warning. |
| `data_out` | If `TRUE`, return the p-value curve at integer steps after the last sighting instead of the upper bound. |
| `equation` | Which equation of the source paper to use: `2` for extinction (the default), `3` (binary runs test) or `4` (frequency runs test) for a change in conservation status. |

## Details

- `equation = 2` (the default) tests extinction. The p-value `(Ce / CT)^N`, where `Ce` is the period of the last sighting, is the probability that the species is not extinct; it is the empty-cells test restricted to the run of empty periods at the end of the record.
- `equation = 3` is the runs test of Bradley (1968) on the binary (occupied / empty) sequence of periods, so counts are ignored. With `n1` occupied periods out of `CT` and `r` the longest run of empty periods anywhere in the record, the probability of an empty run at least `r` long is `sum_{k=1}^{[n0/r]} (-1)^(k+1) C(n1+1, k) C(CT - r k, n1) / C(CT, n1)`, with `n0 = CT - n1` empty periods (inclusion-exclusion over the `n1 + 1` gaps that hold the empty periods).

The equation is illegible in the source, so this form was reconstructed from its definitions and checked against the paper's numbers: it gives the Fig. 1b value 0.02885 and the Table 2 values for the black-footed ferret (0.003) and both stick-nest rats (0.139, 0.350). It does not give the Table 2 value for the monk seal (0.244 here against 0.258 in the paper, which equals the paper's equation 4 value; see Details). Like equation 4, it addresses a change in status rather than extinction.

- `equation = 4` tests for a change in conservation status (Grimson et al., 1992). It computes, by inclusion-exclusion over the occupancy of periods, the probability of a run of empty periods at least as long as the longest observed run anywhere in the record. Burgman et al. (1995, p. 927) note that it addresses a different question from extinction, and that a significant result may be discounted if the longest run is not near the end of the record.

As in the Caribbean monk seal example of the source paper (Fig. 1a and Tables 1-2), the earliest supplied time defines the observation origin and is not one of the `CT` periods; sightings recorded in that period are not counted. To include an earlier start date, prepend a zero-count row. The paper's black-footed ferret values (observation from January 1972) correspond to a zero-count row for 1971, so that 1972 is a period. Burgman et al. (2000, Table 2.1) used the same convention for Western Australian *Acacia*: the year of the first collection is the first period and 1993 the last; equations 2 and 4 reproduce their published values.

For the monk seal, the origin convention above reproduces the paper's equation 2 (0.053) and equation 4 (0.258) values (`CT = 77`, `n1 = N = 4`, and `r` equal to 40), but equation 3 gives 0.244. No choice of `CT` and `r` near these values gives 0.258 exactly (`CT = 78` with `n1 = 4`, i.e. counting 1915 as an empty period, gives 0.2587). Since all seal counts are 1, the paper's equal entries for equations 3 and 4 suggest that the equation 4 value was reported for both.

The paper defines only the tests. Since the periods are discrete, the returned upper bound is the first period after the last sighting at which the null hypothesis is rejected. Under `equation = 3` or `4` the p-value curve need not be monotone, so this bound should be read with the caveat above.

## Value

An ede_estimate object with `estimate = NA` and the one-sided upper bound in `upper`, or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (plus `chance`, a deprecated alias of `p_value`).

## References

Burgman, M. A., Grimson, R. C., & Ferson, S. (1995). Inferring threat from scientific collections. Conservation Biology, 9(4), 923-928.

Burgman, M. A., Maslin, B. R., Andrewartha, D., Keatley, M. R., Boek, C., & McCarthy, M. (2000). Inferring threat from scientific collections: power tests and an application to Western Australian Acacia species. In S. Ferson & M. Burgman (Eds.), Quantitative Methods for Conservation Biology (pp. 7-26). Springer, New York.

Bradley, J. V. (1968). Distribution-free statistical tests. Prentice-Hall, Englewood Cliffs, New Jersey.

`caley_barry2014` *Caley & Barry (2014) Bayesian extinction probability from a detection record*

## Description

Posterior probability that a population is extant at the end of `test_year`, from a record of detections (`y[t] = 1`) and non-detections (`y[t] = 0`) in each period `t = 1, ..., T` (Caley & Barry, 2014). The population is extant at the start of period 1. The latent variable `z` is the period in which it goes extinct (`z > T` if it persists beyond the record). If it goes extinct in period `z`, it cannot be detected from period `z` on.

## Usage

```
caley_barry2014(
  sd,
  test_year,
  model = c("constant", "nonconstant"),
  formulation = c("published", "equations"),
  prior_phi = c(1.1, 1.1),
  prior_lambda = c(1.1, 1.1),
  prior_growth = c(-2.3, 0.69),
  prior_delta = c(0.01, 4.6),
  prior_eps0 = c(-20, 20),
  prior_eps1 = c(0, 20),
  n_iter = 1e+05,
  burn_in = n_iter/10,
  data_out = FALSE
)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object with whole-number times. |
| `test_year` | Last period of observation. Must be later than the last sighting. The posterior refers to persistence beyond the end of this period, given no detections after the last sighting. |
| `model` | `"constant"` (exact computation) or `"nonconstant"` (MCMC). |
| `formulation` | For the constant model: `"published"` (Code S1, which reproduces the paper) or `"equations"` (likelihood as printed). |
| `prior_phi`, `prior_lambda` | Beta prior parameters `c(a, b)` for the constant model. |
| `prior_growth`, `prior_delta`, `prior_eps0`, `prior_eps1` | Uniform prior bounds for the non-constant model. |
| `n_iter`, `burn_in` | Total MCMC iterations and initial iterations discarded (non-constant model). Use `set.seed()` for reproducibility. |

data_out — If TRUE (constant model only), return the posterior probability of persistence at the end of each period from the one after the last sighting to test_year, assuming no further detections.

## Details

**Constant model** (model = "constant", section "Model specification - constant population"). Each period, the population goes extinct with probability phi and, if extant, is detected with probability lambda. The priors are phi ~ Beta(a_phi, b_phi) and lambda ~ Beta(a_lambda, b_lambda). The paper's likelihood is [y | z] = prod_{t < z} lambda^y[t] (1 - lambda)^(1 - y[t]), and [z] = (1 - phi)^(z - 1) phi, or (1 - phi)^T when z > T. The paper samples the posterior with a Gibbs sampler (Code S1). Here it is computed exactly instead: given z, the integrals over phi and lambda are Beta functions, so the posterior of z is a finite set of weights, and the posteriors of phi and lambda are mixtures of Beta distributions.

The published numbers come from Code S1. Its lambda update, Beta(a_lambda + S, b_lambda + z - S), counts the extinction period itself as a non-detection, one period more than the likelihood above. The z update has no such term. The sampler is still a valid Gibbs sampler: the extra factor (1 - lambda) is the same for every z, so it targets the model above with the prior lambda ~ Beta(a_lambda, b_lambda + 1). formulation = "published" (the default) computes that posterior and reproduces the paper's results. formulation = "equations" uses the likelihood as printed in the paper.

**Non-constant model** (model = "nonconstant", Code S2). The population size is N[t] = exp(r (t - 1)), with N[1] = 1 fixed because N[1] and delta are not separately identifiable. Detection is lambda[t] = 1 - exp(-delta N[t]) (equation 3) and extinction is logit(phi[t]) = eps0 - eps1 log(N[t]) (equation 4). The priors are uniform on prior_growth, prior_delta, prior_eps0 and prior_eps1. Each iteration of the Metropolis-within-Gibbs sampler of Code S2 is ported as published:

1. impute z from its conditional distribution;
2. take a log-normal random-walk step for delta (sd 0.5);
3. take a joint step for eps0 (normal, sd 1.5) and eps1 (log-normal, sd 0.15);
4. take a normal random-walk step for r (sd 0.05).

The chain starts from r = 0, delta = log(2), eps0 = 0 and eps1 = 1. As in Code S1, the lambda likelihood used in the delta and r updates includes the extinction period as a non-detection, while the z update does not. Because lambda[t] now varies with t, this factor differs between values of z. **The full conditionals of Code S2 are therefore not those of a single joint model**, and the sampler is reproduced only for comparability with the paper. Code S2 uses r ~ U(-2.3, 0.69), whereas the paper's text gives U(-1.6, 0.69); the code's value is the default here. Code S2 also does not report an implied prior probability of persistence that can be checked, so prior and bayes_factor are NA for this model.

**Input.** Periods are the integer time units from the earliest time in sd to test_year. A period is a detection if its count is positive; periods not listed in sd are non-detections. Unlike in solow1993(), the first period is an observation period even if it has a sighting. To start observation before the first sighting, add a zero-count row (e.g. the 1999 start of the paper's fox example).

## Value

A list of class ede_bayes with components:

- posterior: posterior probability that the population is extant at the end of test_year.
- prior: implied prior probability of that event, B(a_phi, b_phi + T) / B(a_phi, b_phi), for the constant model; NA for the non-constant model.
- bayes_factor: posterior odds over prior odds of persistence; NA for the non-constant model.
- test_year and method.
- extinction_time: a data frame with the posterior probability that extinction occurred in each period after the last sighting.
- Posterior medians and 95% equal-tailed intervals: phi and lambda (constant model), or growth, delta, eps0 and eps1 (non-constant model).
- For the non-constant model, also the chain after burn-in (samples) and the acceptance rates (acceptance).

If data_out = TRUE, a data frame with columns time, bayes_factor and posterior is returned instead.

## References

Caley, P., & Barry, S. C. (2014). Quantifying extinction probabilities from sighting records: inference and uncertainties. PLoS ONE, 9(4), e95857.

## Examples

```
# Tasmanian red fox carcasses (Caley & Barry, 2014, Table 1), 2001 start
fox <- sighting_data(data.frame(year = c(2001, 2003, 2005, 2006),
                                count = 1))
caley_barry2014(fox, test_year = 2012)
# 1999 start
fox99 <- sighting_data(data.frame(year = c(1999, 2001, 2003, 2005, 2006),
                                  count = c(0, 1, 1, 1, 1)))
caley_barry2014(fox99, test_year = 2012)$posterior
```

---

# dekkers1989 *Moment estimator of the extreme-value index and of the end-point*

---

## Description

Estimates the extreme-value index gamma of the sighting-time distribution, and from it the extinction time, with the moment estimator of Dekkers, Einmahl & de Haan (1989) applied to the k latest sightings. The index separates the three domains of attraction: gamma < 0 is the Weibull domain, a finite end-point, the setting of ole() and hall_wang1999() (the Weibull extremal exponent of those functions is -1 / gamma); gamma = 0 is the Gumbel domain, and gamma > 0 an unbounded upper tail. Unlike those estimators it does not assume the Weibull domain, so it also says whether the assumption is plausible.

## Usage

```
dekkers1989(sd, k = NULL, events = FALSE, origin = NULL, alpha = 0.05)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| k | Number of latest sightings, an integer of at least 3 with k + 1 at most the number of sighting times later than origin, or NULL for the 10 latest (fewer if the record is shorter). With the default origin, the earliest sighting cannot serve as threshold, so k is at most the number of sighting times minus two. |
| events | If FALSE (default), count each distinct sighting time once; if TRUE, count every sighting event. |
| origin | Origin of the time axis, earlier than the threshold sighting k + 1; NULL (default) uses the earliest time in sd. |
| alpha | Significance level, in (0, 1), of the normal interval for gamma. |

## Details

With x[i] the sighting times measured from origin, latest first, and the threshold x[k + 1], let M1 = mean(log(x[i] / x[k + 1])) and M2 = mean(log(x[i] / x[k + 1])^2) for i = 1, ..., k. The index estimate is gamma = M1 + 1 - 0.5 / (1 - M1^2 / M2). For gamma < 0 the end-point estimate is x[k + 1] - a / gamma, with scale a = x[k + 1] * M1 * (1 - gamma_minus) and gamma_minus = 1 - 0.5 / (1 - M1^2 / M2) (de Haan & Ferreira, 2006, Section 4.5). The end-point estimate is NA, with a warning, when the estimated index is not negative.

The estimator uses logarithms of the observations, so it requires the times to be positive and, unlike the other estimators in this package, it is not invariant to the origin of the time axis (de Haan & Ferreira, 2006, p. 102). The default origin is the earliest time in the record, the start of observation; the estimate depends on that choice, so check it with different origins. The estimator needs k + 1 sighting times and the threshold must be later than origin.

The asymptotic standard error of the index estimate is sqrt(v / k) with v = 1 + gamma^2 for gamma >= 0 and v = (1 - gamma)^2 (1 - 2 gamma) (1 - gamma + 6 gamma^2) / ((1 - 3 gamma) (1 - 4 gamma)) for gamma < 0. It is a large-sample result, and with the few sightings typical of extinction records the interval is only a guide. No interval is given for the end-point.

By default each distinct sighting time counts once, as in hall_wang1999(); set events = TRUE to count every sighting event.

## Value

An ede_estimate object with estimate (possibly NA, see Details), no interval, and components gamma, gamma_se, gamma_ci (a 1 - alpha normal interval), k and origin.

## References

Dekkers, A. L. M., Einmahl, J. H. J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. Annals of Statistics, 17(4), 1833-1855. doi:10.1214/aos/1176347397

de Haan, L., & Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer, Sections 3.5 and 4.5.

## See Also

ede_evi() for the estimates over a range of k; ede_endpoint_test() to test the Gumbel null; hall_wang1999() and ole() for the Weibull domain.

## Examples

```
sd <- sighting_data(data.frame(year = c(1900, 1904, 1909, 1911, 1913, 1918,
                                        1920, 1925, 1928, 1930, 1934),
                               sightings = 1))
dekkers1989(sd, k = 6)
```

---

# dodo *Sighting record of the dodo*

---

## Description

Reliable sightings of the dodo (*Raphus cucullatus*) in Mauritius, used by Roberts & Solow (2003) to estimate its extinction date with ole(). Two sightings were made in 1628. A later report from 1674 was excluded by the authors as unreliable.

## Usage

```
dodo
```

## Format

A data frame with 9 rows: year and count (number of sightings).

## Source

Roberts, D. L., & Solow, A. R. (2003). Flightless birds: When did the dodo become extinct? Nature, 426(6964), 245.

## Examples

```
ole(sighting_data(dodo))
```

`ede_compare` *Apply every applicable method to one sighting record*

## Description

Runs the estimators and persistence tests that need only a sighting record and returns their results side by side. Methods that need additional input (such as sighting reliabilities) are not included. A method that cannot be applied to the record, or that warns, is kept in the table with the message in `note`.

## Usage

```
ede_compare(sd, test_year, alpha = 0.05)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `test_year` | Latest time to test for the persistence tests. Must be later than the last sighting. |
| `alpha` | Significance level, in (0, 1). |

## Details

The methods rest on different models, so their results answer different questions and should not be averaged; see the vignette for guidance on choosing a method.

## Value

A data frame of class `ede_compare`, with one row per method and columns `method`, `function_name`, `estimate`, `lower`, `upper`, `interval_type` and `note`. Persistence tests report the one-sided upper bound in `upper`. Use `plot()` for a forest plot.

## Examples

```
ede_compare(sighting_data(monk_seal), test_year = 2050)
```

`ede_endpoint_test` *Test whether a sighting record is compatible with a finite end-point*

## Description

Extinction dates are the finite upper end-point of the sighting-time distribution, but a finite end-point is not guaranteed: if the sighting rate declines smoothly to zero (an exponential tail, the Gumbel domain of attraction) or decays like a power law (the Frechet domain), the end-point estimators for the Weibull domain, `ole()` and `hall_wang1999()`, rest on a model the data do not support. This function tests, from the `k` latest sightings, the null hypothesis that the sighting times lie in the Gumbel domain of attraction against the alternative that they lie in the Weibull domain (a finite end-point, the setting of those estimators) or in the Frechet domain (an unbounded upper tail).

## Usage

```
ede_endpoint_test(
  sd,
  k = NULL,
  test = c("neves", "hasofer_wang"),
  alternative = c("finite", "infinite", "two.sided"),
  events = FALSE,
  nsim = 1e+05
)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `k` | Number of latest sightings, an integer of at least 4, or `NULL` for the 10 latest (fewer if the record is shorter; the Neves test needs one more sighting than `k`, and the Rosario test needs `2k + 1` sightings). `ede_select_k()` chooses `k` for the estimators of the Weibull domain. |
| `test` | Which tests to run: any of `"neves"`, `"hasofer_wang"` and `"rosario"`; the first two by default. |
| `alternative` | `"finite"` (default), `"infinite"` or `"two.sided"`; see Details. |
| `events` | If `FALSE` (default), count each distinct sighting time once; if `TRUE`, count every sighting event. |
| `nsim` | Number of draws used for the null of Greenwood's statistic in the Hasofer & Wang test. |

## Details

Three tests are available. All are invariant to the origin and the unit of time, and all depend only on the latest sightings (the `k` latest, or the `2k + 1` latest for `"rosario"`), taken in the same way as for `hall_wang1999()`.

- `"neves"`: the statistic of Neves, Picek & Fraga Alves (2006), the ratio of the maximum to the mean of the excesses over the `(k + 1)`-th latest sighting, minus `log(k)`. With `t[1] > t[2] > ...` the sighting times, latest first, it is `T* = (t[1] - t[k + 1]) / mean(t[1:k] - t[k + 1]) - log(k)`. It needs `k + 1` sighting times. The paper rejects when `T*` falls outside the Gumbel quantiles. Here the p-value is exact for an exponential tail: `T* + log(k)` is then `k` times the maximum over the sum of `k` independent exponentials, whose distribution is Fisher's (1929) statistic for periodogram analysis. The column `p_asymptotic` gives the Gumbel approximation of the paper.
- `"hasofer_wang"`: the statistic of Hasofer & Wang (1992), `W = k (mean(x) - x[k])^2 / ((k - 1) sum((x - mean(x)))` with `x` the `k` latest sightings. Under the null the distances `t[i] - t[k]` of the other `k - 1` sightings are distributed as `k - 1` independent exponentials, and `W = 1 / ((k - 1) (k G - 1))` with `G` Greenwood's statistic of those distances. The p-value is therefore taken from the same simulated null as `hall_wang1999()`.
- `"rosario"`: the statistic of Fraga Alves, Neves & Rosario (2016), built on the end-point estimator of `fraga_alves_neves2014()`. With `xhat` that estimate for `k` and `t[1] > t[2] > ...` the sighting times, latest first, `G = (xhat - t[k + 1]) / (t[k + 1] - t[2 k + 1])` and the statistic is `G* = log(2) G - (log(k) + log(2) / 2)`, which has a Gumbel limit under the null (their Theorem 6) and goes to minus infinity for a finite end-point and plus infinity for an unbounded tail. It needs `2k + 1` sighting times. The paper rejects when `G*` is below the Gumbel quantile of level `alpha` (finite end-point); `p_asymptotic` is that Gumbel p-value. `p_value` is the p-value under an exponential tail, simulated from the same representation as for the other tests (given the `(2k + 1)`-th latest time, the `2k` later ones are the order statistics of `2k` independent exponentials), which is more reliable than the asymptotic one when `k` is small. Like `"neves"` it does not estimate the extreme value index.

Small values of `T*` and `G*` and large values of `W` point to a finite end-point. The p-value is for `alternative`: `"finite"` (default) is the Weibull domain, `"infinite"` the Frechet domain and `"two.sided"` either (twice the smaller one-sided p-value, as in the paper's two-sided rule, capped at 1).

A non-significant result does not show that the end-point is infinite: with only a handful of sightings the tests have little power against a finite end-point (see the simulations in Neves et al., 2006, and in Hasofer & Wang, 1992). Use it to flag records for which the finite end-point model is contradicted, not to confirm it.

By default each distinct sighting time counts once, as in `hall_wang1999()`; set `events = TRUE` to count every sighting event.

## Value

An object of class `ede_endpoint_test`, a list with a data frame `table` (columns `test`, `k`, `statistic`, `p_value`, `p_asymptotic`), `alternative`, `k`, and the number of sighting times `n`.

## References

Neves, C., Picek, J., & Fraga Alves, M. I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction. Journal of Statistical Planning and Inference, 136(4), 1281-1301. doi:10.1016/j.jspi.2004.09.008

Hasofer, A. M., & Wang, Z. (1992). A test for extreme value domain of attraction. Journal of the American Statistical Association, 87(417), 171-177. doi:10.2307/2290466

Fraga Alves, I., Neves, C., & Rosario, P. (2016). A general estimator for the right endpoint with an application to supercentenarian women's records. arXiv:1412.3972 (version 3).

Fisher, R. A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society A, 125(796), 54-59.

## See Also

`hall_wang1999()` and `ole()` for the Weibull domain, `fraga_alves_neves2014()` for the Gumbel domain.

## Examples

```
sd <- sighting_data(data.frame(year = c(1900, 1904, 1909, 1911, 1913, 1918,
                                        1920, 1925, 1928, 1930, 1934),
                               sightings = 1))
ede_endpoint_test(sd)
ede_endpoint_test(sd, k = 6, test = "neves", alternative = "two.sided")
```

---

`ede_estimate` *Extinction date estimate*

---

## Description

Common S3 result class returned by every estimator in this package.

## Format

A list with components:

**estimate** Point estimate, or `NA` if not defined for this method.

**lower, upper** Interval bounds, or `NA` if not defined. An upper bound that does not exist is returned as `Inf`.

**interval_type** Either `"two-sided"` or `"one-sided"`.

**method** Character string identifying the estimator.

**alpha** Significance level used to compute the estimate.

Bayesian estimators (`strauss1989_bayes()`, `weiss_marshall1999()`) return posterior probability (credible) intervals and set `interval_kind` to `"credible"`. Some estimators append method-specific components, documented on their help pages.

## Details

Persistence tests (`solow1993()`, `solow2005()`, `mcinerny2006()`, `burgman1995()`, `solow_roberts2003()` and `jaric2010()`) do not define a point estimate. Inverting the test gives a one-sided `1 - alpha` upper confidence bound for the extinction time, which is returned in `upper` with `estimate = NA`. `solow1993b()` also returns the minimum variance unbiased estimate of its model in `estimate`.

ede_evi *Moment estimates over a range of k (stability diagram)*

## Description

Applies the moment estimator of dekkers1989() for every k in a range, the usual diagnostic for choosing k in extreme-value analysis: look for a range of k over which the estimate of the extreme-value index is stable.

## Usage

```
ede_evi(
  sd,
  kmin = 3L,
  kmax = NULL,
  events = FALSE,
  origin = NULL,
  alpha = 0.05
)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| kmin, kmax | Smallest and largest k; kmax = NULL uses the largest admissible value (the threshold needs one more sighting later than origin). |
| events | If FALSE (default), count each distinct sighting time once; if TRUE, count every sighting event. |
| origin | Origin of the time axis, earlier than the threshold sighting k + 1; NULL (default) uses the earliest time in sd. |
| alpha | Significance level, in (0, 1), of the normal interval for gamma. |

## Value

An object of class ede_evi, a data frame with columns k, gamma, se, lower, upper (normal interval for gamma) and endpoint (NA where the index is not negative), with attributes alpha and origin. plot() draws the estimates against k.

## References

Dekkers, A. L. M., Einmahl, J. H. J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. Annals of Statistics, 17(4), 1833-1855. doi:10.1214/aos/1176347397

## See Also

dekkers1989()

## Examples

```
sd <- sighting_data(data.frame(year = c(1900, 1904, 1909, 1911, 1913, 1918,
                                        1920, 1925, 1928, 1930, 1934),
                               sightings = 1))
ede_evi(sd)
plot(ede_evi(sd))
```

---

`ede_field_test` *Combine the PPCC tests of several records (field significance)*

---

## Description

Combines the individual p-values of tests applied to several records, for example the `vogel2009()` test of a constant sighting rate on many species, with the four field-significance procedures of Vogel et al. (2009, Section 6): a Bonferroni-type test, the false discovery rate procedure of Benjamini & Hochberg (1995), a binomial test of the number of rejections, and a PPCC test of the uniformity of the p-values themselves. With a single short record each test has little power, which is the reason for combining many. The procedures assume the records, hence the p-values, are independent.

## Usage

```
ede_field_test(p, alpha = 0.05, nsim = 1e+05)
```

## Arguments

| | |
|---|---|
| `p` | Numeric vector of p-values, one per record. |
| `alpha` | Significance level of each individual test (used for the number of rejections in the binomial test and the Bonferroni threshold, with the field level `alpha`), in (0, 1). |
| `nsim` | Number of draws used for the null distribution of the PPCC of the p-values. |

## Value

A data frame with one row per procedure and columns `test`, `statistic`, `rejections` (the number of individual tests rejected, where defined) and `p_value` (the field significance level). For the Bonferroni and false discovery rate procedures `p_value` is `NA`: they report the records rejected at field level `alpha`.

## References

Vogel, R. M., Hosking, J. R. M., Elphick, C. S., Roberts, D. L., & Reed, J. M. (2009). Goodness of fit of probability distributions for sightings as species approach extinction. Bulletin of Mathematical Biology, 71(3), 701-719. doi:10.1007/s11538-008-9377-3

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B, 57(1), 289-300.

## See Also

vogel2009()

## Examples

```
p <- c(0.92, 0.65, 0.08, 0.90, 0.43, 0.017, 0.001, 0.015, 0.74, 0.99)
ede_field_test(p)
```

---

ede_select_k *Choose the number of recent sightings for extreme-value estimators*

---

## Description

Extreme-value estimators such as ole() and hall_wang1999() use only the k most recent sightings, because only the extreme order statistics of the sighting times follow the Weibull extreme-value model. Too large a k biases the estimate; too small a k makes it imprecise. ede_select_k() chooses k automatically, adapting to this problem the threshold selection of Bader, Yan & Zhang (2018): goodness-of-fit tests of the Weibull model are run for k = kmax, kmax - 1, ..., kmin, from the most to the least data, and the ForwardStop rule of G'Sell et al. (2016) controls the false discovery rate of the ordered sequence of tests. The chosen k is the largest one that is not rejected.

## Usage

```
ede_select_k(
  sd,
  kmin = 4L,
  kmax = NULL,
  alpha = 0.1,
  events = FALSE,
  nsim = 499L
)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| kmin, kmax | Smallest and largest k examined. kmin is at least 4; kmax defaults to the number of sightings, at most 30. |
| alpha | False discovery rate of the ForwardStop rule, in (0, 1). |
| events | If FALSE (default), count each distinct sighting time once; if TRUE, count every sighting event. |
| nsim | Number of bootstrap draws per value of k.<br>In simulations from the Weibull model the single tests reject at close to their nominal level and the procedure keeps the largest k in about nine cases out of ten; against a record whose older sightings come from a much more diffuse regime its power is moderate. A k for which the minimum-distance estimator |

of hall_wang1999() does not exist can still be selected, because a record with widely spread older sightings is compatible with the Gumbel limit of the model; the estimable column of the table flags these cases.

## Details

The test for a given k is a parametric bootstrap. Its statistic is the minimum over candidate endpoints of Greenwood's statistic of the log-spacings (the criterion of hall_wang1999()), which is large when the k latest sightings do not follow the model. The null distribution is simulated from the Weibull model fitted to the same k sightings (extremal exponent estimated by the minimum-distance method), so each p-value is approximate: it ignores the uncertainty of the fitted exponent.

This is a package construct that applies the ordered-testing idea of Bader et al. (2018), developed for generalised Pareto thresholds, to the sighting-record model; it is not a published method for sighting records. The procedure is random: call set.seed() first for reproducible results.

## Value

An object of class ede_select_k: a list with k (the selected value, NA if every k is rejected), table (a data frame with columns k, statistic, shape, p_value, estimable and rejected), alpha and method.

## References

Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. The Annals of Applied Statistics, 12(1), 310-329. doi:10.1214/17-AOAS1092

G'Sell, M. G., Wager, S., Chouldechova, A., & Tibshirani, R. (2016). Sequential selection procedures and false discovery rate control. Journal of the Royal Statistical Society B, 78(2), 423-444.

## See Also

hall_wang1999(), ole()

## Examples

```
sd <- sighting_data(data.frame(year = c(1880, 1900, 1904, 1909, 1911, 1913,
                                        1918, 1920, 1925, 1928, 1930, 1934),
                               sightings = 1))
set.seed(1)
sel <- ede_select_k(sd, nsim = 99)
sel
```

`extinct_birds` *Sighting records of four possibly extinct birds, by evidence quality*

## Description

Sighting years of the Eskimo Curlew, the Ivory-billed Woodpecker (U.S.), the Nukupu'u and the O'ahu 'Alauahio, classified by Roberts et al. (2010) into physical evidence, independent expert opinion and controversial sightings, as tabulated by Jaric & Roberts (2014). Jaric & Roberts assign reliabilities of 0.8-0.9, 0.6-0.8 and 0.1-0.4 to the three classes. A year can appear under more than one class.

## Usage

```
extinct_birds
```

## Format

A data frame with columns `species`, `scientific_name`, `year` and `evidence` (`"physical"`, `"expert"` or `"controversial"`).

## Source

Jaric, I., & Roberts, D. L. (2014). Accounting for observation reliability when inferring extinction based on sighting records. Biodiversity and Conservation, 23(11), 2801-2815, Table 2.

Roberts, D. L., Elphick, C. S., & Reed, J. M. (2010). Identifying anomalous reports of putatively extinct species and why it matters. Conservation Biology, 24(1), 189-196.

## Examples

```
curlew <- extinct_birds[extinct_birds$species == "Eskimo Curlew", ]
# keep the most reliable class for each year
curlew <- curlew[order(curlew$year, match(curlew$evidence,
  c("physical", "expert", "controversial"))), ]
curlew <- curlew[!duplicated(curlew$year), ]
reliability <- c(physical = 0.85, expert = 0.7, controversial = 0.25)
jaric_roberts2014(
  sighting_data(data.frame(year = curlew$year, count = 1)),
  reliability = reliability[curlew$evidence]
)
```

## ferret *Sighting record of the black-footed ferret*

### Description

Monthly sightings of the black-footed ferret (*Mustela nigripes*) in South Dakota from 1972 to 1984, one row per month with a sighting, used by Solow (1993b) and Jaric & Ebenhard (2010).

### Usage

```
ferret
```

### Format

A data frame with 28 rows: `year`, `month` (1-12) and `count`.

### Details

The two papers index months differently. Solow (1993b) counts months from January 1972 (= 1), starts observation at 0 and tests to `T = 229`; Jaric & Ebenhard (2010) count from June 1972 (= 1) and test to December 1990 (`T = 223`).

### Source

Solow, A. R. (1993b). Inferring extinction in a declining population. Journal of Mathematical Biology, 32(1), 79-82.

Jaric, I., & Ebenhard, T. (2010). A method for inferring extinction based on sighting records that change in frequency over time. Wildlife Biology, 16(3), 267-275.

### Examples

```
# Solow (1993b): months from January 1972, observation starting at 0
months <- (ferret$year - 1972) * 12 + ferret$month
sd <- sighting_data(data.frame(time = c(0, months), count = c(0, ferret$count)))
tail(solow1993b(sd, test_year = 229, data_out = TRUE), 1)
```

## fraga_alves_neves2014 *Fraga Alves & Neves (2014) end-point estimator for the Gumbel domain*

### Description

Estimates the extinction time as the finite right end-point of the sighting time distribution when that distribution lies in the Gumbel domain of attraction: a density that vanishes at the end-point faster than any power of the distance to it (for example an exponentially declining sighting rate that is cut off at the extinction time). It complements `ole()` and `hall_wang1999()`, which assume the Weibull domain, a density that behaves like a power of the distance to the end-point; the Weibull models become increasingly Gumbel-like as their extremal exponent grows.

## Usage

```
fraga_alves_neves2014(sd, k = NULL, events = FALSE)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| k | Number of top order statistics defining the high threshold, a positive integer with 2k at most the number of sightings, or NULL for the largest admissible value (floor(n / 2)). |
| events | If FALSE (default), count each distinct sighting time once; if TRUE, count every sighting event. |

## Details

With the sighting times ordered t[1] > t[2] > ... (latest first) the estimate is t[1] + sum_{i = 0}^{k - 1} a[i] * (t[k + with weights a[i] = (log(k + i + 1) - log(k + i)) / log(2) that are positive and sum to one. It is therefore never earlier than the last sighting, which the usual estimators for the Weibull domain do not guarantee, and it uses the 2k latest sightings.

The limiting distribution of the estimator depends on unknown auxiliary functions of the sampling distribution, so Fraga Alves & Neves (2014) provide no interval, and neither does this function: lower and upper are NA.

Each distinct sighting time counts once by default, as in robson1964(); set events = TRUE to count every sighting event, as ole() does.

## Value

An ede_estimate object with estimate and no interval, plus the component k.

## References

Fraga Alves, I., & Neves, C. (2014). Estimation of the finite right endpoint in the Gumbel domain. Statistica Sinica, 24(4), 1811-1835. doi:10.5705/ss.2013.183

## See Also

hall_wang1999() and ole() for the Weibull domain.

## Examples

```
sd <- sighting_data(data.frame(year = c(1900, 1904, 1909, 1911, 1913, 1918,
                                        1920, 1925, 1928, 1930, 1934),
                               sightings = 1))
fraga_alves_neves2014(sd)
fraga_alves_neves2014(sd, k = 3)
```

hall_wang1999 *Hall & Wang (1999) minimum-distance end-point estimator*

## Description

Estimates the extinction time as the end-point of the sighting-time distribution, from the k most recent sightings, by the minimum-distance method of Hall & Wang (1999). The k latest sightings are taken to follow the limiting Weibull extreme-value model, the same model that underlies ole(). If t[1] > t[2] > ... > t[k] are the sighting times, latest first, and theta is a candidate end-point, then Y[i] = i * (log(theta - t[i + 1]) - log(theta - t[i])) for i = 1, ..., k - 1. At the true end-point the Y[i] are independent exponential variables with a common mean. The estimate is the theta that minimises Greenwood's statistic G = sum(Y^2) / sum(Y)^2 of the Y[i].

## Usage

```
hall_wang1999(
  sd,
  alpha = 0.05,
  k = NULL,
  events = FALSE,
  nsim = 1e+05,
  data_out = FALSE,
  test_year = NULL
)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| alpha | Significance level, in (0, 1). The interval has coverage 1 - alpha. |
| k | Number of most recent sightings to use, an integer of at least 4, or NULL for the 10 latest (all, if there are fewer). |
| events | If FALSE (default), count each distinct sighting time once; if TRUE, count every sighting event. |
| nsim | Number of draws used to compute the critical values of Greenwood's statistic. |
| data_out | If TRUE, return instead the curve of the statistic and of the p-value of candidate end-points between the last sighting and test_year (the confidence interval is the set where p_value > alpha). |
| test_year | Upper end of the curve, later than the last sighting; required when data_out = TRUE. |

## Details

Unlike maximum likelihood, the criterion does not involve the extremal exponent (the shape of the Weibull model), which makes the estimator well defined much more often than maximum likelihood when k is small. For the same reason the distribution of G at the true end-point does not depend on

any nuisance parameter, so the confidence interval `{theta : G(theta) < g}`, with `g` the `1 - alpha` quantile of that distribution, has exact coverage when the model holds. The package computes the quantile by simulation (`nsim` draws, with a fixed internal seed, so the result is reproducible and the global random number stream is left untouched). The distribution of Greenwood's statistic has no closed form; with the default `nsim` the simulated quantiles agree to three decimals with the exact points of Burrows (1979) for 10 uniform spacings (as quoted by Stephens, 1981), and are more accurate than the Pearson-curve points of Stephens (1981) in the lower tail for small samples. The Monte Carlo error of a tail probability near 0.05 is about 0.0007.

The estimator does not exist when `G` has no minimum below the end-point at infinity, which happens for some records; `estimate` is then `NA` with a warning. The interval is still reported. An upper limit that does not exist is returned as `Inf` with a warning, as in `ole()`.

The model assumes distinct (continuous) sighting times, so by default each distinct sighting time counts once, as in `robson1964()`; set `events = TRUE` to count every sighting event instead, as `ole()` does. With few sightings Hall & Wang recommend a small `k`, because the distribution of the sighting times may change over a long record; by default the 10 latest sightings are used, following the usual practice for `ole()` (Solow 2005). `ede_select_k()` chooses `k` from the data.

Hall & Wang (1999, section 3) report their example with the 90% interval; it is reproduced with `alpha = 0.1`.

## Value

An ede_estimate object with components `estimate`, `lower`, `upper`, plus `shape` (estimated extremal exponent), `k`, `statistic` (minimum of Greenwood's statistic) and `critical` (its critical value). With `data_out = TRUE`, a data frame with columns `time`, `statistic` and `p_value`.

## References

Hall, P., & Wang, J. Z. (1999). Estimating the end-point of a probability distribution using minimum-distance methods. Bernoulli, 5(1), 177-189. doi:10.2307/3318618

Solow, A. R. (2005). Inferring extinction from a sighting record. Mathematical Biosciences, 195(1), 47-55.

## See Also

`ole()`, which fits the same model by optimal linear estimation; `fraga_alves_neves2014()` for the Gumbel domain of attraction; `ede_select_k()` to choose `k`.

## Examples

```
sd <- sighting_data(data.frame(year = c(1900, 1904, 1909, 1911, 1913, 1918,
                                        1920, 1925, 1928, 1930, 1934),
                               sightings = 1))
hall_wang1999(sd)
hall_wang1999(sd, k = 6, alpha = 0.1)
```

`jaric2010` *Jaric & Ebenhard (2010) sighting-trend index*

## Description

Index of persistence from the average interval `I` between distinct sighting times, optionally adjusted by the trend coefficient `c`. For candidate time `T`, the index is `I / (I + T - t[n])` (equations 3-4 of Jaric & Ebenhard, 2010) or, with `trend = TRUE`, `(I + c) / (I + c + T - t[n])` (equations 5-6). The paper calls this the probability that the species is still present; it is an index rather than a p-value from a fitted probability model.

## Usage

```
jaric2010(sd, alpha = 0.05, test_year, data_out = FALSE, trend = TRUE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). The returned upper bound is the earliest candidate time at which persistence is rejected at this level. |
| `test_year` | Latest time to test. Must be later than the last sighting. A bound beyond `test_year` is reported as `NA` with a warning. |
| `data_out` | If `TRUE`, return the p-value curve at integer steps after the last sighting instead of the upper bound. |
| `trend` | Logical. If `TRUE` (the default), adjust the average interval by the trend coefficient (equations 5-6 and 8). |

## Details

Several sightings in one time unit count as one sighting, as in the paper. `I` is the mean interval between the first and last sighting times, so the observation origin is not used. `c` is the mean change in consecutive interval lengths, which reduces to the difference between the last and first intervals divided by `n - 2`; it is positive when sightings become less frequent.

The returned upper bound is `t[n] + (I + c) (1 - alpha) / alpha`, the upper bound of the confidence interval in equations 7-8 of the paper (`c = 0` when `trend = FALSE`).

## Value

An ede_estimate object with `estimate = NA`, the one-sided upper bound in `upper`, and additional components `average_interval` and `trend_coefficient`; or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (index values; `chance` is a deprecated alias) carrying the same two quantities as attributes.

## References

Jaric, I., & Ebenhard, T. (2010). A method for inferring extinction based on sighting records that change in frequency over time. Wildlife Biology, 16(3), 267-275.

`jaric_roberts2014` *Reliability-adjusted extinction inference*

## Description

Modifies the constant-rate Solow (1993a) model by assigning a reliability `p[i]` in `[0, 1]` to each sighting. The number of sightings is replaced by the effective number `r = sum(p[i])` (equation 4 of Jaric & Roberts, 2014) and the last sighting time by its expectation `t[r]` over which record is the last valid one (equations 5-6). The point estimate is `(r + 1) / r * t[r]` (equation 9) and the upper confidence bound is `t[r] / alpha^(1 / r)` (equation 10). For a candidate time `T`, the modified Solow p-value is `(t[r] / T)^r` (equations 7-8), with times measured from the observation origin.

## Usage

```
jaric_roberts2014(
  sd,
  reliability,
  alpha = 0.05,
  test_year = NULL,
  data_out = FALSE
)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. Counts must be zero or one. This is a restriction of the package, not of the paper: reliability belongs to an individual record, and the paper does not say how to combine several records of different reliability in one time unit. |
| `reliability` | Numeric reliabilities in `[0, 1]`, supplied either for every row of `sd` or for the positive-count rows only. |
| `alpha` | Significance level, in (0, 1), for the upper confidence bound. |
| `test_year` | Optional candidate time at which to compute the p-value of equations 7-8. Must be later than the last sighting. Required when `data_out = TRUE`. |
| `data_out` | If `TRUE`, return the p-value curve at integer steps after the last sighting up to `test_year` instead of the estimate. |

## Details

The earliest supplied time defines the observation origin. A sighting at that time fixes the origin and does not enter `r` or `t[r]`, so its reliability is not used; prepend a zero-count row to include it. Reliabilities given for zero-count rows are ignored.

## Value

An ede_estimate object with additional components `effective_sightings` (`r`), `effective_endpoint` (`t[r]`, on the original time scale) and, if `test_year` is given, `p_value`; or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (plus `chance`, a deprecated alias of `p_value`).

## References

Jaric, I., & Roberts, D. L. (2014). Accounting for observation reliability when inferring extinction based on sighting records. Biodiversity and Conservation, 23(11), 2801-2815.

---

`kodikara2020` *Kodikara et al. (2020) Bayesian extinction year from yearly sighting records*

---

## Description

Posterior distribution of the extinction year from a record of yearly certain and (optionally) uncertain sightings, following Kodikara, Demirhan, Wang, Solow & Stone (2020). Each year `t = 1, ..., T` of the observation period is a trial. `tau_E` is the first year following extinction: the species is extant in years `1, ..., tau_E - 1`, so `tau_E` is at least one year after the last certain sighting `c_n`.

## Usage

```
kodikara2020(certain, uncertain = NULL, test_year, model = NULL, level = 0.95)
```

## Arguments

| | |
|---|---|
| `certain` | A sighting_data object with whole-number times (years) and the certain sightings. Its earliest time is the first year of observation. At least one certain sighting is required. |
| `uncertain` | A sighting_data object with the uncertain sightings, or `NULL`. Ignored by Model 1. |
| `test_year` | Last year `T` of the observation period. Must be later than the last certain sighting and not earlier than the last uncertain sighting. |
| `model` | `1`, `2` or `3` (see Description). Defaults to `1` without uncertain sightings and `2` otherwise. |
| `level` | Probability content of the HDIs. |

## Details

- **Model 1** (certain sightings only, section 2.1). An extant species is sighted (with certainty) in a year with probability `pc`. With `nc` certain sighting years, the likelihood is equation 3: `0` for `tau_E <= c_n`, `pc^nc (1 - pc)^(tau_E - 1 - nc)` for `c_n < tau_E <= T`, and `pc^nc (1 - pc)^(T - nc)` for `tau_E > T`.
- **Model 2** (certain and uncertain sightings, section 2.2). Before extinction, valid and invalid uncertain sightings occur independently with probabilities `puv` and `pui`, so an uncertain sighting occurs with probability `pu = puv (1 - pui) + pui (1 - puv) + puv pui`, and is recorded only in years without a certain sighting (the "censoring" process). After extinction, only invalid uncertain sightings occur, with probability `pui`. With `nu(tau_E)` uncertain sightings before `tau_E` and `Nu` in total, the likelihood is equation 9: `pc^nc ((1 - pc) pu)^nu(tau_E) ((1 - pc)(1 - pu))^(tau_E - 1 - nc) pui^(Nu - nu(tau_E)) (1 - pui)^(T - (tau_E - 1) - (Nu - nu(tau_E)))` for `c_n < tau_E <= T`, and `pc^nc ((1 - pc) pu)^Nu ((1 - pc)(1 - pu))^(T - nc - Nu)` for `tau_E > T`.

- **Model 3** (section 4.2, Table 4): Model 2 with puv = pc.

The prior of tau_E is geometric with parameter theta (equation 4), p(tau_E | theta) = (1 - theta)^(tau_E - 1) theta, and theta, pc, puv and pui are standard uniform (equations 5-6); the joint posterior is equation 7.

The paper samples this posterior with JAGS. Because tau_E is discrete and all priors are uniform, the posterior of tau_E is computed here exactly instead (a package construct): integrating theta gives the marginal prior p(tau_E) = 1 / (tau_E (tau_E + 1)), whose tail beyond T has mass 1 / (T + 1); the integrals over pc, puv and pui are finite sums of Beta functions, obtained by writing 1 - (1 - puv)(1 - pui) = puv + (1 - puv) pui and expanding its power binomially (all terms positive). Since the likelihood is constant for tau_E > T, the posterior of tau_E beyond T is proportional to 1 / (tau_E (tau_E + 1)), so its quantiles there are analytic. The posterior median and highest posterior density interval (HDI, the shortest run of consecutive years with posterior probability at least level) of tau_E are reported as years, as in Tables 2-4 of the paper.

**Input.** Periods are the integer years from the earliest time in certain to test_year (T); include a zero-count row in certain when observation began before the first certain sighting (as for the O'ahu 'Alauahio record, which starts in 1837). A year is a sighting year if its count is positive (several sightings in a year count once); unlisted years have no sighting. Uncertain sightings in a year that also has a certain sighting are ignored, as the censoring process of Model 2 assumes they cannot be recorded.

## Value

A list of class ede_bayes with components:

- posterior: posterior probability p(tau_E > T | S) that the species was extant in every year of the record (one minus the paper's p(tau_E <= T | S)).
- prior: the implied prior of that event, 1 / (T + 1).
- bayes_factor: posterior odds over prior odds of persistence.
- test_year, method and model.
- extinction_year: posterior median and level HDI of tau_E, as years (c(median, lower, upper)); level.
- extinction_time: data frame with the posterior probability of each value of tau_E (time, as a year) from the year after the last certain sighting to test_year.
- p_certain: posterior median and level HDI of pc (Models 1 and 2; NA for Model 3).

## References

Kodikara, S., Demirhan, H., Wang, Y., Solow, A. & Stone, L. (2020). Inferring extinction year using a Bayesian approach. Methods in Ecology and Evolution, 11(8), 964-973. doi:10.1111/2041-210X.13408

## Examples

```
# Ivory-billed Woodpecker, physical-evidence sightings (Kodikara et al.,
# 2020, Table 2)
ibw_c <- sighting_data(data.frame(
  year = c(1898:1902, 1904:1910, 1913, 1914, 1917, 1924, 1925, 1932,
           1935, 1938, 1939),
```

```
  count = 1))
fit <- kodikara2020(ibw_c, test_year = 2010)
fit$extinction_year
```

---

`kodikara2021` *Kodikara et al. (2021) extinction time with non-homogeneous Poisson sightings*

---

## Description

Posterior distribution of the extinction time `tau_E` when certain and uncertain sightings follow independent non-homogeneous Poisson processes with Weibull (power-law) rates, and `tau_E` is a change-point for the uncertain sightings (Kodikara, Demirhan, Wang & Stone, 2021). Times are measured from the observation origin, which is the earliest time in `certain`. The observation period is `(0, T)`.

## Usage

```
kodikara2021(
  certain,
  uncertain = NULL,
  test_year,
  formulation = c("published", "equations"),
  homogeneous = FALSE,
  rate = 0.005,
  sigma_max = 1000,
  alpha_max = 1000,
  level = 0.95,
  n_grid = 200
)
```

## Arguments

| | |
|---|---|
| `certain` | A sighting_data object with the certain sightings. Its earliest time is the origin; at least one certain sighting must be later than it. |
| `uncertain` | A sighting_data object with the uncertain sightings, or `NULL` (certain sightings only, equation 10). Zero-count rows are ignored. |
| `test_year` | End `T` of the observation period. Must be later than the last certain sighting and not earlier than any uncertain sighting. |
| `formulation` | `"published"` (the authors' code; reproduces the paper) or `"equations"` (equation 9 as printed). |
| `homogeneous` | If `TRUE`, all shape parameters are fixed at 1. Only available without uncertain sightings: the paper's value for the homogeneous model with uncertain sightings (Figure 7) is not reproduced. |
| `rate` | Rate of the exponential prior of `tau_E`, per time unit. |

sigma_max, alpha_max
: Upper bounds of the uniform priors of the scale and shape parameters.

level
: Probability content of the HDIs.

n_grid
: Approximate number of grid points for tau_E in (last certain sighting, T).

## Details

- Certain sightings occur before tau_E with rate lambda_c(t) = (alpha_c / sigma_c) (t / sigma_c)^(alpha_c - 1) and mean function m_c(t) = (t / sigma_c)^alpha_c (equations 3 and 5).
- Uncertain sightings occur with rate lambda_u1 (parameters alpha_u1, sigma_u1) before tau_E and lambda_u2 (alpha_u2, sigma_u2) after it (equations 4 and 6).
- The likelihood is equation 9: with Kc certain sightings at times y_ci, and Ku(tau_E) of the Ku uncertain sightings before tau_E, prod_i lambda_c(y_ci) exp(-m_c(tau_E)) x prod_(j <= Ku(tau_E)) lambda_u x prod_(j > Ku(tau_E)) lambda_u2(y_uj) exp(-(m_u2(T) - m_u2(tau_E))); without uncertain sightings only the first factor is used (equation 10). The likelihood is zero for tau_E before the last certain sighting, and is evaluated at tau_E = T for tau_E >= T.
- Priors: Uniform(0, alpha_max) for the shape parameters, Uniform(0, sigma_max) for the scale parameters (both 1000 in the paper) and an exponential prior with rate rate (0.005 in the paper) for tau_E (equation 11). homogeneous = TRUE fixes all shape parameters at 1 (constant rates; section 3.2 and Figure 7).

**Formulation.** The paper samples the posterior with JAGS. In the authors' code (R_script_Ferret.R, R_script_IBW.R), the likelihood is a product over sightings in which each sighting's factor contains the exponential term of its block. The term exp(-m_c(tau_E)) is therefore raised to the power Kc, exp(-m_u1(tau_E)) to the power Ku(tau_E) and exp(-(m_u2(T) - m_u2(tau_E))) to the power Ku - Ku(tau_E). The code also uses m_u1(tau_E) without truncating tau_E at T. formulation = "published" (the default) uses this likelihood and reproduces the published results. formulation = "equations" uses equation 9 as printed. Its results differ from the published ones (for the Ivory-billed Woodpecker, the posterior median of tau_E is 1946 instead of 1950), so it is not verified against the paper.

**Computation (package construct).** This function does not use MCMC. Given tau_E, the likelihood factorises into the certain, pre- and post-extinction uncertain blocks. In each block, the integral over the scale sigma is an upper incomplete gamma function, int_0^sigma_max sigma^(-alpha K) exp(-c sigma^(-alph The shape alpha is integrated numerically on a log-spaced grid. The posterior density of tau_E is evaluated on a grid between the discontinuities at the uncertain sighting times, and integrated with the trapezoidal rule. The density beyond T is integrated exactly for the equations (constant likelihood), and on a grid for the published formulation. The median and the HDI (the shortest interval with posterior probability level) are as reported in Tables 1 and 2 of the paper. The marginal posteriors of the shape parameters are summarised in the same way.

**Input.** Times are in the units of certain and uncertain. The priors depend on these units: the paper uses months for the black-footed ferret and years for the Ivory-billed Woodpecker. A count greater than one is that number of sightings at the same time. Every sighting must be later than the origin. Add a zero-count row in certain at the start of observation.

## Value

A list of class ede_bayes with components:

- posterior: posterior probability P(tau_E > T) of persistence beyond the observation period.
- prior: the prior probability of that event, exp(-rate (T - origin)).
- bayes_factor: posterior odds over prior odds of persistence.
- test_year, method, formulation and level.
- extinction_time: posterior median and HDI of tau_E, as times (c(median, lower, upper)).
- shape: posterior medians and HDIs of alpha_c, alpha_u1 and alpha_u2 (a matrix with one row for each estimated shape).
- density: data frame with the posterior density (density) and distribution function (cdf) of tau_E on the grid (time). Use it for P(tau_E <= t).

## References

Kodikara, S., Demirhan, H., Wang, Y. & Stone, L. (2021). Inferring extinction date of a species using non-homogeneous Poisson processes with a change-point. Methods in Ecology and Evolution, 12(3), 530-538. doi:10.1111/2041-210X.13542

## Examples

```
# Black-footed ferret, months since January 1972 (Kodikara et al., 2021,
# Table 1; dates from Solow, 1993)
months <- (ferret$year - 1972) * 12 + ferret$month
sd <- sighting_data(data.frame(month = c(0, months),
                               count = c(0, ferret$count)))
fit <- kodikara2021(sd, test_year = 229, n_grid = 200)
1972 + fit$extinction_time / 12
fit$shape
```

---

# lee2014 *Lee et al. (2014) Bayesian extinction probability with certain and uncertain sightings*

---

## Description

Posterior probability that a species is extant at the end of test_year, from a discrete record of certain sightings and, optionally, a second record of uncertain sightings that may be false positives (Lee, McCarthy, Wintle, Bode, Roberts & Burgman, 2014). The model is the OpenBUGS model of the paper's Appendix S1, computed exactly instead of by MCMC.

## Usage

```
lee2014(certain, uncertain = NULL, test_year, prior = 0.5, rate_max = 100)
```

## Arguments

| | |
|---|---|
| certain | A sighting_data object with the certain sightings, with whole-number times. It may contain only zero-count rows (no certain sightings). |
| uncertain | A sighting_data object with the uncertain sightings, or NULL (certain sightings only). |
| test_year | Last period of observation. Must be a whole number later than every sighting. |
| prior | Prior probability that the species is extant at the end of test_year, in (0, 1). |
| rate_max | Upper bound M of the uniform priors on the sighting rates (Appendix S1: 100). |

## Details

**Model** (Appendix S1; equations 5-12). The record is a series of periods `i = 1, ..., T` (e.g. years). In each period at most one sighting of each kind is recorded (equation 1). The extinction time is `te ~ U(0, T)` given extinction (equation 12) and the species is present in period `i` if it is extant at the end of the record or if `i <= te` (`step(te - i)` in Appendix S1). A certain sighting occurs in a period with probability `1 - exp(-mc)` while the species is present and `0` after extinction (equation 5); an uncertain sighting occurs with probability `1 - exp(-(mu + f))` while it is present and `1 - exp(-f)` after extinction (equation 6). The two records are independent given `te` (equations 7-8). The rates have independent uniform priors on `(0, rate_max)` (equation 11; Appendix S1 uses `rate_max = 100` for `mc`, `mu` and `f`), and the prior probability that the species is extant is `prior` (`Extant ~ dbern(0.5)` in Appendix S1, `pi` in the paper). The posterior is equation 2 with the Bayes factor of equation 3.

**Exact computation.** Given extinction, the likelihood depends on `te` only through the number of periods `k = floor(te)` with the species present, which is uniform on `0, ..., T - 1`. For each `k`, the certain likelihood is `(1/M) int_0^M (1 - e^-m)^S e^(-m (k - S)) dm` (equations 9-10 in discrete form), an incomplete Beta function, and is zero if a certain sighting occurs after period `k`. The uncertain likelihood is a double integral over `mu` and `f`: the integral over `mu` is again an incomplete Beta function and the integral over `f` is computed with `stats::integrate()`. The likelihood under extinction (equation 10) is the average over `k`; the likelihood under persistence is the case `k = T`.

**Input.** Periods are the integer time units from the observation origin, the earliest time in `certain` or `uncertain` (zero-count rows included), to `test_year`, both included. A period is a sighting if its count is positive (several sightings in one period count once). As in Appendix S1 (whose example record starts with a sighting), a sighting in the first period counts as a sighting; add a zero-count row to start observation before the first sighting. Uncertain sightings may occur anywhere in the record, including before the first certain sighting.

**Published conventions.** With these conventions the bars of Fig. 4 for the paper's own model (Birdlife records, observation from `t0` in Table 3 to 2011) are reproduced. The bars of Fig. 3 and the 0.697 quoted for the greater stick-nest rat are reproduced when the sighting at the origin is not counted (the origin given as a zero-count row, as in Table 2) and the record has `2011 - t0` periods, i.e. `test_year = 2010`.

Package constructs: the exact computation (the paper uses MCMC), the `extinction_time` output, and the requirement that `test_year` be later than every sighting. The paper's eqn 14 (continuous extinction time) and its prior for the false-positive rate built from the probability that a sighting is true (eqn 13, Table 1; used for the light grey bars of Fig. 4) are not implemented: the paper does

not give enough detail to reproduce those results. Appendix S1 has a single uncertain record; so has this function (the uncertain sightings are grouped together, as in the dark grey bars of Fig. 4).

## Value

A list of class `ede_bayes` with components `bayes_factor` (`l(s | H0) / l(s | H1)`, equation 3), `posterior` (probability that the species is extant, equation 2), `prior`, `test_year`, `method`, `periods` (`T`), `rate_max`, and `extinction_time`: a data frame with, for each period after the last certain sighting, the posterior probability that it is the first period in which the species is absent (these probabilities sum to `1 - posterior`).

## References

Lee, T. E., McCarthy, M. A., Wintle, B. A., Bode, M., Roberts, D. L. & Burgman, M. A. (2014). Inferring extinctions from sighting records of variable reliability. Journal of Applied Ecology, 51(1), 251-258.

## Examples

```
# Greater stick-nest rat (Lee et al. 2014, Table 2), observation from 1835
yrs <- c(1844, 1857, 1879, 1896, 1897, 1907, 1920, 1921, 1930, 1933,
         1938, 1986)
rat <- sighting_data(data.frame(year = c(1835, yrs),
                                count = c(0, rep(1, 12))))
lee2014(rat, test_year = 2010)$posterior  # Fig. 3: 0.697
# the 1986 sighting treated as uncertain
rat_c <- sighting_data(data.frame(year = c(1835, yrs[-12]),
                                  count = c(0, rep(1, 11))))
rat_u <- sighting_data(data.frame(year = 1986, count = 1))
lee2014(rat_c, rat_u, test_year = 2010)
```

---

`marshall1994` *Marshall (1994) distribution-free confidence intervals on a range end-point*

---

## Description

Confidence interval on the end-point of a stratigraphic (or sighting) range that does not assume randomly placed finds, only that gap sizes come from a continuous distribution with no trend along the section. The `C` confidence interval equals the `C`th percentile of the gap-size distribution (Marshall, 1994, p. 460), which is bracketed by order statistics of the `N` observed gaps. With `F(x)` the binomial `Binomial(N, C)` distribution function and `gamma = (1 - conf_prob) / 2`: the lower limit on the interval length is the `(x + 1)`th smallest gap, `x` the largest integer `< N` with `F(x) <= gamma` (equation 1); the upper limit is the `(x + 1)`th smallest gap, `x` the smallest integer with `F(x) >= 1 - gamma` (equation 2). With probability `conf_prob` (the paper's confidence probability `1 - 2 gamma`) the length of the `C` interval lies between the two. If no `x < N` satisfies equation 1 there is no lower limit; if only `x = N` satisfies equation 2 there is no upper limit and the interval is unbounded (e.g. a 95% interval needs at least 72 gaps at `conf_prob = 0.95`).

## Usage

```
marshall1994(
  sd,
  alpha = 0.05,
  conf_prob = 0.95,
  horizons = c("distinct", "count")
)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| alpha | One minus the confidence level C of the interval, in (0, 1). |
| conf_prob | Confidence probability 1 - 2 gamma with which the length of the C interval is bracketed, in (0, 1). The paper uses 0.95 or 0.99. |
| horizons | "distinct" (each distinct positive-count time is one horizon) or "count" (a count of k is k horizons at that time). |

## Details

The method gives a range of possible interval lengths rather than one interval. The returned upper is the conservative end, last find plus the upper-limit gap (Inf if there is none); upper_min is the last find plus the lower-limit gap, so the C bound lies in [upper_min, upper] with probability conf_prob. This mapping onto ede_estimate is a package construct.

Gaps are the distances between successive find horizons. By default (horizons = "distinct") each distinct time with a positive count is one horizon, as in strauss1989(). With horizons = "count" a row with count k stands for k horizons at the same (binned) time, i.e. k - 1 gaps of size zero, as in the paper's Metrarabdotos ages binned to 0.05 Myr. Zero-count rows are not used. The end point is the one at the largest time: for a first appearance, or for ages in Ma, code time = -age (or -height). The method assumes continuous sampling; it is not meant for discretely sampled cores (use weiss_marshall1999() for those). Check for a trend in gap size with marshall1994_trend() first.

## Value

An ede_estimate object with estimate = NA, lower = NA, upper (last find plus upper-limit gap, Inf if not available) and interval_type = "one-sided", plus components upper_min (last find plus lower-limit gap, NA if not available), ci_length (named lower/upper limits on the interval length), gap_ranks (ranks l and u of the gaps used, NA if not available), n_gaps, conf_prob and last_sighting.

## References

Marshall, C. R. (1994). Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. Paleobiology, 20(4), 459-469.

`marshall1994_trend` *Marshall (1994) transposition test for a trend in gap size*

## Description

Tests the assumption of `marshall1994()` (and of the classical intervals) that gap size is not correlated with stratigraphic position. Following Marshall (1994, p. 462-463), the gaps are listed in stratigraphic order and the number of transpositions `T` (pairs in which a larger gap precedes a smaller one) is counted; each group of `m` equal gaps adds `m (m - 1) / 4`. `T` near 0 indicates gaps increasing in size, `T` near `T_max = N (N - 1) / 2` gaps decreasing in size.

## Usage

```
marshall1994_trend(sd, horizons = c("distinct", "count"))
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `horizons` | `"distinct"` (each distinct positive-count time is one horizon) or `"count"` (a count of `k` is `k` horizons at that time). |

## Details

The paper compares `min(T, T_max - T)` with tabulated critical values (its Table 3, from Kreyszig 1979). Here a p-value is computed instead from the exact null distribution of the number of transpositions in a random ordering of `N` distinct gaps: `p = min(1, 2 P(I <= min(T, T_max - T)))`, with a non-integer `T` (from ties) rounded down. This two-sided p-value is a package construct. Table 3's critical values are two-sided (each tail `p / 2`): the exact distribution reproduces 67 of its 72 entries, and the other five differ by one at borderline probabilities.

## Value

A list of class `marshall1994_trend` with components `transpositions` (`T`, including the tie correction), `t_max`, `n_gaps`, `direction` (`"increasing"` if `T < T_max / 2`, `"decreasing"` if `T > T_max / 2`, otherwise `"none"`), `p_value` and `gaps` (in stratigraphic order).

## References

Marshall, C. R. (1994). Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. Paleobiology, 20(4), 459-469.

marshall1997 *Marshall (1997) generalized confidence interval for a range end-point*

## Description

Confidence interval on the end-point of a stratigraphic (or sighting) range when the probability of recovering a find varies along the axis according to a user-supplied fossil recovery potential `f(t)`. With `H` distinct find horizons, the observed range from the first (FAD) to the last (LAD) find carries recovery mass `R* = integral of f from FAD to LAD` (equation 10 of Marshall, 2010). The `C = 1 - alpha` interval extends beyond the LAD by the distance `r` whose recovery mass `r* = integral of f from LAD to LAD + r` (equation 11) equals `R* [(1 - C)^(-1 / (H - 1)) - 1]` (equation 9, the generalization of the classical interval, equation 4). If even the whole recovery mass beyond the LAD (up to `max_time`) is smaller than that, the interval is unbounded and `upper = Inf`, as in the short half-life case of Marshall (2010, Case 3). With constant `f` the method reduces to the classical interval of `strauss1989()`.

## Usage

```
marshall1997(
  sd,
  recovery = function(t) rep(1, length(t)),
  alpha = 0.05,
  test_year = NULL,
  max_time = Inf
)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `recovery` | A vectorized function of time giving the (relative) fossil recovery potential, non-negative. Only ratios of its integrals matter, so it need not be normalized. The default (constant) gives the classical interval. |
| `alpha` | Significance level, in (0, 1); the confidence level is `1 - alpha`. |
| `test_year` | Optional time after the last find at which to compute the confidence that the end-point had already been reached. |
| `max_time` | Upper limit of the axis on which `recovery` is defined (e.g. the top of the section). Defaults to `Inf`. |

## Details

The confidence that the end-point lies before a candidate time `T`, `C = 1 - [R* / (R* + r*)]^(H - 1)` with `r*` the recovery mass from the LAD to `T` (equations 3 and 6 with `R*`, `r*` in place of the range and its extension), is returned as `confidence` when `test_year` is given.

Each distinct time with a positive count is one find horizon; counts are ignored and zero-count rows are not used, as in `strauss1989()`. The end point estimated is always the one at the largest `time`:

to place an interval on a last appearance dated in Ma, code `time = -age`; for a first appearance, code `time = -height` (or `time = age`). `recovery` must be written on the same coded axis.

The formulas are from Marshall (2010, Box 3), who attributes the method to Marshall (1997); they have been checked against the worked example of Marshall (2010, section 4.1), not against Marshall (1997) itself. The integrals are computed numerically with `stats::integrate()`, so a discontinuous (e.g. stepwise) recovery potential should be supplied with breakpoints smoothed or with care.

## Value

An ede_estimate object with `estimate = NA`, `lower = NA`, `upper =` the end of the one-sided `1 - alpha` interval (possibly `Inf`) and `interval_type = "one-sided"`, plus components `last_sighting` (the LAD), `recovery_range` (`R*`), `recovery_extension` (the `r*` required) and, if `test_year` is given, `confidence`.

## References

Marshall, C. R. (1997). Confidence intervals on stratigraphic ranges with non-random distributions of fossil horizons. Paleobiology, 23(2), 165-173.

Marshall, C. R. (2010). Using confidence intervals to quantify the uncertainty in the end-points of stratigraphic ranges. In J. Alroy & G. Hunt (Eds.), Quantitative Methods in Paleobiology (pp. 291-316). The Paleontological Society Papers, 16.

---

# mccarthy1998 *McCarthy (1998) persistence test adjusted for sampling effort*

---

## Description

Persistence test of Solow (1993a) in which the time axis is replaced by the cumulative sampling effort, the "partial Solow equation" of McCarthy (1998, equation 2). If the probability that a record falls in a given time unit is proportional to the effort `e[i]` spent in it, then, under the hypothesis that the species persisted to the candidate time `T`, the probability that all `N` records fall no later than the last record `t[n]` is `p = (sum(e[0:t[n]]) / sum(e[0:T]))^N`, where the sums run over the time units from the observation origin. With constant effort it is the equation of `solow1993()` up to the discretisation (`(t + 1) / (T + 1)` in place of `t / T`, because the origin unit counts). Roberts & Jaric (2020) use the same test for species known from a single specimen, with the number of specimens collected as the effort index.

## Usage

```
mccarthy1998(sd, effort, alpha = 0.05, test_year, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object whose times are whole numbers. |
| effort | The sampling effort in each time unit, non-negative and not all zero: either a numeric vector with one value for each time unit from the origin (min(sd$time)) to at least test_year, in order, or a vectorised function of time returning the effort in each time unit. |
| alpha | Significance level, in (0, 1). The upper bound is the earliest time unit at which persistence is rejected at this level. |
| test_year | Latest time to test, a whole number later than the last sighting. A bound beyond test_year is reported as NA with a warning; extend effort and test_year to project further. |
| data_out | If TRUE, return the p-value at each time unit after the last sighting up to test_year instead of the upper bound. |

## Details

The count N and the origin are those of solow1993(): the earliest time of sd defines the origin, one sighting there is omitted, and every other sighting event is counted. The time unit is one whole number of the time scale (typically a year), as in McCarthy's data; sd$time must be integers and effort must be given for every time unit from the origin to test_year. Inverting the test gives the one-sided upper bound reported in upper: the earliest time unit after the last sighting at which p <= alpha.

Effort is only a proxy, and McCarthy and Roberts & Jaric report that the method is well calibrated when the extinction is sudden and effort is stable or rising, but biased when the decline is gradual or effort falls. The choice of the effort index (for instance the number of specimens of all species, or of related species, collected per year) is part of the analysis.

## Value

An ede_estimate object with estimate = NA and the one-sided upper bound in upper, or (if data_out = TRUE) a data frame with columns time and p_value (plus the deprecated alias chance).

## References

McCarthy, M. A. (1998). Identifying declining and threatened species with museum data. Biological Conservation, 83(1), 9-17. doi:10.1016/S0006-3207(97)00048-7

Roberts, D. L., & Jaric, I. (2020). Inferring the extinction of species known only from a single specimen. Oryx, 54(2), 161-166.

Solow, A. R. (1993a). Inferring extinction from sighting data. Ecology, 74(3), 962-964.

## See Also

solow1993(), the test with constant effort; marshall1997() for stratigraphic ranges with a recovery function.

## Examples

```
sd <- sighting_data(data.frame(year = c(1900, 1904, 1909, 1911, 1913, 1918,
                                        1920, 1925),
                               sightings = 1))
# effort doubles after 1915
effort <- ifelse(1900:1950 < 1915, 1, 2)
mccarthy1998(sd, effort, test_year = 1950)
solow1993(sd, test_year = 1950)
mccarthy1998(sd, effort, test_year = 1950, data_out = TRUE)[1:3, ]
```

---

`mcinerny2006` *McInerny, Roberts, Davy & Cribb (2006) sighting-rate persistence test*

---

## Description

A modification of the Solow (1993a) persistence test that conditions on the sighting rate observed up to the last sighting, `s = n / t[n]`, instead of on the length of the whole observation window. For candidate time `T`, the p-value is `(1 - s)^(T - t[n])` (equation 2 of the source paper). This makes the test comparable across records with very different total observation periods.

## Usage

```
mcinerny2006(sd, alpha = 0.05, test_year, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). |
| `test_year` | Latest time to test. Must be later than the last sighting. A bound beyond `test_year` is reported as `NA` with a warning. |
| `data_out` | If `TRUE`, return the p-value curve instead of the upper bound. |

## Details

The record is treated as binary: `n` counts time units with at least one sighting, as in the source paper. The earliest supplied time defines the observation origin. If that time has a sighting, it is used as the start date and not counted, so `n` is reduced by one, as in the paper. If the record starts with a zero-count row, that row is the known start date and every sighting time is counted.

The model is in whole time units, so the returned upper bound is the first time unit after the last sighting at which persistence is rejected, which is how the paper itself reports the test (Fig. 1).

## Value

An ede_estimate object with `estimate = NA` and the one-sided upper bound in `upper`, or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (plus `chance`, a deprecated alias of `p_value`).

## References

McInerny, G. J., Roberts, D. L., Davy, A. J., & Cribb, P. J. (2006). Significance of sighting rate in inferring extinction and threat. Conservation Biology, 20(2), 562-567.

---

`monk_seal` *Sighting record of the Caribbean monk seal*

---

## Description

Reliable sightings of the Caribbean monk seal (*Monachus tropicalis*) used as the example in Solow (1993a), Burgman et al. (1995) and Solow (2005). There is no natural start of the observation period, so the first sighting defines the origin.

## Usage

```
monk_seal
```

## Format

A data frame with 5 rows: `year` and `count`.

## Source

Solow, A. R. (1993a). Inferring extinction from sighting data. Ecology, 74(3), 962-964.

## Examples

```
solow1993(sighting_data(monk_seal), test_year = 2000)
solow1993_bayes(sighting_data(monk_seal), test_year = 1992)
```

---

`nz_shells` *Radiocarbon ages of shell samples from New Zealand*

---

## Description

Approximate calibrated ages and error standard deviations of 22 shell samples from archaeological sites in and around New Zealand, read by Solow (1997, Table 1) from Figure 3 of Higham & Hogg (1997), who used them as evidence of a late Polynesian settlement. The example of `solow2006()` for the settlement time. The ages are approximate, as in the source, and several are tied.

## Usage

```
nz_shells
```

## Format

A data frame with 22 rows: `sample` (laboratory code), `age_bp` (calibrated age in years before present) and `sd_yr` (standard deviation of the dating error, in years).

## Source

Solow, A. R. (1997). Estimating settlement time. Radiocarbon, 39(3), 351-354, Table 1.

## Examples

```
solow2006(data.frame(time = nz_shells$age_bp, error = nz_shells$sd_yr))
```

---

# ole *Optimal Linear Estimation of extinction date*

---

## Description

Estimates the extinction date by optimal linear estimation (Cooke, 1980), as applied by Roberts & Solow (2003). The joint distribution of the k most recent sighting times (the extreme order statistics) is assumed to have the Weibull extreme-value form; the estimate is the weighted sum of those times given by equations 12-15 of Solow (2005).

## Usage

```
ole(sd, alpha = 0.05, k = NULL)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level for the confidence interval, in (0, 1). |
| `k` | Number of most recent sighting events to use. Counts greater than one are treated as independent events at the recorded time; Solow (2005) recommends combining sightings that are obviously related. By default, all sighting events are used. Must be at least 3; Solow (2005) suggests at least 5, and notes that the extreme-value approximation may fail if `k` is too large. |

## Details

The confidence interval is the two-sided `1 - alpha` interval of Roberts & Solow (2003). Its upper limit does not exist when `-log(alpha / 2) >= k`; it is then returned as `Inf` with a warning. `solow2005()` inverts the corresponding test one-sidedly (Solow, 2005, equation 17), so its bound at level `alpha` equals the upper limit here at level `2 * alpha`.

## Value

An ede_estimate object.

## References

Cooke, P. (1980). Optimal linear estimation of bounds of random variables. Biometrika, 67(1), 257-258.

Roberts, D. L., & Solow, A. R. (2003). Flightless birds: When did the dodo become extinct? Nature, 426(6964), 245.

Solow, A. R. (2005). Inferring extinction from a sighting record. Mathematical Biosciences, 195(1), 47-55.

---

# orchid *Collection record of a Mauritian orchid*

---

## Description

Collections of *Angraecum* aff. *rutenbergianum*, one of the nine Mauritian orchids analysed by Solow & Roberts (2003). With 1996 as the end of the observation period, their nonparametric test gives a p-value of 0.10.

## Usage

```
orchid
```

## Format

A data frame with 11 rows: `year` and `count`.

## Source

Solow, A. R., & Roberts, D. L. (2003). A nonparametric test for extinction based on a sighting record. Ecology, 84(5), 1329-1332.

## Examples

```
solow_roberts2003(sighting_data(orchid), test_year = 2030)
```

---

# plot.ede_compare *Plot the results of* ede_compare()

---

## Description

A forest plot with one row per method: point estimate, interval and, for the persistence tests, the one-sided upper confidence bound (a triangle). Methods that could not be applied are omitted.

## Usage

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

### Arguments

| | |
|---|---|
| x | A result of ede_compare(). |
| ... | Further arguments passed to graphics::plot(). |

### Value

x, invisibly.

### Examples

```
plot(ede_compare(sighting_data(monk_seal), test_year = 2050))
```

---

## plot.ede_estimate *Plot an extinction date estimate*

---

### Description

Draws the point estimate and its interval on a time axis. An interval without a finite end (Inf) is drawn with an arrow. For a persistence test, which has no point estimate, only the upper confidence bound is drawn.

### Usage

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

### Arguments

| | |
|---|---|
| x | An ede_estimate object. |
| ... | Further arguments passed to graphics::plot(). |

### Value

x, invisibly.

### Examples

```
plot(ole(sighting_data(dodo)))
```

## plot.sighting_data *Plot a sighting record*

### Description

Draws the number of sightings recorded at each time, with the observation origin and the last sighting marked.

### Usage

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

### Arguments

| | |
|---|---|
| x | A sighting_data object. |
| ... | Further arguments passed to graphics::plot(). |

### Value

x, invisibly.

### Examples

```
plot(sighting_data(dodo))
```

## robson1964 *Robson & Whitlock (1964) truncation point estimator*

### Description

Estimates the extinction date by a Quenouille (jackknife-type) bias reduction of the most recent sighting. If the two most recent distinct sighting times are t[n-1] and t[n], the estimate is t[n] + (t[n] - t[n-1]) and the one-sided upper confidence bound is t[n] + (t[n] - t[n-1]) (1 - alpha) / alpha.

### Usage

```
robson1964(sd, alpha = 0.05)
```

### Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| alpha | Significance level, in (0, 1), for the approximate one-sided confidence interval. |

### Details

The estimator is mean-unbiased to order `n^-2` in general and exactly unbiased for a uniform distribution. The bound is exact for a uniform distribution and asymptotically valid when the density is continuous and positive at the endpoint; its coverage is below `1 - alpha` when the density decreases towards the endpoint, as with a declining sighting rate.

Only distinct sighting times are used; counts are ignored, so the paper's order statistics are applied to distinct times.

### Value

An ede_estimate object.

### References

Robson, D. S., & Whitlock, J. H. (1964). Estimation of a truncation point. Biometrika, 51(1/2), 33-39.

---

## `rout2009` *Rout et al. (2009) Bayes factor with a declining sighting rate*

---

### Description

Bayesian test of persistence (e.g. of an invasive species after an eradication programme) in which the pre-extinction sighting rate declines as `lambda(s) = m * s^(-a)` (equation 6 of Rout, Salomon & McCarthy, 2009), with `m` a constant and `0 <= a < 1`. Integrating over `m` (with the prior of Solow, 1993a) gives the Bayes factor in favour of persistence to `S`: `B = (n(1 - a) - 1) / ((S / s[n])^(n(1 - a) - 1) - 1)` (Rout et al., 2009, unnumbered equation after equation 6; derivation in their Appendix S1). With `a = 0` it is the Bayes factor of Solow (1993a) (Rout et al., equation 2), and this function then equals `solow1993_bayes()`. When `n(1 - a) = 1` the limit `1 / log(S / s[n])` is used. The posterior probability of persistence is `1 / (1 + (1 - prior) / (prior * B))` (equation 1).

### Usage

```
rout2009(sd, test_year, a = 0, prior = 0.5, data_out = FALSE)
```

### Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `test_year` | Candidate time `S` (e.g. the current survey number). Must be later than the last sighting. |
| `a` | Rate of decline of the sighting rate, in `[0, 1)`. Rout et al. (2009) estimated `a = 0.1` for *Helenium amarum*. |
| `prior` | Prior probability that the species persisted to `test_year`, in (0, 1). |
| `data_out` | If `TRUE`, return the Bayes factor and posterior probability at integer steps after the last sighting up to `test_year`. |

## Details

Rout et al. (2009) measure time in surveys: `S` surveys have been made and the species was seen in `n` of them, the last being survey `s[n]`. Here times are measured from the observation origin, the earliest time in `sd`, and sighting events are counted as in `solow1993()`: a sighting at the origin defines it and is not counted. To follow the paper, supply one row per survey with binary counts and a leading zero-count row for survey 0.

The paper's model has the rate declining with the number of surveys, so the result depends on where the origin is placed; this is a property of the model, not of the implementation.

## Value

A list of class `ede_bayes` with components `bayes_factor`, `posterior`, `prior`, `test_year`, `method` and `a`; or (if `data_out = TRUE`) a data frame with columns `time`, `bayes_factor` and `posterior`.

## References

Rout, T. M., Salomon, Y., & McCarthy, M. A. (2009). Using sighting records to declare eradication of an invasive species. Journal of Applied Ecology, 46(1), 110-117.

Solow, A. R. (1993a). Inferring extinction from sighting data. Ecology, 74(3), 962-964.

## See Also

`rout2009_absences()` for the number of absent surveys after which eradication should be declared.

---

# `rout2009_absences` *Optimal number of absent surveys before declaring eradication*

---

## Description

Decision rule of Rout, Salomon & McCarthy (2009): after `d` consecutive surveys without a sighting, the net expected cost of stopping, in units of the cost of escape and damage `C[e]`, is `NEC(d) / C[e] = (d - 1) R + p(extant` (equation 3, the "rule of thumb"), where `R = C[s] / C[e]` is the ratio of the cost of one survey to the cost of escape, and `p(extant | d)` is the posterior probability of persistence of `rout2009()` with `S = s[n] + d`. The optimal number of absent surveys `d*` minimises `NEC(d)` over `d = 0, 1, ..., max_absent`.

## Usage

```
rout2009_absences(sd, cost_ratio, a = 0, prior = 0.5, max_absent = 1000)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `cost_ratio` | The cost ratio `R = C[s] / C[e]`, positive. |
| `a` | Rate of decline of the sighting rate, in `[0, 1)`. Rout et al. (2009) estimated `a = 0.1` for *Helenium amarum*. |
| `prior` | Prior probability that the species persisted to `test_year`, in (0, 1). |
| `max_absent` | Largest number of absent surveys considered. |

## Details

For the constant-rate model (`a = 0`) the closed-form approximation `d* ~ s[n] ((n - 1)^2 prior / (s[n] (1 - prior) R))^` (equation 5) is also returned. The paper gives no approximation for `a > 0`, so it is `NA` then. The paper's stochastic dynamic programme is not implemented.

## Value

A list with components `optimal_absent` (`d*`), `stop_time` (last sighting time plus `d*`), `approximation` (equation 5, or `NA`), `nec` (a data frame with columns `absent`, `time`, `posterior` and `nec`, the latter in units of `C[e]`), `n`, `sn`, `a`, `prior` and `cost_ratio`.

## References

Rout, T. M., Salomon, Y., & McCarthy, M. A. (2009). Using sighting records to declare eradication of an invasive species. Journal of Applied Ecology, 46(1), 110-117.

---

# `seymour_ammonites` *Ammonite finds from Seymour Island*

---

## Description

Number of finds and approximate stratigraphic location of the lowest and highest find of the ten uppermost ammonite taxa of Seymour Island, Antarctica, as given in Table 1 of Solow (1996), who took them from Figure 5 of Macellari (1986) and, for taxa 3 and 8, from Strauss & Sadler (1989). The example of `solow1996()`. Taxa are numbered as in the table.

## Usage

```
seymour_ammonites
```

## Format

A data frame with 10 rows: `taxon` (number in the table), `n` (number of finds), `lowest` and `highest` (locations in meters of the lowest and the highest find).

## Source

Solow, A. R. (1996). Tests and confidence intervals for a common upper endpoint in fossil taxa. Paleobiology, 22(3), 406-410, Table 1.

## Examples

```
solow1996(seymour_ammonites)
```

## sighting_data *Construct a validated sighting record*

### Description

Builds the common input object used by every estimator in the package: a time-ordered table of sighting counts, checked for the conditions each estimator in the sighting-record literature assumes (numeric, non-negative, no duplicated times).

### Usage

```
sighting_data(data, time_col = 1L, count_col = 2L)
```

### Arguments

| | |
|---|---|
| data | A data frame or matrix. By default the first column is read as time (e.g. year) and the second as the number of sightings recorded at that time. The earliest supplied time defines the observation origin; include an initial zero-count row when observation began before the first sighting. |
| time_col, count_col | Column name or position for time and sighting count. |

### Value

An object of class sighting_data: a data frame with columns time and count, sorted by time.

## smith1985 *Maximum likelihood end-point and profile-likelihood interval (Smith, 1985)*

### Description

Fits the generalised Pareto model to the excesses of the k latest sightings over the (k + 1)-th latest, in the Weibull domain, and returns the maximum likelihood estimate of the end-point (the extinction time) with a profile-likelihood interval. This is the likelihood treatment of non-regular end-point problems of Smith (1985, Section 7), applied to sighting times.

### Usage

```
smith1985(sd, alpha = 0.05, k = NULL, events = FALSE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1); the interval has nominal coverage `1 - alpha`. |
| `k` | Number of latest sightings, an integer of at least 4 with `k + 1` at most the number of sighting times, or `NULL` for the 10 latest (fewer if the record is shorter). |
| `events` | If `FALSE` (default), count each distinct sighting time once; if `TRUE`, count every sighting event. |

## Details

With `u = t[k + 1]` the threshold (sightings ordered latest first) and `theta` the end-point, the excesses `y = t - u` satisfy `y / (theta - u) ~ Beta(1, a)`, where the shape `a = -1 / gamma` is the extremal exponent of `ole()` and `hall_wang1999()` (`gamma` is the extreme-value index, negative in the Weibull domain). Maximising over `a` for fixed `theta` gives the closed-form profile log-likelihood `k * log(a) - k * log(theta - u) - k - sum(log((theta - t) / (theta - u)))` with `a = -k / sum(log((theta - t) / (theta - u)))`, which is maximised over `theta`. The interval is `{theta : 2 * (l(theta_hat) - l(theta)) <= qchisq(1 - alpha, 1)}` on the profile `l`.

**Regularity.** The likelihood is unbounded as `theta` approaches the last sighting, so the maximum likelihood estimate is a local interior maximum, and it exists, and is asymptotically normal, only when the shape is larger than 2 (`gamma > -1/2`; Smith, 1985, Theorems 1 and 3). For `a < 2` the problem is non-regular, and for `a < 1` (`gamma < -1`) no consistent maximum likelihood estimator exists. A sighting rate that is constant until extinction gives `a = 1`, so in the setting that motivates most extinction models the likelihood treatment is outside its regular range. When the fitted shape is not larger than 2 the function warns that the chi-squared calibration of the interval is not justified; `hall_wang1999()`, whose interval is exact for any shape, is then the better choice. The estimate is `NA`, with a warning, when the profile likelihood has no interior maximum.

In simulations (300 records each, nominal 95% interval) the estimate existed in 61% to 97% of records with shape 3 to 4 and `k` of 15 to 40, in 53% with shape 2 and in 6% with shape 1 (`k` 10 to 15). Where it existed with shape above 2 the interval covered the true end-point in 91% to 96% of records, against 93% to 97% for `hall_wang1999()` in all cases.

An upper limit that does not exist is returned as `Inf`, as in `ole()`. By default each distinct sighting time counts once; set `events = TRUE` to count every sighting event.

## Value

An ede_estimate object with `estimate` (possibly `NA`), `lower`, `upper`, and components `shape` (the extremal exponent `a`), `gamma` (`-1 / a`), `k`, `loglik` (maximised profile log-likelihood) and `regular` (`TRUE` when the shape exceeds 2).

## References

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika, 72(1), 67-90. doi:10.1093/biomet/72.1.67

## See Also

`hall_wang1999()` and `ole()` for the same model with exact intervals; `ede_endpoint_test()` to test the Gumbel null.

## Examples

```
# the dodo record: shape near 2, so the interval is not justified (warning)
smith1985(sighting_data(dodo), k = 8)
```

---

`solow1993` *Solow (1993a) constant-rate persistence test*

---

## Description

Parametric test of the null hypothesis that a species was still extant at a candidate test year, under a stationary Poisson sighting process. For candidate time `T`, the p-value is `(t[n] / T)^n` (equation 2 of Solow, 1993a), with times measured from the observation origin.

## Usage

```
solow1993(sd, alpha = 0.05, test_year, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). The returned upper bound is the earliest candidate time at which persistence is rejected at this level. |
| `test_year` | Latest time to test. Must be later than the last sighting. A bound beyond `test_year` is reported as `NA` with a warning. |
| `data_out` | If `TRUE`, return the p-value curve at integer steps after the last sighting instead of the upper bound. |

## Details

The earliest supplied time defines the observation origin. If that time has a sighting, one sighting defines the origin and is omitted, as in the Caribbean monk seal example of the source paper; any further sightings in the same time unit are counted. Every sighting event after that is counted, so a time unit with several sightings contributes several events to `n`.

Solow (1993a) defines only the test. Inverting it gives the one-sided upper confidence bound `t[n] / alpha^(1/n)` for the extinction time (Solow, 2005, equation 5), which is returned as `upper`.

## Value

An ede_estimate object with `estimate = NA` and the one-sided upper bound in `upper`, or (if `data_out = TRUE`) a data frame with columns `time` and `p_value`. The column `chance`, the name used up to EDE 0.3.0, is a deprecated alias of `p_value`.

## References

Solow, A. R. (1993a). Inferring extinction from sighting data. Ecology, 74(3), 962-964.

Solow, A. R. (2005). Inferring extinction from a sighting record. Mathematical Biosciences, 195(1), 47-55.

solow1993b *Solow (1993b) declining-population persistence test*

## Description

Conditional test of the null hypothesis that a declining species was still extant at a candidate test year, under a non-stationary Poisson process with an exponentially declining rate. The test conditions on `n` and on the sum `S` of the sighting times, so the rate parameters are not estimated; the p-value is `F(t[n]) / F(T)`, where `F` is Fisher's distribution of the largest gap (equations 3-4 of the source paper).

## Usage

```
solow1993b(sd, alpha = 0.05, test_year, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). The returned upper bound is the earliest candidate time at which persistence is rejected at this level. |
| `test_year` | Latest time to test. Must be later than the last sighting. A bound beyond `test_year` is reported as `NA` with a warning. |
| `data_out` | If `TRUE`, return the p-value curve at integer steps after the last sighting instead of the upper bound. |

## Details

Sighting events are counted as in `solow1993()`: one sighting at the observation origin defines it and is omitted, and every other sighting event is counted, including repeated sightings in one time unit. Include a zero-count first row when observation began before the first sighting, as in the black-footed ferret example of the source paper.

Solow (1993b) defines only the test. Under the assumption that the species is extinct, the minimum variance unbiased estimate of the extinction time under the same truncated exponential model (Beg, 1982) is equation 8 of Solow (2005), also listed in Table 1 of McCrea et al. (2024): `T_E = t[n] + sum_{i=0}^{[s/t[n]]} (-1)^i C(n, i) (s - i t[n])^(n-1) / (n (n-1) sum_{i=0}^{[s/t[n]]-1} (-1`
which equals `t[n] + F(t[n]) / F'(t[n])`. It is returned as `estimate` (in the original time units, i.e. shifted by the origin). Both sums are evaluated as Fisher's largest-gap probabilities through the stable B-spline (Irwin-Hall) recursion rather than the alternating series, which loses all precision when the sighting times are concentrated near `t[n]`. The estimate is `NA` when it is undefined (all sightings other than the last at the origin with `n > 2`, or all at `t[n]`).

Inverting the test numerically gives a one-sided upper confidence bound for the extinction time, returned as `upper` (a package construct; Solow, 2005, Section 3, describes the numerical inversion). Under this model the bound may not exist: the p-value cannot fall below `F(t[n])`, however late the candidate time (Solow, 2005, Section 3).

## Value

An ede_estimate object with the minimum variance unbiased estimate in `estimate` and the one-sided upper bound in `upper` (`NA`, with a warning, if it does not exist before `test_year`), or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (plus `chance`, a deprecated alias of `p_value`).

## References

Solow, A. R. (1993b). Inferring extinction in a declining population. Journal of Mathematical Biology, 32(1), 79-82.

Solow, A. R. (2005). Inferring extinction from a sighting record. Mathematical Biosciences, 195(1), 47-55.

Beg, M. A. (1982). Optimal tests and estimators for truncated exponential families. Metrika, 29, 103-113.

Solow, A. R., & Helser, T. (2000). Detecting extinction in sighting data. In S. Ferson & M. Burgman (Eds.), Quantitative Methods for Conservation Biology (pp. 1-6). Springer, New York.

McCrea, R. S., Cheale, T., Campillo-Funollet, E., & Roberts, D. L. (2024). Inferring species extinction from sighting data. Cambridge Prisms: Extinction, 2, e19.

---

`solow1993_bayes` *Solow (1993a) Bayes factor for persistence*

---

## Description

Bayesian counterpart of `solow1993()` under the same stationary Poisson model. With a uniform prior on the extinction time over the observation period and the noninformative prior `dm / m` on the sighting rate, the Bayes factor in favour of persistence to `T` is `B = (n - 1) / ((T / t[n])^(n - 1) - 1)` (equation 3 of Solow, 1993a), which tends to `1 / log(T / t[n])` when `n = 1`. Given the prior probability `prior` of persistence, the posterior probability is `1 / (1 + (1 - prior) / (prior * B))`.

## Usage

```
solow1993_bayes(sd, test_year, prior = 0.5, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `test_year` | Candidate time `T`. Must be later than the last sighting. |
| `prior` | Prior probability that the species persisted to `test_year`, in (0, 1). |
| `data_out` | If `TRUE`, return the Bayes factor and posterior probability at integer steps after the last sighting up to `test_year`. |

## Details

Times are measured from the observation origin and sighting events are counted as in `solow1993()`.

## Value

A list of class ede_bayes with components bayes_factor, posterior, prior, test_year and method; or (if data_out = TRUE) a data frame with columns time, bayes_factor and posterior.

## References

Solow, A. R. (1993a). Inferring extinction from sighting data. Ecology, 74(3), 962-964.

---

solow1996 *Test and confidence interval for a common end-point of several taxa (Solow 1996)*

---

## Description

Likelihood ratio test of whether several taxa share the same upper end-point, such as the extinction of all of them in one event, and the confidence interval for the common end-point (Solow, 1996). Each taxon's finds are taken to be a stationary Poisson process, so that, given their number, they are independent uniform draws over the unknown range of the taxon. The test improves on the conservative test of Strauss & Sadler (1989) because its significance level is known exactly.

## Usage

```
solow1996(taxa, alpha = 0.05, end = c("upper", "lower"))
```

## Arguments

| | |
|---|---|
| taxa | The taxa, as a data frame with one row per taxon and columns n (number of finds, at least 2), lowest and highest (the locations of the lowest and the highest find), or a list of sighting_data objects, one per taxon, in which each distinct time with a positive count is one find (counts are ignored, as in strauss1989()). |
| alpha | Significance level, in (0, 1), of the confidence interval. |
| end | "upper" (the default) or "lower": which end-point is common. |

## Details

For taxon j with n[j] finds, lowest find L[j], highest find H[j], and Z the highest find of all, let Q[j] = ((H[j] - L[j]) / (Z - L[j]))^(n[j] - 1). The likelihood ratio statistic, conditional on the lowest find of each taxon, is the product of the Q[j], and under the null hypothesis of a common upper end-point -2 log(product) has a chi-square distribution with 2 (k - 1) degrees of freedom (Hogg, 1956), for k taxa; with two taxa the statistic is itself the p-value. The confidence interval for the common end-point U, assuming it holds, is (Z, U_u), where U_u is the value at which -2 log of the product of ((H[j] - L[j]) / (U - L[j]))^(n[j] - 1) equals the upper alpha quantile of the chi-square distribution with 2 k degrees of freedom; it is found numerically. For one taxon it is the interval of strauss1989().

The upper end-point is the one at the largest time. For a common lower end-point (for instance a common time of first appearance) use end = "lower", which, as the paper recommends, multiplies

the locations by minus one. The paper's assumption of independent stationary Poisson processes is crucial and should be checked (see vogel2009()); the interval is conditional on a common end-point and becomes too narrow if one taxon ended earlier than the others.

## Value

An ede_estimate object (class ede_solow1996) with estimate = NA and the confidence interval for the common end-point in lower and upper (for end = "upper", lower is the highest find), plus statistic (-2 log of the likelihood ratio statistic), df, p_value (NA for a single taxon) and taxa (the data used, in the original orientation).

## References

Solow, A. R. (1996). Tests and confidence intervals for a common upper endpoint in fossil taxa. Paleobiology, 22(3), 406-410.

Strauss, D., & Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology, 21(4), 411-427.

## See Also

strauss1989() for one taxon, seymour_ammonites.

## Examples

```
# ten uppermost ammonite taxa of Seymour Island (Solow 1996, Table 1)
solow1996(seymour_ammonites)
# two of them, with a 90% interval
solow1996(seymour_ammonites[c(3, 8), ], alpha = 0.1)
```

---

# solow2005 *Solow (2005) Weibull extreme-value persistence test*

---

## Description

Tests persistence using the Weibull extreme-value model for the k most recent sighting events, the model behind ole(). For a candidate time T, the approximate p-value is equation 16 of Solow (2005), with the shape parameter estimated by equation 15.

## Usage

```
solow2005(sd, alpha = 0.05, test_year, data_out = FALSE, k = NULL)
```

### Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| alpha | Significance level, in (0, 1). The returned upper bound is the earliest candidate time at which persistence is rejected at this level. |
| test_year | Latest time to test. Must be later than the last sighting. A bound beyond test_year is reported as NA with a warning. |
| data_out | If TRUE, return the p-value curve at integer steps after the last sighting instead of the upper bound. |
| k | Number of most recent sighting events to use. Counts greater than one are treated as independent events at the recorded time; Solow (2005) recommends combining sightings that are obviously related. By default, all sighting events are used. Must be at least 3; Solow (2005) suggests at least 5. |

### Details

Inverting the test gives the one-sided upper confidence bound of equation 17, `(t[1] - c t[k]) / (1 - c)` with `c = (-log(alpha) / k)^m`, returned as `upper`. The bound does not exist when `-log(alpha) >= k`. Note that `ole()` reports the two-sided interval of Roberts & Solow (2003), so its upper limit at level `alpha` equals this bound at level `alpha / 2`.

### Value

An ede_estimate object with `estimate = NA` and the one-sided upper bound in `upper`, or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (plus `chance`, a deprecated alias of `p_value`).

### References

Solow, A. R. (2005). Inferring extinction from a sighting record. Mathematical Biosciences, 195(1), 47-55.

---

## solow2006 *Extinction time from dated remains, allowing for dating error (Solow et al. 2006)*

---

### Description

Maximum likelihood estimate of the extinction time of a taxon from the radiometric ages of its most recent remains, with the dating error taken into account (Solow, Roberts & Robbirt, 2006). The true ages of the remains are assumed independent and uniformly distributed between a lower end-point `gamma` and the extinction time `theta`; each recorded age is the true age plus a normal dating error with a known standard deviation. The recorded age then has density `(pnorm((x - gamma) / s) - pnorm((x - theta) / s)) / (theta - gamma)` (equation 2 of the paper, written for a time axis that increases towards the present). Because the dating error is normal and positive the usual likelihood theory applies, which is not the case for the same model without dating error (`strauss1989()`, `robson1964()`).

## Usage

```
solow2006(sd, error = NULL, alpha = 0.05, k = NULL)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object whose positive-count rows are the dated remains, or, without error, a data frame with columns time and error. |
| error | Standard deviation of the dating error, positive, in the unit of the time axis: one value, or one per row of sd in the order of its rows (sd is sorted by time). Omit it when sd is a data frame with an error column. |
| alpha | Significance level for the profile-likelihood interval, in (0, 1). |
| k | Number of most recent ages to use; at least 3. By default all. |

## Details

The uniform distribution is a strong assumption and holds, if at all, only for the most recent part of the record; k selects the number of most recent ages used, and solow2006_check() implements the simulation check of the paper. For two taxa, solow2006_region() gives the joint confidence region of their extinction times (equations 5 to 7 of the paper).

The model for one taxon, the profile log-likelihood and the interval of the extinction time (the set of values at which twice the drop of the profile log-likelihood is at most the upper alpha quantile of the chi-square distribution on one degree of freedom) are those of Solow (1997), who applies them to the settlement time of New Zealand; Solow et al. (2006) use the same likelihood for two taxa and a joint region, see solow2006_region().

The endpoint estimated is the one at the largest time. For an extinction, ages in years before present enter as minus the age, so that the most recent end of the range is the largest time; for a settlement or other first appearance, as in Solow (1997), they enter as the age, so that the earliest end is the largest time. The ages are on the scale of the source, without calibration, which is outside this function.

The dated remains are given in one of two ways. With error, the rows of a sighting_data object with a positive count are the remains (a count above one repeats the age with the same error). Without error, sd is a data frame with one row per remain and columns time and error, which allows ties in the times with different errors, as in radiocarbon data.

## Value

An ede_estimate object (class ede_solow2006) with the maximum likelihood estimate of the extinction time in estimate, the interval in lower and upper (Inf if the upper limit does not exist), and the components gamma (estimated lower end-point), loglik, k, and data (the ages used, most recent first, with their errors).

## References

Solow, A. R. (1997). Estimating settlement time. Radiocarbon, 39(3), 351-354. doi:10.1017/S0033822200053327

Solow, A. R., Roberts, D. L., & Robbirt, K. M. (2006). On the Pleistocene extinctions of Alaskan mammoths and horses. Proceedings of the National Academy of Sciences, 103(19), 7351-7353. doi:10.1073/pnas.0509480103

## See Also

solow2006_region(), solow2006_check(), alaska_megafauna, nz_shells.

## Examples

```
mammoth <- alaska_megafauna[alaska_megafauna$taxon == "mammoth", ]
# extinction: ka before present enters as minus the age
dates <- data.frame(time = -mammoth$age_ka_bp, error = mammoth$sd_ka)
fit <- solow2006(dates)
fit
fit$gamma

# settlement of New Zealand (Solow 1997): the earliest end is the one wanted,
# so the age enters as it is
solow2006(data.frame(time = nz_shells$age_bp, error = nz_shells$sd_yr))
```

---

# solow2006_check *Simulation check of the fitted model of Solow et al. (2006)*

---

## Description

The check of the paper (its Figure 2). The fitted model is used to simulate `nsim` records: `k` true ages are drawn uniformly between the estimated lower end-point and extinction time, ordered, and the `j`-th is given the normal dating error with the standard deviation of the `j`-th recorded age (the ages ordered as recorded); the simulated ages are then ordered again. If the model fits, the ordered recorded ages are close to the expected ordered simulated ages and within their lower and upper `0.025` quantiles.

## Usage

```
solow2006_check(fit, nsim = 1000, level = 0.95)
```

## Arguments

| | |
|---|---|
| `fit` | A result of solow2006(). |
| `nsim` | Number of simulated records. |
| `level` | Coverage of the band; the paper uses the `0.025` and `0.975` quantiles, that is `0.95`. |

## Value

A data frame of class `solow2006_check` with one row per age and columns `observed` (ordered recorded ages, oldest first), `expected` (mean of the ordered simulated ages), `lower` and `upper`. `plot()` draws the observed against the expected ages with the band and the 45 degree line.

## References

Solow, A. R., Roberts, D. L., & Robbirt, K. M. (2006). On the Pleistocene extinctions of Alaskan mammoths and horses. Proceedings of the National Academy of Sciences, 103(19), 7351-7353. doi:10.1073/pnas.0509480103

## See Also

solow2006().

## Examples

```
mam <- alaska_megafauna[alaska_megafauna$taxon == "mammoth", ]
fit <- solow2006(data.frame(time = -mam$age_ka_bp, error = mam$sd_ka))
chk <- solow2006_check(fit, nsim = 200)
head(chk, 3)
```

---

# solow2006_region *Joint confidence region for the extinction times of two taxa*

---

## Description

Approximate joint confidence region for the extinction times of two taxa from their dated remains, by the profile likelihood of the two-taxon model of Solow, Roberts & Robbirt (2006, equations 5 to 7). Each taxon follows the model of solow2006() and the two are independent, so the log-likelihood is the sum and the profile log-likelihood of the pair of extinction times is the sum of the single-taxon profiles. The region is the set of pairs for which twice the drop of the profile log-likelihood from its maximum is at most the upper alpha quantile of the chi-square distribution on two degrees of freedom.

## Usage

```
solow2006_region(
  sd1,
  sd2,
  error1 = NULL,
  error2 = NULL,
  alpha = 0.05,
  k1 = NULL,
  k2 = NULL,
  n = 100L
)
```

## Arguments

| | |
|---|---|
| sd1, sd2 | The dated remains of each taxon, as the sd argument of solow2006(): sighting_data objects, or data frames with columns time and error. |
| error1, error2 | Dating error standard deviations for sd1 and sd2, as the error argument of solow2006(); omit them for data frames. |
| alpha | Significance level, in (0, 1). |
| k1, k2 | Number of most recent ages of each taxon; by default all. |
| n | Number of grid points on each axis. |

## Details

The region answers questions about the relation of the two extinction times, for instance whether it contains pairs in which the second taxon went extinct after the first, or after a given event (the arrival of humans in the paper), through the columns of grid.

## Value

An object of class solow2006_region: a list with fit1 and fit2 (the single-taxon fits, with unadjusted 1-df intervals), grid (a data frame with theta1, theta2, statistic and inside), critical and alpha. Use plot() to draw the region.

## References

Solow, A. R., Roberts, D. L., & Robbirt, K. M. (2006). On the Pleistocene extinctions of Alaskan mammoths and horses. Proceedings of the National Academy of Sciences, 103(19), 7351-7353. doi:10.1073/pnas.0509480103

## See Also

solow2006(), alaska_megafauna.

## Examples

```
mam <- alaska_megafauna[alaska_megafauna$taxon == "mammoth", ]
hor <- alaska_megafauna[alaska_megafauna$taxon == "horse", ]
hor <- hor[-which.max(hor$age_ka_bp), ] # the paper omits the oldest horse age
mk <- function(d) data.frame(time = -d$age_ka_bp, error = d$sd_ka)
reg <- solow2006_region(mk(mam), mk(hor), n = 40)
reg
# pairs in the region with the horse extinct later than 12.0 ka BP
range(reg$grid$theta2[reg$grid$inside])
```

`solow_roberts2003` *Solow & Roberts (2003) nonparametric persistence test*

## Description

Tests the null hypothesis that a species persisted to a candidate time using only the two most recent distinct sighting times. For candidate time `T`, the p-value is `(t[n] - t[n-1]) / (T - t[n-1])` (equation 4 of the source paper). It is exact for a constant sighting rate and asymptotically correct otherwise.

## Usage

```
solow_roberts2003(sd, alpha = 0.05, test_year, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). The returned upper bound is the earliest candidate time at which persistence is rejected at this level. |
| `test_year` | Latest time to test. Must be later than the last sighting. A bound beyond `test_year` is reported as `NA` with a warning. |
| `data_out` | If `TRUE`, return the p-value curve at integer steps after the last sighting instead of the upper bound. |

## Details

The paper derives the test from the Robson & Whitlock (1964) interval; inverting it gives the one-sided upper confidence bound `t[n] + (t[n] - t[n-1]) (1 - alpha) / alpha`, returned as `upper`. It is the same bound as `robson1964()`.

## Value

An ede_estimate object with `estimate = NA` and the one-sided upper bound in `upper`, or (if `data_out = TRUE`) a data frame with columns `time` and `p_value` (plus `chance`, a deprecated alias of `p_value`).

## References

Solow, A. R., & Roberts, D. L. (2003). A nonparametric test for extinction based on a sighting record. Ecology, 84(5), 1329-1332.

## strauss1989 *Strauss & Sadler (1989) confidence interval for the end of a range*

### Description

Unbiased point estimator and classical one-sided confidence interval for the true endpoint of a temporal range, assuming finds are uniformly distributed between the true endpoints. With `n` finds spanning the range `y` to `z`, the estimate is `z + (z - y) / (n - 1)` (equation 8 of the source paper) and the interval is `(z, z + c (z - y))` with `c = alpha^(-1 / (n - 1)) - 1` (equations 13, 16 and 20).

### Usage

```
strauss1989(sd, alpha = 0.05)
```

### Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |
| `alpha` | Significance level, in (0, 1). |

### Details

Each distinct sighting time is one find, as each fossil horizon is in the paper; counts are ignored. The paper's model is continuous, so with dense yearly records the unbiasedness and the coverage hold only approximately. Both endpoints are estimated from the data, so the observation origin (a leading zero-count row) is not used.

### Value

An ede_estimate object containing the unbiased point estimate and a one-sided confidence interval.

### References

Strauss, D., & Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology, 21(4), 411-427.

## strauss1989_bayes *Strauss & Sadler (1989) Bayesian estimate of the end of a range*

## Description

Bayesian counterpart of strauss1989() (Section 4.3 of the source paper). Finds are independent and uniform between the true endpoints theta1 < theta2 (equation 23), and the prior on the endpoints is uniform over a < theta1 < theta2 < b (equation 22, there scaled to (0, 1)), where (a, b) = prior_range is the interval that the endpoints are known to lie in. With n finds spanning y to z, the posterior density of the extinction time theta2 is h(theta2) = (n - 2) [(theta2 - y)^(1-n) - (theta2 - a)^(1-n)] / u_n on z < theta2 < b (equation 26), with u_n = (z-y)^(2-n) - (b-y)^(2-n) - (z-a)^(2-n) + (b-a)^(2-n) (equation 25). The estimate is the posterior mean (equation 27), the Bayes point estimate under quadratic loss. Since the density decreases from its maximum at z, the shortest 1 - alpha posterior probability interval is (z, q), with q the 1 - alpha posterior quantile (as shaded in Fig. 3 of the paper); it is returned in lower and upper.

## Usage

```
strauss1989_bayes(sd, alpha = 0.05, prior_range, data_out = FALSE)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| alpha | Posterior probability outside the interval, in (0, 1). |
| prior_range | Numeric vector c(a, b): the prior is uniform for the endpoints over a < theta1 < theta2 < b. a must be finite and earlier than the first find; b must be later than the last find and may be Inf. |
| data_out | If TRUE, return the posterior on a grid of 200 times from z to the 0.999 posterior quantile instead of the estimate. |

## Details

The paper writes the equations for (a, b) = (0, 1); they are applied here after the corresponding linear change of scale, which leaves the posterior unchanged. The paper suggests (0, Inf) for taxa that may be extant; b = Inf is allowed (the posterior is proper for n >= 3, but its mean is infinite for n = 3). The paper's example takes (0, 1500) m for the ammonite section of Seymour Island.

As in strauss1989(), each distinct sighting time is one find and counts are ignored; the observation origin (a leading zero-count row) is not used, but it may be a natural choice for a. The paper requires n > 2.

## Value

An ede_estimate object with the posterior mean in estimate and the one-sided posterior probability interval (z, q) in lower and upper, plus components density and cdf (functions of time giving the posterior density and distribution function of the extinction time) and prior_range. If data_out = TRUE, a data frame with columns time, density and chance, where chance is the posterior probability that the extinction time is at or before time.

## References

Strauss, D., & Sadler, P. M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology, 21(4), 411-427.

---

`strauss1989_curve` *Confidence curve for the Strauss & Sadler (1989) upper bound*

---

## Description

Evaluates the upper confidence bound of `strauss1989()`, `z + c (z - y)` with `c = alpha^(-1 / (n - 1)) - 1`, for alpha from 0.01 to 1 in steps of 0.01, for plotting instead of a single bound.

## Usage

```
strauss1989_curve(sd)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object. |

## Value

A data frame with columns `time` and `chance`, where `chance` is the confidence level `1 - alpha` of the bound `time`, i.e. the probability that the true endpoint is at or before `time`.

---

`thompson2013` *Thompson et al. (2013) probability of persistence with uncertain sightings*

---

## Description

Discrete-time model of Thompson, Lee, Stone, McCarthy & Burgman (2013) for a record of one or more observation types (e.g. specimens, photographs, sound recordings) seen or not seen in each time unit. The first type is definite (no false sightings); the others may be false. For type `alpha`, `p[alpha]` is the probability of not seeing that type in a time unit when the species is extant and `q[alpha]` the probability of not seeing it when the species is extinct (`q = 1` for the definite type). Time units are independent, so the probability of the record given persistence to the last time unit `T` is the product of the `p_i` (equation 5), and given extinction in time unit `j` it is `p_1 ... p_(j-1) q_j ... q_T` (equation 7), where `p_i` and `q_i` are products over types of `p` or `1 - p` (`q` or `1 - q`) according to whether the type was seen (equation 27). The probability of persistence is equation 9,

## Usage

```
thompson2013(
  definite,
  uncertain = NULL,
  test_year,
  p,
  q = NULL,
  prior_extant = NULL,
  prior_extinct = NULL,
  average = c("annealed", "quenched"),
  nodes = NULL
)
```

## Arguments

| | |
|---|---|
| definite | A sighting_data object for the definite (certain) observation type. |
| uncertain | NULL, a sighting_data object, or a list of them, one per uncertain observation type. |
| test_year | Last time unit T of the record. Must be later than the last definite sighting and not earlier than any other sighting. |
| p | List (or numeric vector, for point values) with one element per observation type, definite first: the probability of not observing that type in a time unit given the species is extant. Each element is a point value in (0, 1) or a range c(a, b) with 0 <= a < b <= 1. |
| q | As p, for the uncertain types only: the probability of not observing that type given the species is extinct. Point values may be 1. |
| prior_extant | Prior probability P(X_T) that the species is extant at test_year. Default T_N / T (equation 19). |
| prior_extinct | Prior probability P_E of extinction in each time unit after the last definite sighting. Default 1 / T (equation 19). |
| average | "annealed" (default) or "quenched"; ignored when all parameters are point values. |
| nodes | Number of Gauss-Legendre nodes per random parameter for the quenched average with more than one random parameter. Default chooses about 2e5 nodes in total. |

## Details

P(X_T | s) = prod(p_i) P(X_T) / (prod(p_i) P(X_T) + sum_j p_1...p_(j-1) q_j...q_T P_E),

with j running over the time units after the last definite sighting T_N, P(X_T) the prior probability of persistence and P_E the prior probability of extinction in each of those time units.

Each p[alpha] and q[alpha] is either a point value or a range c(a, b) read as a uniform distribution, as in the paper. With point values the result is the deterministic estimate P_D (equation 14 when there is only the definite type). With ranges the result is averaged (Section 2.2): "annealed" averages the numerator and the denominator of equation 9 separately (a ratio of averages; equation 12), "quenched" averages the ratio itself (equation 11). Parameters of different types, and p

and q of the same type, are independent. Each parameter is held fixed across time units within a realisation.

The record is binary: a time unit counts as a sighting of a type when its count is positive. Time units are numbered from 1 at the earliest time of any record (so, unlike solow1993(), the first sighting is part of the record), and T = test_year - origin + 1. Times must be whole numbers.

Package constructs: the paper estimates the averages by simulation; here the annealed averages are computed exactly, as products of one-dimensional integrals (stats::integrate(), on the log scale), and the quenched average by stats::integrate() for a single random parameter or by tensor-product Gauss-Legendre quadrature otherwise. The Bayes factor (deterministic and annealed only) is the likelihood under persistence divided by the mean likelihood over the extinction years, which gives the paper's equation 13 in its two-year example; it is NA for the quenched average, which is not a ratio of likelihoods.

## Value

A list of class ede_bayes with components bayes_factor (in favour of persistence, or NA), posterior (probability of persistence from equation 9), prior (prior_extant), test_year, method, plus prior_extinct and average ("deterministic", "annealed" or "quenched").

## References

Thompson, C. J., Lee, T. E., Stone, L., McCarthy, M. A. & Burgman, M. A. (2013). Inferring extinction risks from sighting records. Journal of Theoretical Biology, 338, 16-22.

---

thompson2017 *Thompson et al. (2017) iterative probability that a taxon is extant*

---

## Description

Year-by-year model of the probability P(X[t]) that a taxon is extant, combining records of varying reliability with unsuccessful dedicated and passive surveys (Thompson et al., 2017; the model behind the IUCN Red List extinction-probability tool). Starting from an initial probability p_initial at the beginning of start_year, each year updates the probability by equation 5 of the paper:

## Usage

```
thompson2017(
  records,
  surveys,
  passive,
  p_initial,
  start_year = NULL,
  end_year = NULL,
  n_sim = 10000,
  data_out = FALSE
)
```

## Arguments

| | |
|---|---|
| records | Data frame of record years with columns year and pci_lower, pci_upper (or a single column pci), the probability of correct identification. NULL if there are no records. |
| surveys | Data frame of unsuccessful dedicated survey years with columns year, eps_lower, eps_upper, pi_lower, pi_upper, pr_lower, pr_upper (or single columns eps, pi, pr). NULL if there are no dedicated surveys. |
| passive | Passive-survey inputs applied to every gap year: a named numeric vector or list with elements eps_lower, eps_upper, pi_lower, pi_upper, pr_lower, pr_upper (or eps, pi, pr). Use c(eps = 0, pi = 0, pr = 0) for no passive surveillance. |
| p_initial | Probability that the taxon is extant at the beginning of start_year: one value or a c(lower, upper) range. |
| start_year | First year of the iteration; p_initial applies at its beginning. Defaults to the earliest year in records and surveys. |
| end_year | Last year of the iteration. Defaults to the latest year in records and surveys; years after the last record or survey are gap years. Records or surveys outside start_year:end_year are ignored. Both years are required when there are no records or surveys (e.g. the passive-only projection of section 8.2 and equation 14 of the paper). |
| n_sim | Number of Monte Carlo replicates for the percentile band; 0 skips the simulation. |
| data_out | If TRUE, return the year-by-year trajectory instead of the summary object. |

## Details

- record year: P(X[t+1]) = p(ci) + (1 - p(ci)) * P(X[t]) (equation 1), where p(ci) is the probability that the record is a correct identification of the extant taxon;
- unsuccessful survey year: P(X[t+1]) = (1 - eps * p(r) * p(i)) * P(X[t]) (equations 3-4), where eps is the proportion of the habitat surveyed, p(r) the probability that the taxon would have been recorded and p(i) the probability that it would have been reliably identified;
- gap year (neither a record nor a dedicated survey): the survey rule with the passive-survey inputs eps', p'(r) and p'(i) (section 4.2).

The model is not Bayesian (section 3 of the paper): P(X[t]) is an absolute probability updated by the rules above. Successful surveys are supplied as records.

Every input is given as a lower and upper bound, as in Table 1 of the paper, and uncertainty is propagated in the two ways of section 6:

- p_extant iterates equation 5 with the midpoint of every bound (the solid line of Figs. 1-3);
- lower and upper are the results of interval arithmetic. Equation 5 increases with P(X[t]) and p(ci) and decreases with eps, p(r) and p(i), so the exact bounds are obtained by iterating it with the corresponding extreme inputs (the light shaded region of Figs. 1-3);
- if n_sim > 0, each bound pair is also treated as the 5th and 95th percentiles of an independent normal distribution, inputs are drawn independently for every year (including each gap year)

and every replicate, and the 5th and 95th percentiles of the simulated `P(X[t])` are returned as `mc_lower` and `mc_upper` (the dark shaded region). The paper does not say how draws outside `[0, 1]` are handled; here they are truncated to `[0, 1]` (a package construct).

A point value can be supplied instead of a range by using a single column (e.g. `pci` instead of `pci_lower`/`pci_upper`).

## Value

A list of class `ede_extant` with components `p_extant` (midpoint estimate of the probability that the taxon is extant at the end of `end_year`), `lower`, `upper` (interval-arithmetic bounds), `mc_lower`, `mc_upper` (Monte Carlo 5th and 95th percentiles, or `NA` when `n_sim = 0`), `p_extinct` (`1 - p_extant`), `end_year` and `method`; or (if `data_out = TRUE`) a data frame with one row per year and columns `time`, `type` (`"record"`, `"survey"` or `"passive"`), `p_extant`, `lower`, `upper`, `mc_lower` and `mc_upper`, where each row gives the probability that the taxon is extant at the end of year `time`, after that year's data (as plotted in Figs. 1-3).

## References

Thompson, C. J., Koshkina, V., Burgman, M. A., Butchart, S. H. M., & Stone, L. (2017). Inferring extinctions II: a practical, iterative model based on records and surveys. Biological Conservation, 214, 328-335.

## Examples

```
# Alaotra Grebe, Table 1 of Thompson et al. (2017)
rec <- data.frame(
  year = c(1929, 1960, 1963, 1969, 1970, 1971, 1972, 1982, 1985, 1986, 1988),
  pci_lower = c(0.99, 0.95, 0.75, 0.75, 0.10, 0.10, 0.60, 0.60, 0.20, 0.20, 0.20),
  pci_upper = c(1.00, 0.99, 0.94, 0.94, 0.40, 0.40, 0.80, 0.80, 0.80, 0.70, 0.50)
)
sv_years <- c(1989, 1990, 1993, 1994, 1997, 1998, 1999, 2000, 2004, 2009)
sv <- data.frame(
  year = sv_years,
  eps_lower = ifelse(sv_years == 2000, 0.70, 0.80),
  eps_upper = ifelse(sv_years == 2000, 0.90, 0.95),
  pi_lower = 0.90, pi_upper = 0.95, pr_lower = 0.70, pr_upper = 0.90
)
pas <- c(eps_lower = 0, eps_upper = 0.05, pi_lower = 0.10, pi_upper = 0.65,
         pr_lower = 0.40, pr_upper = 0.60)
thompson2017(rec, sv, pas, p_initial = c(0.95, 0.99), end_year = 2017,
             n_sim = 0)
```

---

`thompson2019` *Thompson et al. (2019) Bayesian updating of the probability of persistence*

---

## Description

Bayesian counterpart of `thompson2017()`. Each year's posterior is the next year's prior (Thompson et al., 2019, equations 1-6), so the probability that the taxon is extant after year `t` is `P(X[t+1]) = 1 / (1 + B[t] * (1 / P(X[1]) - 1))` (equation 7), where `P(X[1])` is the initial probability at the beginning of `start_year` and `B[t]` is the cumulative Bayes factor for extinction, the product of the yearly factors `b[j]` (equation 8). The yearly factors (equations 9-11) are

## Usage

```
thompson2019(
  records,
  surveys,
  passive,
  p_initial = 0.5,
  start_year = NULL,
  end_year = NULL,
  data_out = FALSE
)
```

## Arguments

| | |
|---|---|
| `records` | Data frame of record years with columns `year` and `pci_lower`, `pci_upper` (or a single column `pci`), the probability of correct identification. `NULL` if there are no records. |
| `surveys` | Data frame of unsuccessful dedicated survey years with columns `year`, `eps_lower`, `eps_upper`, `pi_lower`, `pi_upper`, `pr_lower`, `pr_upper` (or single columns `eps`, `pi`, `pr`). `NULL` if there are no dedicated surveys. |
| `passive` | Passive-survey inputs applied to every gap year: a named numeric vector or list with elements `eps_lower`, `eps_upper`, `pi_lower`, `pi_upper`, `pr_lower`, `pr_upper` (or `eps`, `pi`, `pr`). Use `c(eps = 0, pi = 0, pr = 0)` for no passive surveillance. |
| `p_initial` | Probability `P(X[1])` that the taxon is extant at the beginning of `start_year`: one value or a `c(lower, upper)` range, whose midpoint is used. |
| `start_year` | First year of the iteration; `p_initial` applies at its beginning. Defaults to the earliest year in `records` and `surveys`. |
| `end_year` | Last year of the iteration. Defaults to the latest year in `records` and `surveys`; years after the last record or survey are gap years. Records or surveys outside `start_year:end_year` are ignored. Both years are required when there are no records or surveys (e.g. the passive-only projection of section 8.2 and equation 14 of the paper). |
| `data_out` | If `TRUE`, return the year-by-year Bayes factors and probabilities instead of the summary object. |

## Details

- record year: `b = 1 - p(ci)`;
- unsuccessful dedicated survey year: `b = 1 / (1 - eps * p(r) * p(i))`;

- gap year (passive survey): the survey factor with `eps'`, `p'(r)` and `p'(i)`.

Inputs use the same format as `thompson2017()`. Following section 2.1 and the case study of the paper, ranges are replaced by their midpoints (including a range for `p_initial`); the paper does not propagate the ranges any further.

The result is an `ede_bayes` object, as returned by `solow1993_bayes()`. Its `bayes_factor` is in favour of persistence, `1 / B[t]`, so that it has the same orientation as in `solow1993_bayes()`; the paper's cumulative Bayes factor for extinction is returned as `bayes_factor_extinction`.

## Value

A list of class `ede_bayes` with components `bayes_factor` (cumulative Bayes factor in favour of persistence, `1 / B`), `bayes_factor_extinction` (`B`, equation 8), `posterior` (probability that the taxon is extant at the end of `end_year`, equation 7), `prior` (`P(X[1])`), `test_year` (`end_year`) and `method`; or (if `data_out = TRUE`) a data frame with one row per year and columns `time`, `type` (`"record"`, `"survey"` or `"passive"`), `b` (yearly Bayes factor for extinction, equation 11), `bayes_factor_extinction` (cumulative `B[t]`) and `posterior` (probability extant at the end of year `time`).

## References

Thompson, C. J., Kodikara, S., Burgman, M. A., Demirhan, H., & Stone, L. (2019). Bayesian updating to estimate extinction from sequential observation data. Biological Conservation, 229, 26-29.

## Examples

```
# Alaotra Grebe from 1990 (Thompson et al., 2019, Table 1)
sv <- data.frame(year = c(1990, 1993, 1994, 1997, 1998, 1999),
                 eps_lower = 0.80, eps_upper = 0.95, pi_lower = 0.90,
                 pi_upper = 0.95, pr_lower = 0.70, pr_upper = 0.90)
pas <- c(eps_lower = 0, eps_upper = 0.05, pi_lower = 0.10, pi_upper = 0.65,
         pr_lower = 0.40, pr_upper = 0.60)
thompson2019(NULL, sv, pas, p_initial = 0.5, end_year = 1994)
```

---

# vogel2009 *Probability plot correlation test of constant sighting rate*

---

## Description

Tests whether the sighting times of a record are compatible with a constant sighting rate before extinction, that is, with times uniformly distributed on an interval ending at the extinction time (the assumption of `solow1993()` and `solow2005()`), with the probability plot correlation coefficient (PPCC) test of Vogel et al. (2009), who apply it to the sighting records of North American and Hawaiian birds.

## Usage

```
vogel2009(sd, events = FALSE, nsim = 1e+05)
```

## Arguments

| | |
|---|---|
| sd | A sighting_data object. |
| events | If FALSE (default), count each distinct sighting time once; if TRUE, count every sighting event. |
| nsim | Number of draws used for the null distribution. |

## Details

With the n sighting times in increasing order, t[1] <= ... <= t[n], the statistic is the product-moment correlation r between t[i] and the plotting positions i / (n + 1) (equation 14 of the paper). Low values of r are evidence against the uniform model. r is invariant to the origin and the unit of time, and to the unknown end of the interval, so its null distribution depends only on n. The p-value is the probability, under the uniform model, of a correlation at most r; it is computed by simulation (nsim draws, with a fixed internal seed, so the result is reproducible and the global random number stream is left untouched) in place of the tables of Vogel & Kroll (1989).

As Vogel et al. stress, a single short record gives the test little power, so failing to reject is not evidence for the uniform model. For a set of records, ede_field_test() combines the p-values.

By default each distinct sighting time counts once, as in hall_wang1999(); set events = TRUE to count every sighting event, in which case ties are present and the model is only approximate.

## Value

An object of class htest, with statistic (r), p.value, parameter (the number of sighting times n) and alternative ("less": the correlation is smaller than under the uniform model).

## References

Vogel, R. M., Hosking, J. R. M., Elphick, C. S., Roberts, D. L., & Reed, J. M. (2009). Goodness of fit of probability distributions for sightings as species approach extinction. Bulletin of Mathematical Biology, 71(3), 701-719. doi:10.1007/s11538-008-9377-3

Vogel, R. M., & Kroll, C. N. (1989). Low-flow frequency analysis using probability-plot correlation coefficients. Journal of Water Resources Planning and Management, 115(3), 338-357.

## See Also

ede_field_test() to combine tests over several records; solow1993() and solow2005(), which assume a constant rate.

## Examples

```
# Eskimo curlew, the example of Vogel et al. (2009, Figure 2)
yrs <- c(1946, 1947, 1948, 1950, 1955, 1956, 1959, 1960, 1961, 1962, 1963,
         1964, 1968, 1970, 1972, 1973, 1974, 1976, 1977, 1980, 1981, 1982,
         1983, 1985)
```

```
vogel2009(sighting_data(data.frame(year = yrs, sightings = 1)))
```

---

`weiss_marshall1999` *Weiss & Marshall (1999) Bayesian end point for discretely sampled sections*

---

## Description

Posterior distribution of the true end point `S` of a stratigraphic range (or of a sighting record) when the section is sampled at discrete points `t[1] < ... < t[J]`, each recorded as presence or absence. While the taxon is extant (`t[j] <= S`) it is found in a sample with probability `p`; after `S` it is never found (equations 1-2 of Weiss & Marshall, 1999). With `a[m]` presences up to the last presence `t[m]` and `b(S)` absences up to `S`, the likelihood is `p^a[m] (1 - p)^b(S)` for `S >= t[m]` (equation 3). The prior on `S` is exponential with mean longevity `l` (equation 4) and the prior on `p` is `Beta(c, d)`. Integrating `p` out gives the step-function posterior of equation 6: on `[t[j], t[j+1])` (with `t[J+1] = Inf`) its mass is `g[j]` of equation 8, `B(a[m] + c, b[j] + d) (exp(-t[j] / l) - exp(-t[j+1] / l))`, with `b[j] = j - a[m]`. Quantiles, moments and the posterior of `p` (a mixture of `Beta(a[m] + c, b[j] + d)` densities, equations 7-8) follow in closed form (Appendix).

## Usage

```
weiss_marshall1999(sd, prior_p, mean_longevity, alpha = 0.05, test_year = NULL)
```

## Arguments

| | |
|---|---|
| `sd` | A sighting_data object, one row per sampled horizon (count 0 = absence). |
| `prior_p` | Numeric vector `c(c, d)` of the `Beta(c, d)` prior on the detection probability `p`; both must be positive. |
| `mean_longevity` | Mean `l` of the exponential prior on the time to the end point (equation 4), in the units of `time`; positive and finite. The paper uses 800 m (a 21 Myr mean foraminiferal longevity converted to section thickness). |
| `alpha` | Posterior tail probability, in (0, 1), for `upper`. |
| `test_year` | Optional time after the last presence; if given, the posterior probability `P(S > test_year \| Y)` that the taxon was still extant at that time is returned as `persistence_prob`. |

## Details

Each row of `sd` is one sampled horizon: a positive count is a presence and a zero count an absence. Every sampled horizon must be supplied, including the barren samples above the last presence, which carry the information about the end point; unsampled positions must not be given as rows. The earliest row is `t[1]`, where the paper assumes the taxon is already extant (it need not be found there). Counts above one are treated as presence; the model uses presence/absence only.

Times may be stratigraphic heights or ages, but the end point estimated is always the one at the largest `time`. To estimate a last appearance from ages in Ma (which decrease upwards), code `time`

= -age; to estimate a first appearance from heights, code time = -height. mean_longevity and the results are on the coded scale.

The reported estimate is the posterior median and upper is the 1 - alpha posterior quantile, so [t[m], upper] is the minimum-length 1 - alpha posterior interval the paper reports (S.95). The paper tabulates these as distances beyond the last presence (residual lifetime); here they are on the original time scale. Choosing c and d is part of the analysis: the paper pools the records of similar taxa, see weiss_marshall1999_prior().

## Value

An ede_estimate object with estimate (posterior median of S), lower (last presence t[m]), upper (1 - alpha posterior quantile) and interval_type = "one-sided", plus components posterior_mean and posterior_sd of S, p_mean and p_sd (posterior mean and SD of p), prior_p, mean_longevity and, if requested, persistence_prob.

## References

Weiss, R. E., & Marshall, C. R. (1999). The uncertainty in the true end point of a fossil's stratigraphic range when stratigraphic sections are sampled discretely. Mathematical Geology, 31(4), 435-453.

Marshall, C. R. (2010). Using confidence intervals to quantify the uncertainty in the end-points of stratigraphic ranges. In J. Alroy & G. Hunt (Eds.), Quantitative Methods in Paleobiology (pp. 291-316). The Paleontological Society Papers, 16.

---

# weiss_marshall1999_prior

*Beta prior for p pooled from related records (Weiss & Marshall, 1999)*

---

## Description

Builds the Beta(c, d) prior on the detection probability used by weiss_marshall1999() the way Weiss & Marshall (1999, p. 442-443) do for their Eastbourne foraminifera: starting from Beta(0, 0), it adds the presences (c) and absences (d) of other, comparable taxa from the first sample up to, but excluding, each taxon's last presence (the last presence is excluded because it is a presence by definition).

## Usage

```
weiss_marshall1999_prior(records)
```

## Arguments

| | |
|---|---|
| records | A list of sighting_data objects, one per comparable taxon, each with one row per sampled horizon (count 0 = absence). |

## Value

A named numeric vector c(c = , d = ) to pass as prior_p.

## References

Weiss, R. E., & Marshall, C. R. (1999). The uncertainty in the true end point of a fossil's stratigraphic range when stratigraphic sections are sampled discretely. Mathematical Geology, 31(4), 435-453.

# Index


∗ **datasets**
  alaska_megafauna, 3
  dodo, 17
  extinct_birds, 26
  ferret, 27
  monk_seal, 46
  nz_shells, 46
  orchid, 48
  seymour_ammonites, 53

alaska_megafauna, 3, *8*, *63*, *65*
alroy2014, 4
as.data.frame.ede_estimate, 6

base::as.data.frame(), *6*
bradshaw2012, 7
brook2019, 9
burgman1995, 11
burgman1995(), *21*

caley_barry2014, 13

dekkers1989, 15
dekkers1989(), *22*
dodo, 17

ede_compare, 18
ede_compare(), *48, 49*
ede_endpoint_test, 19
ede_endpoint_test(), *17*, *55*
ede_estimate, *5, 6*, *8–10*, *12*, *16*, 21, *28*, *30–32*, *40*, *43–45*, *47*, *49*, *51*, *55, 56*, *58*, *60–62*, *66–68*, *78*
ede_evi, 22
ede_evi(), *17*
ede_field_test, 23
ede_field_test(), *76*
ede_select_k, 24
ede_select_k(), *19*, *30*
extinct_birds, 26

ferret, 27
fraga_alves_neves2014, 27
fraga_alves_neves2014(), *20, 21*, *30*

graphics::plot(), *49, 50*

hall_wang1999, 29
hall_wang1999(), *15–17*, *19–21*, *24, 25*, *27, 28*, *55*, *76*

jaric2010, 31
jaric2010(), *21*
jaric_roberts2014, 32

kodikara2020, 33
kodikara2021, 35

lee2014, 37

marshall1994, 39
marshall1994(), *41*
marshall1994_trend, 41
marshall1994_trend(), *40*
marshall1997, 42
marshall1997(), *44*
mccarthy1998, 43
mcinerny2006, 45
mcinerny2006(), *7, 8*, *21*
monk_seal, 46

nz_shells, 46, *63*

ole, 47
ole(), *9*, *15*, *17*, *19*, *21*, *24, 25*, *27–30*, *55*, *60, 61*
orchid, 48

plot.ede_compare, 48
plot.ede_estimate, 49
plot.sighting_data, 50

robson1964, 50

robson1964(), *9*, *28*, *30*, *61*, *66*
rout2009, 51
rout2009(), *52*
rout2009_absences, 52
rout2009_absences(), *52*

set.seed(), *13*, *25*
seymour_ammonites, 53, *60*
sighting_data, *5*, *7*, *9*, *11*, *13*, *16*, *18, 19*, *22*, *24*, *28, 29*, *31–33*, *35*, *38*, *40–42*, *44, 45*, *47*, *50–52*, 54, *55–59*, *61, 62*, *65–70*, *76–78*
smith1985, 54
solow1993, 56
solow1993(), *9*, *14*, *21*, *43, 44*, *52*, *57, 58*, *71*, *75, 76*
solow1993_bayes, 58
solow1993_bayes(), *51*, *75*
solow1993b, 57
solow1993b(), *21*
solow1996, 59
solow1996(), *53*
solow2005, 60
solow2005(), *21*, *47*, *75, 76*
solow2006, 61
solow2006(), *3*, *8*, *46*, *63–65*
solow2006_check, 63
solow2006_check(), *62, 63*
solow2006_region, 64
solow2006_region(), *3*, *62, 63*
solow_roberts2003, 66
solow_roberts2003(), *21*
stats::integrate(), *38*, *43*, *71*
strauss1989, 67
strauss1989(), *9*, *40*, *42*, *59–61*, *68, 69*
strauss1989_bayes, 67
strauss1989_bayes(), *6*, *21*
strauss1989_curve, 69

thompson2013, 69
thompson2017, 71
thompson2017(), *74, 75*
thompson2019, 73

vogel2009, 75
vogel2009(), *23, 24*, *60*

weiss_marshall1999, 77
weiss_marshall1999(), *6*, *21*, *40*, *78*
weiss_marshall1999_prior, 78
weiss_marshall1999_prior(), *78*