

Package ‘Sequential’

August 2, 2017

Type Package

Title Exact Sequential Analysis for Poisson and Binomial Data

Version 2.3.2

Date 2017-08-02

Author Ivair Ramos Silva and Martin Kulldorff

Maintainer Ivair Ramos Silva <jamesivair@yahoo.com.br>

Description Functions to calculate exact critical values, statistical power, expected time to signal and required sample sizes for performing exact sequential analysis. All these calculations can be done for either Poisson or binomial data, for continuous or group sequential analyses, and for different types of rejection boundaries. In case of group sequential analyses, the group sizes do not have to be specified in advance and the alpha spending can be arbitrarily settled.

License GPL-2

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2017-08-02 05:14:46 UTC

R topics documented:

Sequential-package	2
Analyze.Binomial	11
Analyze.CondPoisson	16
Analyze.Poisson	21
AnalyzeSetUp.Binomial	25
AnalyzeSetUp.CondPoisson	28
AnalyzeSetUp.Poisson	31
CV.Binomial	34
CV.CondPoisson	36
CV.G.Binomial	39
CV.G.Poisson	41
CV.Poisson	43

Performance.Binomial	45
Performance.CondPoisson	48
Performance.G.Binomial	50
Performance.G.Poisson	53
Performance.Poisson	55
SampleSize.Binomial	57
SampleSize.CondPoisson	59
SampleSize.Poisson	62

Index	65
--------------	-----------

Sequential-package	<i>Analysis Support, Critical Values, Power, Time to Signal and Sample Size for Sequential Analysis with Poisson and Binomial Data.</i>
--------------------	---

Description

Sequential is designed for continuous and group sequential analysis, where statistical hypothesis testing is conducted repeatedly on accumulating data that gradually increases the sample size. This is different from standard statistical analysis, where a single analysis is performed using a fixed sample size. It is possible to analyze either Poisson type data or binomial 0/1 type data. For binomial data, it is possible to incorporate an off-set term to account for variable matching ratios. For Poisson data, the critical value is based on a Wald-type upper boundary, which is flat on the scale of the log-likelihood ratio, and on a predetermined maximum sample size. For data distributions, it is also possible to apply a user defined alpha spending function. For group sequential analyses, there are functions for pre-specified group sizes and for the situation when the group sizes are not known a priori. It is also possible to perform mixed continuous/group sequential analysis, where, for example, there is at first a big batch of data that arrives in one group, followed by continuous sequential analysis. All results are exact, based on iterative numerical calculations, rather than asymptotic theory or computer simulations.

In the package, there are functions to calculate critical values, statistical power, expected time to signal when the null hypothesis is rejected, and expected sample size at the end of the sequential analyses whether the null hypothesis was rejected or not. For example, for any desired power, relative risk and alpha level, the package can calculate the required upper limit on the sample size, the critical value needed, and the corresponding expected time to signal when the null hypothesis is rejected.

Details

Package:	Sequential
Type:	Package
Version:	2.3.2
Date:	2017-08-02
License:	GPL 2
LazyLoad:	yes

Index:

<code>Analyze.Binomial</code>	Function to Conduct Group Sequential Analyses for Binomial Data When the Group Sizes are not Known a Priori.
<code>AnalyzeSetUp.Binomial</code>	Function to Set Up the Input Parameters Before Using the <code>Analyze.Binomial</code> Function for the First Time.
<code>Analyze.Poisson</code>	Function to Conduct Group Sequential Analyses for Poisson Data When the Group Sizes are not Known a Priori.
<code>AnalyzeSetUp.Poisson</code>	Function to Set Up the Input Parameters Before Using the <code>Analyze.Poisson</code> Function for the First Time.
<code>Analyze.CondPoisson</code>	Function to Conduct Group Sequential Analyses for Conditional Poisson Data When the Group Sizes are not Known a Priori.
<code>AnalyzeSetUp.CondPoisson</code>	Function to Set Up the Input Parameters Before Using the <code>Analyze.CondPoisson</code> Function for the First Time.
<code>CV.Binomial</code>	Critical Values for Continuous Sequential Analysis with Binomial Data.
<code>CV.G.Binomial</code>	Critical Values for Group Sequential Analysis with Binomial Data.
<code>CV.G.Poisson</code>	Critical Values for Group Sequential Analysis with Poisson Data.
<code>CV.Poisson</code>	Critical Values for Continuous Sequential Analysis with Poisson Data.
<code>CV.CondPoisson</code>	Critical Values for continuous sequential CMaxSPRT for Poisson data with limited information from historical cohort.
<code>Performance.Binomial</code>	Power, Expected Signal Time and Sample Size for Continuous Sequential Analysis with Binomial Data.
<code>Performance.G.Binomial</code>	Power, Expected Signal Time and Sample Size for Group Sequential Analysis with Binomial Data.
<code>Performance.G.Poisson</code>	Power, Expected Signal Time and Sample Size for Group Sequential Analysis with Poisson Data.
<code>Performance.Poisson</code>	Power, Expected Signal Time and Sample Size for Continuous Sequential Analysis from Limited Historical Cohort Poisson Data.
<code>Performance.CondPoisson</code>	Power, Expected Signal Time and Sample Size for Continuous Sequential CMaxSPRT with Poisson Data.
<code>SampleSize.Binomial</code>	Sample Size Calculation for Continuous Sequential Analysis with Binomial Data.
<code>SampleSize.Poisson</code>	Sample Size Calculation for Continuous Sequential Testing with Poisson Data.
<code>SampleSize.CondPoisson</code>	Sample Size Calculation for Continuous Sequential CMaxSPRT with Poisson Data.

Overview

Most of the sequential analysis methods found in the literature are based on asymptotic results. In contrast, this package contains functions for the exact calculation of critical values, statistical power, expected time to signal when the null is rejected and the maximum sample size needed when the null is not rejected. This is done for Poisson and binomial type data with a Wald-type upper boundary, which is flat with respect to the likelihood ratio function, and a predetermined upper limit on the

sample size. For a desired statistical power, it is also possible to calculate the latter. The motivation for this package is post-market near real-time drug and vaccine safety surveillance, where the goal is to detect rare but serious safety problems as early as possible, in many cases after only a handful of adverse events. The package can also be used in other application areas, such as clinical trials.

The basis for this package is the Maximized Sequential Probability Ratio Test (MaxSPRT) statistic (Kulldorff et al., 2011), which is a variant of Wald's Sequential Probability Ratio Test (SPRT) (Wald, 1945,47). MaxSPRT uses a composite alternative hypothesis, and upper boundary to reject the null hypothesis when there are more events than expected, no lower boundary, and an upper limit on the sample size at which time the sequential analyses end without rejecting the null. MaxSPRT was developed for post-market vaccine safety surveillance as part of the Vaccine Safety Datalink project run by the Centers for Disease Control and Prevention.

In this package, all critical values, alpha spending strategies, statistical power, expected time to signal and required sample size to achieve a certain power, are obtained exactly to whatever decimal precision desired, using iterative numerical calculations. None of the results are based on asymptotic theory or computer simulations.

Poisson Data

To start, consider continuous sequential analysis for Poisson data. Let C_t be the random variable that counts the number of events up to time t . Suppose that, under the null hypothesis, C_t has a Poisson distribution with mean μ_t , where μ_t is a known function reflecting the population at risk. Under the alternative hypothesis, suppose that C_t has a Poisson distribution with mean $RR\mu_t$, where "RR" is the unknown increased relative risk due to the vaccine. The MaxSPRT statistic defined in terms of the log likelihood ratio is given by:

$$LLR_t = (\mu_t - c_t) + c_t \log c_t / \mu_t,$$

when c_t is at least μ_t , and $LLR_t = 0$, otherwise. For continuous sequential analysis, the test statistic, LLR_t , is monitored at all times $t \in (0, T]$, where $T = \text{SampleSize}$. `SampleSize` is defined a priori by the user in order to achieve the desired statistical power, which can be calculated using the `SampleSize.Poisson` function. The sequential analysis ends, and H_0 is rejected if, and when, $LLR_t \geq CV$, where CV is calculated using the `CV.Poisson` function. If $\mu_t = \text{SampleSize}$, the sequential analysis ends without rejecting the null hypothesis. To calculate other important performance metrics, such as the expected time to signal when the null hypothesis is rejected, use the `Performance.Poisson` function.

If the first event occurs sufficiently early, the sequential analysis may end with the null hypothesis rejected after a single event. There is an option to require a minimum number of observed events, $c_t = M$, before the null can be rejected. Setting M in the range $[3,6]$ is often a good choice (Kulldorff and Silva, 2012). If there is a delay until the sequential analysis starts, but it continues continuously thereafter, there is an option for that as well, requiring a minimum number $\mu_t = D$ of expected events before the null can be rejected.

With continuous sequential analysis, investigators can repeatedly analyze the data as often as they want, ensuring that the overall probability of falsely rejecting the null hypothesis at any time during the analysis is controlled at the desired nominal significance level (Wald, 1945, 1947). Continuous sequential methods are suitable for real-time or near real-time monitoring. When data is only analyzed intermittently, group sequential methods are used instead (Chin, 2012; Cook and DeMets, 2007; Xia, 2007; Friedman et al., 2010; Ghosh and Sen, 1991; Jennison and Turnbull, 2000; Mukhopadhyay and Silva, 2002; Whitehead, 1997). The data is then analyzed at regular or

irregular discrete time intervals after a certain amount of data is accessible. Group sequential statistical methods are commonly used in clinical trials, where a trial may be stopped early due to either efficacy or unexpected adverse events (Jennison and Turnbull, 2000).

The same test statistic, LLR_t , is used for group sequential analyses (Silva and Kulldorff, 2012). The times when LLR_t is evaluated can be defined in several ways, using regular or irregular time intervals that are referenced by calendar period, sample size or some scale involving the distribution of the data. For Poisson data, the group sequential analysis must be conducted with equal size groups, with a constant expected number of adverse events between looks at the accumulating data. In another words, LLR_t is compared against CV whenever μ_t is a multiple of $\text{SampleSize}/\text{Looks}$, where 'Looks' is the total number of looks at the data. To do group sequential analysis for Poisson data, use the `CV.G.Poisson` and `Performance.G.Poisson` functions.

Binomial Data

The MaxSPRT method can also be applied to binomial/Bernoulli data. Let n denote the total number of events that has been observed in a sequential monitoring up to a certain moment in time. Suppose that these n events are categorized as cases and controls. For example, cases may be adverse events happening to a person taking drug A, while controls may be the same adverse event happening to someone in a matched set of individuals taking drug B. As another example, in a self-control sequential analysis, cases may be adverse events happening during the 1-28 days following vaccination, while controls are the same adverse events happening 29-56 days after vaccination.

Let C_t to denote the number of cases among the n events, and assume that C_t follows a binomial distribution with success probability equal to p , where $p = 1/(1+z)$, and z is the matching ratio between the occurrence of a case and of a control under the null hypothesis. For example, if the probability of having a case (instead of a control) is $p = 1/(1+z) = 0.5$, then $z=1$ (1:1 matching ratio), or, $p = 0.25$ for $z=3$ (1:3 matching ratio), etc.

The MaxSPRT statistic (Kulldorff et al., 2011) for a continuous binomial surveillance is:

$$LR_n = \frac{(c_n/n)^{c_n} [(n - c_n)/n]^{n-c_n}}{[1/(1+z)]^{c_n} [z/(1+z)]^{n-c_n}},$$

if $zc_n/(n - c_n) > 1$, and $LR_n = 1$ otherwise.

The monitoring is continued until either there is a signal rejecting the null hypothesis ($LR_n > CV$) or until $n = N$, which indicates that the null is not to be rejected. To perform the calculations, use the `CV.Binomial`, `SampleSize.Binomial` and `Performance.Binomial` functions.

To calculate the critical value for a Wald type rejection boundary, and when the group sizes are fixed a priori, use the `CV.G.Binomial` function. For statistical power, expected time to signal, expected time of surveillance, use the `Performance.G.Binomial` function.

The main assumptions behind the method above are: (i) the monitoring is truly performed in a continuous fashion; (ii) the matching ratio (z) is constant for all of the n events, and (iii) it uses a Wald type rejection boundary that is flat in terms of the likelihood function. Relaxing these assumptions, Fireman et al. (2013) developed exact sequential analysis for group sequential data with varying matching ratios, and for any user specified alpha rejection plan.

Conditional Poisson data with limited information from historical cohort - CMaxSPRT

In Poisson MaxSPRT, the expected mean μ_t is assumed to be a known function reflecting the baseline adverse event risk in the absence of the exposure of interest. In practice, it is estimated with historical data and the uncertainty associated with the estimated counts may or may not have a non-negligible impact on the performance of the sequential analysis method. Li and Kulldorff (Li and

Kulldorff, 2010) showed in their simulation study that uncertainty in the estimated baseline means can be ignored when the total number of events in the historical data is at least 5 times the specified upper limit T . Otherwise, it is recommended to implement a statistical procedure that takes in account the variation from the historical data, i.e. a procedure that conditions the likelihood function of the historical Poisson data, here simply called "conditional Poisson data". For this, the Conditional Maximized Sequential Probability Ratio Test (CMaxSPRT) to account for variation in both the historical and surveillance cohorts.

Let c and V denote the total number of events and the cumulative person-time in the historical data, let P_k denote the cumulative person-time observed in the surveillance population when the k th event occurred. The CMaxSPRT statistic defined in terms of the log likelihood ratio is given by

$$U_k = c \log\left(\frac{c(1 + P_k/V)}{c + k}\right) + k \log\left(\frac{k(1 + P_k/V)}{P_k/V(c + k)}\right),$$

when $k/c > P_k/V$, and $U_k = 0$, otherwise. In the original publication (Li and Kulldorff, 2010), the method was introduced as a continuous sequential analytic approach with the upper limit defined in terms of the maximum number of observed events, i.e., $k \leq K$, and the critical value calculated via a Monte Carlo approach. A large number of Monte Carlo simulations (e.g., 10 million) might be needed to calculate the critical values with a reasonable precision.

In Silva et al. (2016), the method was extended i) with another option of defining the surveillance length in terms of the maximum cumulative person-time divided by the total cumulative person-time in the historical cohort, i.e., $P_k/V \leq T$, ii) with an exact calculation of the critical values for both surveillance length definitions, and iii) for group sequential analysis with data updated and analyzed intermittently instead of continuously. The exact critical values are calculated using the interval halving method to solve for the root of a complex, non-linear equation such that the overall Type I error rate is preserved at the nominal level. As K increases, the computing time for the exact critical values increases exponentially.

Silva et al. (2016) also proposed two approximation methods to calculate the critical values that require substantially less computing time. One approach may overestimate the critical values and thus is referred to as the conservative approach as it may yield lower-than-nominal Type I error rates; the other approach may underestimate the critical values and thus is referred to as the liberal approach as it may yield higher-than-nominal Type I error rates. The recommendation is to use the exact approach when K is small (e.g., 10), use the conservative approach when K is medium or large but c is small, and use the liberal approach when c is medium (e.g., 50) or large. Exact calculations for selected tuning parameters show that the three approaches yield very similar results when K and c are reasonably large.

For calculating critical values for a Wald type rejection boundary, use the `CV.CondPoisson` function. For statistical power, expected time to signal, expected time of surveillance, and maximum sample size requirements, use the `Performance.CondPoisson` and `SampleSize.CondPoisson` functions.

Alpha spending function for unpredictable group sizes

The alpha spending function specifies the cumulative amount, $F_\alpha(t)$, of Type I error probability related to each of the possible values of n . Thus, at the end of the monitoring the alpha spending corresponds to a value smaller than or equal to the overall amount of Type I error probability defined for the overall nominal significance level, α .

Denote the single probability of rejecting the null hypothesis at the j -th test by α_j . Then, the alpha spending at test i is given by $F_\alpha(t_i) = \sum_{j=1}^i \alpha_j \leq \alpha$.

There is a vast number of proposals for choosing the shape of the alpha spending function. Jennison and Turnbull (2000) present a rich discussion about this topic. They dedicated a special attention to the alpha spending of the form: $F_\alpha(t) = \alpha t^\rho$, where $\rho > 1$, and t represents a fraction of the maximum length of surveillance.

To run continuous or group sequential analysis with a user defined alpha spending function, and/or, when the group sizes are not known a priori, `Analyze.Binomial`, `Analyze.Poisson`, and `Analyze.CondPoisson` should be used for binomial and Poisson, and conditional Poisson data, respectively. These functions work differently than the other functions mentioned above. Those other functions are designed to be used before the start of the sequential analysis, in order to determine what the maximum sample size and critical value should be. Once the sequential analysis is under way, the test statistic is then calculated using a hand calculator or an excel spread sheet, and compared with the critical value. The functions `Analyze.Binomial`, `Analyze.Poisson`, and `Analyze.CondPoisson` work very differently, in that they are run at each look at the accumulating data, whenever a new group of data arrives, and it is meant to perform the test itself, i.e., there is no need to use hand calculators or excel spread sheets or any other auxiliar code. The results and conclusions, including a descriptive table and illustrative graphics, are automatically provided after running `Analyze.Binomial`, `Analyze.Poisson`, or `Analyze.CondPoisson`.

Important: before using these functions, though, it is necessary to first run the functions `AnalyzeSetup.Binomial`, `AnalyzeSetup.Poisson`, or `AnalyzeSetup.CondPoisson` once in order to set everything up for the sequential analysis.

Comparison with Other R Packages for Sequential Analysis

The R `Sequential` package is designed for sequential analysis where statistical hypothesis testing is performed using gradually accumulating data. It is not designed for quality control problems, where a process is monitored over time to detect an emerging problem due to a sudden increase in the excess risk. Although the methods for sequential analysis and quality control may seem similar, as they both analyze gradually accumulating data, they are actually very different in both their purpose and design. Under the sequential hypothesis testing approach, the objective is to quickly determine if there is some intrinsic excess risk, with the assumption that this risk does not change over time. For example, we may want to know if drug A is better than drug B, and there is no reason to believe that the behavior of the drugs change over time. In the quality control setting, the objective is instead to detect a possible change in a stochastic process that may occur in the future, and to detect that change as soon as possible after it occurs. For example, the heart of a hospital patient is beating as it should, but if there is a sudden deterioration, the alarm should sound as soon as possible without generating a lot of false alarms. This package is only meant for sequential analysis of the former type, and it should not be used for quality control type problems. For quality control type analyses, there are other R packages available, such as `graphicsQC`, `IQCC`, `MetaQC`, `MSQC`, `qcc`, and `qcr`.

In a number of ways, the R `Sequential` package differs from other R packages for sequential analyses. Historically, most sequential analysis has been conducted using asymptotic statistical theory, and that is also what is used in the `gsDesign`, `lbound`, `PwrGSD`, `seqDesign`, `seqmon`, and `sglr` R packages. In contrast, the R `Sequential` package is based on exact results, using iterative numerical calculations, rather than using asymptotic theory or computer simulations.

With this package, it is only possible to analyze binomial/Bernoulli, Poisson, or conditional Poisson data. For other probability distributions, such as normal or exponential data, other R packages should be consulted, such as `GroupSeq` or `SPRT`. Moreover, all functions in this package uses a one-sided upper bound to reject the null hypothesis, while the analyses end without rejecting the null

when an upper limit on the sample size is reached. For two sided sequential analysis, or other types of rejection boundaries, other R packages must be used, such as e.g. `ldbounds` and `Binseqtest`. Finally, in this package, there are functions for both continuous and group sequential analysis, and it is also possible to analyze situations where some of the data arrives continuously while other parts of the data arrives in groups. Most other R packages are exclusively designed for group sequential analysis, but there are some that also do continuous sequential analysis, such as `Binseqtest` and `SPRT`, but `Binseqtest` is only for binomial data type, and `SPRT` is for simple alternative hypothesis, while `Sequential` can be used for binomial and Poisson data and is meant to composite alternative hypothesis. The present package offers the possibility to calculate the expected time to signal through the `Performance.Poisson`, `Performance.G.Poisson`, `Performance.Binomial`, `Performance.G.Binomial`, and `Performance.CondPoisson` functions, which is not offered by the other packages cited above. Another important advantage of the `Sequential` package is the possibility of eliciting, through exact calculations, the minimum sample size needed to accomplish with target statistical powers through the functions `SampleSize.Poisson`, `SampleSize.CondPoisson`, and `SampleSize.Binomial`.

Acknowledgements

Development of the R `Sequential` package has been funded and supported by:

- Food and Drug Administration, USA, through the Mini-Sentinel Project (v1.0,1.1,2.0).
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0).
- Federal University of Ouro Preto (UFOP), through contract under internal UFOP's resolution CEPE 4600 (v2.0).
- National Council of Scientific and Technological Development (CNPq), Brazil (v1.0).
- Bank for Development of the Minas Gerais State (BDMG), Brazil (v1.0).

Feedback from users is greatly appreciated. Very valuable suggestions concerning the R `Sequential` package have been received from various individuals, including:

- Ron Berman, University of California Berkeley.
- Claudia Coronel-Moreno, Harvard Pilgrim Health Care Institute.
- Bruce Fireman, Kaiser Permanente Northern California.
- Josh Gagne, Harvard Medical School and Brigham and Women's Hospital.
- Ned Lewis, Kaiser Permanente Northern California.
- Judith Maro, Harvard Medical School and Harvard Pilgrim Health Care Institute.
- Azadeh Shoaibi, Food and Drug Administration.
- Katherine Yih, Harvard Medical School and Harvard Pilgrim Health Care Institute.
- Jie Tang, Clinical biostatistics, Janssen R and D US, Johnson and Johnson LLC.
- Tuomo A. Nieminen, The National Institute for Health and Welfare (THL), Finland.
- Andreia Leite, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine.
- Laura Hou, Harvard Pilgrim Health Care Institute, Boston, USA.
- Abdurrahman Abdurrob, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School.
- Kirk Snyder, Information Management Services, Inc.

Version History of the R Sequential Package

Version 1.1, February 2013

Exact sequential analysis for Poisson data:

- Exact continuous sequential analysis.
- Exact group sequential analysis with pre-defined and constant groups sizes.
- Wald type rejection boundary.
- Statistical power, expected time to signal and sample size calculations.
- User guide.

Version 1.2, January 2014

- Improved code structure and efficiency.
- More extensive user guide.

Version 2.0, June 2015

Exact sequential analysis for binomial data:

- Continuous sequential analysis.
- Group sequential analysis with pre-defined group sizes.
- Group sequential analysis with unpredictable group sizes, not specified a priori.
- Fixed or variable binomial probabilities (matching ratios).
- User specified alpha spending function.
- Statistical power, expected time to signal and sample size calculations.
- Updated user guide.

Version 2.0.1, June 2015

- Correction of bugs in `CV.Poisson` function.
- Updated user guide.

Version 2.0.2, October 2015

- Improved user guide.

Version 2.1, May 2016

Exact sequential analysis for Poisson data:

- Group sequential analysis with unpredictable group sizes, not specified a priori.
- User specified alpha spending function.
- Mixed group-continuous sequential analysis.
- Statistical power, expected time to signal and sample size calculations for non-constant groups sizes.

Other:

- Directory address parameter in `AnalyzeSetUp` functions.
- Probability parameter in binomial functions.
- Updated user guide.

Version 2.1.1, June 2016

- Correction of bugs in Poisson functions.
- Updated user guide.

Version 2.2, July 2016

- Critical Value, Performance, and SampleSize calculations for `CMaxSPRT` with Poisson data.

- Updated user guide.

Version 2.2.1, September 2016

- Correction of bugs in `CV.Poisson` and `CV.G.Poisson` functions.

- Updated user guide.

Version 2.3, Dec 2016

- Correction of bugs in the `SampleSize.Binomial` function.

- Improvement of `SampleSize` functions for considering vectors for the input parameters `R` and `power`.

- Inclusion of the new functions `AnalyzeSetUp.CondPoisson` and `Analyze.CondPoisson`.

- Updated user guide.

Version 2.3.1, Feb 2017

- Correction of bugs in `Analyze.Binomial` and `AnalyzeSetUp.Poisson` functions.

- Adjustment on the relative risk estimation method for `Analyze.Binomial` function.

- Updated user guide.

Version 2.3.2, Aug 2017

- Correction of bugs in `Analyze.Binomial`.

- Updated user guide.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

Maintainer: Ivair Ramos Silva <jamesivair@yahoo.com.br>

References

Chin R. (2012), *Adaptive and Flexible Clinical Trials*, Boca Raton, FL: Chapman and Hall/CRC.

Cook TD, DeMets DL. (2007), *Introduction to Statistical Methods for Clinical Trials*: Chapman and Hall/CRC Texts in Statistical Science.

Fireman B, et al. (2013) Exact sequential analysis for binomial data with timevarying probabilities. Manuscript in Preparation.

Friedman LM, Furberg CD, DeMets D. (2010), *Fundamentals of Clinical Trials*, 4th ed.: Springer.

Ghosh BK, Sen PK. (1991), *Handbook of Sequential Analysis*, New York: MARCEL DEKKER, Inc.

Ghosh M, Mukhopadhyay N, Sen PK. (2011), *Sequential Estimation*: Wiley.

Jennison C, Turnbull B. (2000), *Group Sequential Methods with Applications to Clinical Trials*, London: Chapman and Hall/CRC.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, 30: 58–78.

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Mukhopadhyay N, Silva BM. (2002), Sequential Methods and Their Applications, 1th ed.: Chapman and Hall/CRC.

Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Silva IR, Li L, Kulldorff M. (2016). Exact Conditional Sequential Testing for Poisson Data. Working paper.

Xia Qi. (2007), A Procedure for Group Sequential Comparative Poisson Trials. *Journal of Biopharmaceutical Statistics*, 17, 869–881.

Wald A. (1945), Sequential Tests of Statistical Hypotheses, *Annals of Mathematical Statistics*, 16, 117–186.

Wald A. (1947), *Sequential Analysis*. New York: John Wiley and Sons.

Whitehead J. (1997), *The Design and Analysis of Sequential Clinical Trials*, 2th ed.: Wiley.

Examples

```
## Critical value for continuous sequential analyses for Poisson Data.
## Maximum sample size = 10, alpha = 0.05 and minimum number of events = 3:

cvt<- CV.Poisson(SampleSize=10,D=0,M=3,alpha=0.05)

## Statistical power and the expected time to signal for relative risk RR=2:

result<- Performance.Poisson(SampleSize=10,D=0,M=3,cv=cvt,RR=2)

# And if you type:
result

# Then you will see the following:
#           Power ESignalTime ESampleSize
# [1,] 0.7329625   4.071636   5.654732
```

Analyze.Binomial	<i>Function to Conduct Group Sequential Analyses for Binomial Data When the Group Sizes are not Known a Priori.</i>
------------------	---

Description

The function `Analyze.Binomial` is used for either continuous or group sequential analysis, or for a combination of the two. Unlike `CV.Binomial` and `CV.G.Binomial`, it is not necessary to pre-specify the group sizes before the sequential analysis starts. Moreover, under the null hypothesis, the binomial probability, p , can be different for different observations. In a matched case-control setting, this means that the matching ratios can be different for different matched sets. It is possible to use either a Wald type rejection boundary, which is flat with respect to the likelihood ratio, or a user defined alpha spending function. `Analyze.Binomial` is run at each look at the data. Before running it by the first time, it is necessary to run the [AnalyzeSetUp.Binomial](#) function.

Usage

```
Analyze.Binomial(name, test, z="n", p="n", cases, controls, AlphaSpend="n")
```

Arguments

name	The name of the sequential analysis. Must be identical for all looks at the data, and it must be the same as the name given by the AnalyzeSetup.Binomial function. Should never be the same as another sequential analysis that is run simultaneously on the same computer.
test	An integer indicating the number of hypothesis tests performed up to and including the current test. For example, if there were four prior looks at the data, and this is the fifth one, then "test=5". This number should be increased by one each time that the Analyze.Binomial function is run for a new group of data, when it is part of the same sequential analysis. If not, there is an error message.
z	For a matched case-control analysis, z is the number of controls matched to each case. For example, if there are 3 controls matched to each case, "z=3". In a self-control analysis, z is the ratio of the length of the control interval to the length of the risk interval. For example, if the risk interval is 2 days long and the control interval is 7 days long, "z=7/2". In terms of p, the binomial probability under the null hypothesis, "p=1/(1+z)", or equivalently, "z=1/p-1". The parameter z must be a positive number. The default value is z=1 (p=0.5). If the ratio is the same for all observations, then z can be any positive number. If the ratio is different for different observations, then z is a vector of positive numbers.
p	The probability of having a case under the null hypothesis. There is no default value.
cases	A number or a vector of the same length as z containing the number of cases.
controls	A number or a vector of the same length as z containing the number of controls.
AlphaSpend	The alpha spending function is specified in the AnalyzeSetup.Binomial function. At any look at the data, it is possible to over ride that pre-specified alpha spending plan by using the AlphaSpend parameter. AlphaSpend is a number representing the maximum amount of alpha (Type I error probability) to be spent up to and including the current test. Because of the discrete nature of the binomial distribution, the actual amount of alpha spent may be less than the maximum amount specified. It must be in the range (0,alpha]. The default value is no override, which means that, if AlphaSpend= "n", then the function will use the alpha spending plan specified in the AnalyzeSetup.Binomial function.

Details

The function `Analyze.Binomial` performs continuous or group sequential analysis for Bernoulli or binomial data. It can also be used for mixed continuous-group sequential analysis where some data arrives continuously while other data arrives in groups. Unlike `CV.Binomial` and `CV.G.Binomial`, there is (i) no need to pre-specify the group sizes before the sequential analysis starts, (ii) a variety of alpha spending functions are available, and (iii) it is possible to include an offset term where, under the null hypothesis, different observations have different binomial probabilities p.

In sequential analysis, data is formed by cumulative information, collected in separated chunks or groups, which are observed at different moments in time. `Analyze.Binomial` is run each time a new group of data arrives at which time a new sequential test is conducted. When running `Analyze.Binomial`, only the data from the new group should be included when calling the function. The prior data has been stored, and it will be automatically retrieved by `Analyze.Binomial`, with no need to reenter that data. Before running `Analyze.Binomial` for the first time, it is necessary to set up the sequential analysis using the `AnalyzeSetUp.Binomial` function, which is run once, and just once, to define the sequential analysis parameters. For information about this, see the description of the `AnalyzeSetUp.Binomial` function.

The function `Analyze.Binomial` calculates critical values to determine if the null hypothesis should be rejected or not at each analysis. Critical values are given in the scale of the number of cases. This is done for a pre-specified overall statistical significance level (α), and for an upper limit on the sample size (N). The exact analytical solution is obtained through numerical calculations. Based on the data and the critical value, the function determines if the null hypothesis should be rejected or not, and if subsequent tests should be conducted. After each test, the function also provides information about the amount of α that has been spent, the cumulative number of cases and controls, and the maximum likelihood estimate of the relative risk.

For binomial and Bernoulli data, there are a number of 0/1 observations that can either be a case or a control. Under the null hypothesis, the probability of being a case is p , and the probability of being a control is $1-p$. If data comes from a self-control analysis, the observation is a case if the event occurred in the risk interval, and it is a control if the event occurred in the control interval. Under the null hypothesis, we then have that $p = 1/(1+z)$, where z is the ratio of the length of the control interval to the length of the risk interval. This ratio, and hence p , does not need to be the same for all observations. If data comes from a matched set of exposed and unexposed individuals, then the observation is a case if the event occurred among one of the exposed, and it is a control if it occurred among one of the unexposed. Under the null hypothesis, $p = 1/(1+z)$, where z is the number of unexposed individuals divided by the number of exposed individuals in the matched set. Again, this ratio does not have to be the same for all matched sets. The variable z can be any positive number.

If the ratio parameter z , and hence p , is the same for all observations in the same group of data, then z is just a positive number. On the other hand, if different observations in the same group of data have different values for z , then z is a vector, representing multiple z values. For each value of z , it is necessary to specify the number of cases and the number of controls. This means that for a group of data, the vector of z s has to be of the same length as the vector of cases and the vector of controls. The first entry of the vector z is the matching ratio associated to the first entries of cases and of controls. The second entry of z is the matching ratio with respect to the second entries of cases and of controls, and so on. For example, consider that each of five observations came from four different matching ratios. In this situation, the vectors cases, controls and z are all of length four. For example, suppose " $z=c(2,1,0.5,3)$ ", " $cases=c(1,1,0,0)$ " and " $controls=c(0,0,1,2)$ ". The matching ratio for the first observation, which turned out as a case, is equal to 2. For the second observation, also a case, the matching is equal to 1. With a matching ration of 0.5, the third observation turned out to be a control. The two last observations both had a matching ratio of 3, and both of them were controls. If all observations in the same data group has the same ratio, the vectors are of size one, that is, they are simple numbers. For example, if there were ten observations that all had a ratio of 2, with seven cases and three controls, we have " $z=2$ ", " $cases=7$ ", and " $controls=3$ ".

Alternatively, instead of z the user can specify p directly. Note that only one of these inputs, z or p , has to be specified, but if both are entered the code will only work if z and p are such that $p=1/(1+z)$.

Otherwise, an error message will appear to remind that such condition must be complied.

Before running `Analyze.Binomial`, it is necessary to specify a planned default alpha spending function, which is done using the `AlphaSpendType` parameter in the `AnalyzeSetUp.Binomial` function. The default alpha spending plan can be either, (i) the polynomial power-type alpha spending plan, which is parameterized with ρ , or (ii) the alpha spending associated to the Wald-type rejection boundary, which is flat with respect to the likelihood ratio. See the [AnalyzeSetUp.Binomial](#) for more details.

In most cases, this pre-specified alpha spending function is used throughout the analysis, but if needed, it is possible to override it at any or each of the sequential tests. This is done using the `AlphaSpend` parameter, which specifies the maximum amount of alpha to spend up to and including the current test. In this way, it is possible to use any alpha spending function, and not only those available in `AnalyzeSetUp.Binomial`. It is also possible to use a flexible adaptive alpha spending plan that is not set in stone before the sequential analysis starts. The only requirement is that for a particular test with a new group of data, `AlphaSpend` must be decided before knowing the number of cases and controls in that group. To ensure a statistically valid sequential analysis, `AlphaSpend` can only depend on the number of events (cases + controls) at prior tests and the total number of events in the current test. This is important.

The function `Analyze.Binomial` is meant to perform the binomial sequential analysis with a certain level of autonomy. After running a test, the code offers a synthesis about the general parameter settings, the main conclusions concerning the acceptance or rejection of the null hypothesis, and the historical information from previous tests. A table with the main analyses results is automatically printed in the R console. Each column of the table contains a historical characteristic, including the information for the current test. Each line of the table corresponds to a specific test organized by calendar time. The table is titled with the title input defined through the function `AnalyzeSetUp.Binomial`, and its columns are organized and labeled in the following way: "Test", "Cases", "Controls", "Cumulative Cases", "Cumulative Controls", "Cumulative E[Cases]", "RR", "LLR", "target", "actual", "CV", "Reject H0". Here follows a short description of each column:

- "Test" shows the order of the analysis, i.e., the arrival order of each chunk of data.
- "Cases" and "Controls" present the total of cases and controls that entered at each test, respectively.
- "Cumulative Cases" and "Cumulative Controls" in the i -th line have the cumulative counts of cases and controls up to the i -th test, respectively.
- "Cumulative E[Cases]" in line i is the expected cumulative number of cases for the i -th test under the null hypothesis.
- "RR" is the estimated relative risk for test i .
- "LLR" is the observed log-likelihood ratio test statistic.
- "target" is the target alpha spending for the i -th test.
- "actual" is the actual alpha spent up to the i -th test.
- "CV" is the critical value in the scale of the number of cases, showing how many cases are needed to reject the null hypothesis at this test.
- "Reject H0" is a logical variable that is "Yes" when the null hypothesis is rejected, and the label "No" when H0 is not to be rejected

Observe that, because the binomial distribution is discrete, the target alpha spending will rarely be reached. The actual alpha spending is then shown to facilitate a realistic interpretation of the results.

The function `Analyze.Binomial` was designed to instruct the user with minimal information about bugs from the code, or about non-applicable parameter input usages. Some entries are not applicable for the parameter inputs. For example, the input "z" must be a positive number, and then if the user sets "z= -1", the code will report an error with the message "the entries of the vector "z" must be positive numbers". Thus, messages will appear when mistakes and inconsistencies are detected, and instructions about how to proceed to solve such problems will automatically appear.

Value

`result` A table containing the main characteristics, conclusions concerning the acceptance or rejection of the null hypothesis, and the historical information from previous tests.

Acknowledgements

Development of the `Analyze.Binomial` function was funded by: - Food and Drug Administration, Center for Drug Evaluation and Research, through Mini-Sentinel Project: base version, documentation, unequal matching ratios;
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999: user-defined alpha spending functions, power-type alpha spending function, increased computational speed, confidence intervals for relative risks, end of schedule analysis using left-over alpha, enhanced error handling and messages, improved documentation.

We thank Claudia Coronel-Moreno for valuable editorial support, Bruce Fireman for general guidance, and Josh Gagne for important feedback on the unequal matching ratio feature.

See also

[AnalyzeSetUp.Binomial](#): for setting up sequential analysis with the `Analyze.Binomial` function, before the first look at the data.

[Performance.G.Binomial](#): for calculating the statistical power, expected time to signal and expected sample size for group sequential analysis with binomial data.

[SampleSize.Binomial](#): for calculating the needed sample size to achieve the desired statistical power for continuous sequential analysis with binomial data.

[CV.G.Binomial](#): for calculating critical values for group sequential analysis with binomial data.

[CV.G.Poisson](#): for calculating critical values for group sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva, Ned Lewis, Martin Kulldorff.

References

Fireman B, et al. (2013). Exact sequential analysis for binomial data with time varying probabilities. Manuscript in preparation.

Jennison C, Turnbull B. (2000). Group Sequential Methods with Applications to Clinical Trials. London: Chapman and Hall/CRC.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, 30, 58–78.

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. REVSTAT Statistical Journal, 15(3): 373–394.

Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. Biometrics, 71 (3), 851–858.

Examples

```
### Example. Four chunks of data.

### Firstly, it is necessary to set up the input parameters.
## Here we use the Wald type alpha spending.
## Note: cut off the "#" symbol before running the two lines below.
#   AnalyzeSetUp.Binomial(name="VaccineA",N=200,alpha=0.05,zp=1,M=3,
#   AlphaSpendType="Wald", title="Monitoring_vaccineA",
#   address="C:/Users/Ivair/Documents")

### Now we apply sequential tests to each of four chunks of data.
# -----

## Test 1 - Situation where each individual event came from a different
## matching ratio.
## This first test uses the default Wald type alpha spending (AlphaSpend="n").
## Note: cut off the "#" symbol before running the line below.
# Analyze.Binomial(name= "VaccineA",test=1,z=c(1.1,1.3,1.2,1),
# cases= c(1,0,0,0), controls= c(0,1,1,1) )

## Test 2 - Situation where some of the events came from the same matching
## ratio.
## Observe that here we use an arbitrary alpha spending of 0.02.
## Note: cut off the "#" symbol before running the line below.
# Analyze.Binomial(name= "VaccineA",test=2,z=c(1,1.5),cases= c(12,1),
# controls= c(0,10), AlphaSpend=0.02)

## Test 3 - Situation of elevated number of events, but now the
## arbitrary alpha spending is of 0.04, and p is entered instead of z.
## Note: cut off the "#" symbol before running the line below.
# Analyze.Binomial(name= "VaccineA",test=3,p=c(0.4,0.5),cases= c(12,10),
# controls= c(10,14), AlphaSpend=0.04)

## Test 4 - Situation where all the events came from the same matching
## ratio.
## Here the original target alpha spending is used.
## Note: cut off the "#" symbol before running the line below.
# Analyze.Binomial(name= "VaccineA",test=4,z=2,cases= 20,controls= 10)
```

Analyze.CondPoisson *Function to Conduct Group Sequential Analyses for Conditional Poisson Data When the Group Sizes are not Known a Priori.*

Description

The function `Analyze.CondPoisson` is used for either continuous, group, or mixed continuous-group sequential analysis for Poisson data conditioned on observed historical data. Unlike `CV.CondPoisson`, it is not necessary that data arrives in a near-continuous fashion. It is possible to use either a Wald type rejection boundary, which is flat with respect to the likelihood ratio, or a user defined alpha spending function. `Analyze.CondPoisson` is run at each look at the data. Before running it by the first time, it is necessary to run the [AnalyzeSetUp.CondPoisson](#) function.

Usage

```
Analyze.CondPoisson(name, test, events, PersonTimeRatio, AlphaSpend="n")
```

Arguments

name	The name of the sequential analysis. Must be identical for all looks at the data, and it must be the same as the name given by the <code>AnalyzeSetUp.CondPoisson</code> function. Should never be the same as another sequential analysis that is run simultaneously on the same computer.
test	An integer indicating the number of hypothesis tests performed up to and including the current test. For example, if there were four prior looks at the data, and this is the fifth one, then "test=5". This number should be increased by one each time that the <code>Analyze.CondPoisson</code> function is run for a new group of data, when it is part of the same sequential analysis. If not, there is an error message.
events	The number of observed events during the surveillance period.
PersonTimeRatio	The observed ratio between the punctual, instead of cumulative from previous tests, person-time observed in the current test, by the total person-time observed in the historical period.
AlphaSpend	The alpha spending function is specified in the <code>AnalyzeSetUp.CondPoisson</code> function. At any look at the data, it is possible to over ride that pre-specified alpha spending plan by using the <code>AlphaSpend</code> parameter. <code>AlphaSpend</code> is a number representing the maximum amount of alpha (Type I error probability) to be spent up to and including the current test. Because of the discrete nature of the Poisson distribution, the actual amount of alpha spent may be less than the maximum amount specified. It must be in the range (0,alpha]. The default value is no override, which means that, if <code>AlphaSpend= "n"</code> , then the function will use the alpha spending plan specified in the <code>AnalyzeSetUp.CondPoisson</code> function.

Details

The function `Analyze.CondPoisson` performs continuous or group sequential analysis for Poisson data conditioned on observed historical data. It can also be used for mixed continuous-group sequential analysis where some data arrives continuously while other data arrives in groups. Unlike `CV.CondPoisson`, there is a variety of alpha spending functions available.

In sequential analysis, data is formed by cumulative information, collected in separated chunks or groups, which are observed at different moments in time. `Analyze.CondPoisson` is run each

time a new group of data arrives at which time a new sequential test is conducted. When running `Analyze.CondPoisson`, only the data from the new group should be included when calling the function. The prior data has been stored, and it will be automatically retrieved by `Analyze.CondPoisson`, with no need to reenter that data. Before running `Analyze.CondPoisson` for the first time, it is necessary to set up the sequential analysis using the `AnalyzeSetUp.CondPoisson` function, which is run once, and just once, to define the sequential analysis parameters. For information about this, see the description of the [AnalyzeSetUp.CondPoisson](#) function.

The function `Analyze.CondPoisson` calculates critical values to determine if the null hypothesis should be rejected or not at each analysis. Critical values are given in the scale of the number of events. This is done for a pre-specified overall statistical significance level (α), and for an upper limit on the sample size, which is given by "T" or "K". Go to the documentation of [AnalyzeSetUp.CondPoisson](#) for more details about the choice between "T" or "K".

The exact analytical solution is obtained through numerical calculations. Based on the data and the critical value, the function determines if the null hypothesis should be rejected or not, and if subsequent tests should be conducted. After each test, the function also provides information about the amount of α that has been spent, the cumulative number of events, and the observed log-likelihood ratio statistic.

Before running `Analyze.CondPoisson`, it is necessary to specify a planned default alpha spending function, which is done using the `AlphaSpendType` parameter in the `AnalyzeSetUp.CondPoisson` function. The default alpha spending plan can be either, (i) the polynomial power-type alpha spending plan, which is parameterized with ρ , or (ii) the alpha spending associated to the Wald-type rejection boundary, which is flat with respect to the likelihood ratio. See the [AnalyzeSetUp.CondPoisson](#) for more details.

In most cases, this pre-specified alpha spending function is used throughout the analysis, but if needed, it is possible to override it at any or each of the sequential tests. This is done using the `AlphaSpend` parameter, which specifies the maximum amount of alpha to spend up to and including the current test. In this way, it is possible to use any alpha spending function, and not only those available in `AnalyzeSetUp.CondPoisson`. It is also possible to use a flexible adaptive alpha spending plan that is not set in stone before the sequential analysis starts. The only requirement is that for a particular test with a new group of data, `AlphaSpend` must be decided before knowing the `PersonTimeRatio` in that group. Hence, in order to ensure a statistically valid sequential analysis, `AlphaSpend` can only depend on the cumulative events. This is important.

The function `Analyze.CondPoisson` is meant to perform the conditional Poisson sequential analysis with a certain level of autonomy. After running a test, the code offers a synthesis about the general parameter settings, the main conclusions concerning the acceptance or rejection of the null hypothesis, and the historical information from previous tests. A table with the main analyses results is automatically printed in the R console. Each column of the table contains a historical characteristic, including the information for the current test. Each line of the table corresponds to a specific test organized by calendar time. The table is titled with the title input defined through the function `AnalyzeSetUp.CondPoisson`, and its columns are organized and labeled in the following way: "Test", "Person-timeR", "events", "Cumulative Person-timeR", "Cumulative events", "LLR", "target", "actual", "CV", "Reject H0". Here follows a short description of each column:

- "Test" shows the order of the analysis, i.e., the arrival order of each chunk of data.
- "Person-timeR" shows the observed ratio between the punctual person-time observed in the current test by the total person-time observed in the historical period.
- "Events" presents the observed number of events from the Poisson counting entered at each test.

- "Cumulative Person-timeR" shows the observed person-time ratio up to the current test.
- "Cumulative events" presents the observed number of events from the Poisson counting up to the current test.
- "LLR" is the observed log-likelihood ratio test statistic.
- "target" is the target alpha spending for the i-th test.
- "actual" is the actual alpha spent up to the i-th test.
- "CV" is the critical value in the scale of the log-likelihood ratio test statistic.
- "Reject H0" is a logical variable that receives the label "Yes" when the null hypothesis is rejected, and the label "No" when H0 is not to be rejected

Observe that, depending on the choices of the input parameters M and alpha through the [AnalyzeSetUp.CondPoisson](#) function, the actual alpha spending can differ from the target one. The actual alpha spending is then shown in order to favor a realistic interpretation of the results.

The function `Analyze.CondPoisson` was designed to instruct the user with minimal information about bugs from the code, or about non-applicable input parameters usage. Some entries are not applicable. For example, the input "Person-timeR" must be a positive number, and then if the user sets "Person-timeR= -1", then the code will report an error with the message "the entry of "Person-timeR" must be a number greater than zero". Thus, messages will appear when mistakes and inconsistencies are detected. Instructions about how to proceed to solve such problems will automatically appear too.

Value

result	A table containing the main characteristics, conclusions concerning the acceptance or rejection of the null hypothesis, and the historical information from previous tests.
--------	---

Acknowledgements

Development of the `Analyze.CondPoisson` function was funded by: - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999. - Foundation for Research Support of Minas Gerais State (FAPEMIG), MG, Brazil, through the grant Demanda Universal.

See also

[AnalyzeSetUp.CondPoisson](#): for setting up sequential analysis with the `Analyze.CondPoisson` function, before the first look at the data.

[AnalyzeSetUp.Poisson](#): for setting up sequential analysis with the `Analyze.Poisson` function, before the first look at the data.

[Performance.G.Poisson](#): for calculating the statistical power, expected time to signal and expected sample size for group sequential analysis with Poisson data.

[SampleSize.Poisson](#): for calculating the needed sample size to achieve the desired statistical power for continuous sequential analysis with Poisson data.

[CV.G.Poisson](#): for calculating critical values for group sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

- Fireman B, et al. (2013). Exact sequential analysis for Poisson data with time varying probabilities. Manuscript in preparation.
- Jennison C, Turnbull B. (2000). Group Sequential Methods with Applications to Clinical Trials. London: Chapman and Hall/CRC.
- Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, 30, 58–78.
- Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.
- Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.
- Silva IR, Li L, Kulldorff M. (2016). Exact Conditional Sequential Testing for Poisson Data. Working paper.

Examples

```
### Example. Four chunks of data.

### Firstly, it is necessary to set up the input parameters.
## Here we use the Wald type alpha spending.
## Note: cut off the "#" symbol before running the two lines below, and,
## Important: choose an actual "address" to save your set up information.
# AnalyzeSetUp.CondPoisson(name="TestA", SampleSizeType="Events", K=100,
# cc=20,alpha=0.05, M=1,AlphaSpendType="power-type",rho=0.5,title="n",
# address="C:/Users/Ivair/Documents")

### Now we apply a test for each one of three chunks of data.
# -----

## Test 1 - Situation where the expected number of events is equal to 5.
## The observed hypothetical information is "PersonTimeRatio=0.5".
## Note: cut off the "#" symbol before running the line below.

# Analyze.CondPoisson(name="TestA",test=1,events=5,PersonTimeRatio=0.5)

## Test 2 - Situation where the expected number of events is equal to 6.
## The observed hypothetical information is "PersonTimeRatio=0.3".

#Analyze.CondPoisson(name="TestA",test=2,events=6,PersonTimeRatio=0.3)

## Test 3 - Situation where the expected number of events is equal to 10.
## The observed hypothetical information is "PersonTimeRatio=0.1".

#Analyze.CondPoisson(name="TestA",test=3,events=10,PersonTimeRatio=0.1)
```

Analyze.Poisson	<i>Function to Conduct Group Sequential Analyses for Poisson Data When the Goup Sizes are not Known a Priori.</i>
-----------------	---

Description

The function `Analyze.Poisson` is used for either continuous or group sequential analysis, or for a combination of the two. Unlike `CV.Poisson` and `CV.G.Poisson`, it is not necessary to pre-specify the group sizes before the sequential analysis starts. Moreover, under the null hypothesis, the expected number of events, μ_0 , can be different for different observations. It is possible to use either a Wald type rejection boundary, which is flat with respect to the likelihood ratio, or a user defined alpha spending function. `Analyze.Poisson` is run at each look at the data. Before running it by the first time, it is necessary to run the [AnalyzeSetUp.Poisson](#) function.

Usage

```
Analyze.Poisson(name, test, mu0, events, AlphaSpend="n")
```

Arguments

<code>name</code>	The name of the sequential analysis. Must be identical for all looks at the data, and it must be the same as the name given by the <code>AnalyzeSetup.Poisson</code> function. Should never be the same as another sequential analysis that is run simultaneously on the same computer.
<code>test</code>	An integer indicating the number of hypothesis tests performed up to and including the current test. For example, if there were four prior looks at the data, and this is the fifth one, then "test=5". This number should be increased by one each time that the <code>Analyze.Poisson</code> function is run for a new group of data, when it is part of the same sequential analysis. If not, there is an error message.
<code>mu0</code>	The expected number of events under the null hypothesis. The parameter μ_0 must be a positive number. There is no default value.
<code>events</code>	The number of observed events.
<code>AlphaSpend</code>	The alpha spending function is specified in the <code>AnalyzeSetUp.Poisson</code> function. At any look at the data, it is possible to over ride that pre-specified alpha spending plan by using the <code>AlphaSpend</code> parameter. <code>AlphaSpend</code> is a number representing the maximum amount of alpha (Type I error probabiliy) to be spent up to and including the current test. Because of the discrete nature of the Poisson distribution, the actual amount of alpha spent may be less than the maximum amount specified. It must be in the range $(0, \alpha]$. The default value is no over-ride, which means that, if <code>AlphaSpend= "n"</code> , then the function will use the alpha spending plan specified in the <code>AnalyzeSetUp.Poisson</code> function.

Details

The function `Analyze.Poisson` performs continuous or group sequential analysis for Poisson data. It can also be used for mixed continuous-group sequential analysis where some data arrives continuously while other data arrives in groups. Unlike `CV.Poisson` and `CV.G.Poisson`, there is (i) no need to pre-specify the group sizes before the sequential analysis starts, (ii) a variety of alpha spending functions are available, and (iii) it is possible to include an offset term where, under the null hypothesis, different observations have different Poisson rates μ_0 .

In sequential analysis, data is formed by cumulative information, collected in separated chunks or groups, which are observed at different moments in time. `Analyze.Poisson` is run each time a new group of data arrives at which time a new sequential test is conducted. When running `Analyze.Poisson`, only the data from the new group should be included when calling the function. The prior data has been stored, and it will be automatically retrieved by `Analyze.Poisson`, with no need to reenter that data. Before running `Analyze.Poisson` for the first time, it is necessary to set up the sequential analysis using the `AnalyzeSetUp.Poisson` function, which is run once, and just once, to define the sequential analysis parameters. For information about this, see the description of the `AnalyzeSetUp.Poisson` function.

The function `Analyze.Poisson` calculates critical values to determine if the null hypothesis should be rejected or not at each analysis. Critical values are given in the scale of the number of events. This is done for a pre-specified overall statistical significance level (α), and for an upper limit on the sample size (N). The exact analytical solution is obtained through numerical calculations. Based on the data and the critical value, the function determines if the null hypothesis should be rejected or not, and if subsequent tests should be conducted. After each test, the function also provides information about the amount of alpha that has been spent, the cumulative number of events, and the estimated relative risk.

Before running `Analyze.Poisson`, it is necessary to specify a planned default alpha spending function, which is done using the `AlphaSpendType` parameter in the `AnalyzeSetUp.Poisson` function. The default alpha spending plan can be either, (i) the polynomial power-type alpha spending plan, which is parameterized with ρ , or (ii) the alpha spending associated to the Wald-type rejection boundary, which is flat with respect to the likelihood ratio. See the `AnalyzeSetUp.Poisson` for more details.

In most cases, this pre-specified alpha spending function is used throughout the analysis, but if needed, it is possible to override it at any or each of the sequential tests. This is done using the `AlphaSpend` parameter, which specifies the maximum amount of alpha to spend up to and including the current test. In this way, it is possible to use any alpha spending function, and not only those available in `AnalyzeSetUp.Poisson`. It is also possible to use a flexible adaptive alpha spending plan that is not set in stone before the sequential analysis starts. The only requirement is that for a particular test with a new group of data, `AlphaSpend` must be decided before knowing the number of events in that group. To ensure a statistically valid sequential analysis, `AlphaSpend` can only depend on cumulative μ_0 values at prior tests and of the μ_0 value in the current test. This is important.

The function `Analyze.Poisson` is meant to perform the Poisson sequential analysis with a certain level of autonomy. After running a test, the code offers a synthesis about the general parameter settings, the main conclusions concerning the acceptance or rejection of the null hypothesis, and the historical information from previous tests. A table with the main analyses results is automatically printed in the R console. Each column of the table contains a historical characteristic, including the information for the current test. Each line of the table corresponds to a specific test

organized by calendar time. The table is titled with the title input defined through the function `AnalyzeSetUp.Poisson`, and its columns are organized and labeled in the following way: "Test", "mu0", "Events", "Cumulative mu0", "Cumulative Events", "RR", "LLR", "target", "actual", "CV", "Reject H0". Here follows a short description of each column:

- "Test" shows the order of the analysis, i.e., the arrival order of each chunk of data.
- "mu0" is the expected number of events under the null hypothesis for the chunk of data to be analyzed at each test.
- "Events" presents the observed number of events from the Poisson counting entered at each test.
- "Cumulative mu0" expected number of events under the null hypothesis up to the i-th test.
- "Cumulative Events" observed number of events up to the i-th test.
- "RR" is the estimated relative risk for test i.
- "LLR" is the observed log-likelihood ratio test statistic.
- "target" is the target alpha spending for the i-th test.
- "actual" is the actual alpha spent up to the i-th test.
- "CV" is the critical value in the scale of the number of events, showing how many events are needed to reject the null hypothesis at this test.
- "Reject H0" is a logical variable that is "Yes" when the null hypothesis is rejected, and the label "No" when H0 is not to be rejected

Observe that, because the Poisson distribution is discrete, the target alpha spending will rarely be reached. The actual alpha spending is then shown to facilitate a realistic interpretation of the results.

The function `Analyze.Poisson` was designed to instruct the user with minimal information about bugs from the code, or about non-applicable parameter input usages. Some entries are not applicable for the parameter inputs. For example, the input "mu0" must be a positive number, and then if the user sets "mu0= -1", the code will report an error with the message "the entry of "mu0" must be a number greater than zero". Thus, messages will appear when mistakes and inconsistencies are detected, and instructions about how to proceed to solve such problems will automatically appear.

Value

result	A table containing the main characteristics, conclusions concerning the acceptance or rejection of the null hypothesis, and the historical information from previous tests.
--------	---

Acknowledgements

Development of the `Analyze.Poisson` function was funded by: - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999.

See also

[AnalyzeSetUp.Poisson](#): for setting up sequential analysis with the `Analyze.Poisson` function, before the first look at the data.

[Performance.G.Poisson](#): for calculating the statistical power, expected time to signal and expected sample size for group sequential analysis with Poisson data.

[SampleSize.Poisson](#): for calculating the needed sample size to achieve the desired statistical

power for continuous sequential analysis with Poisson data.

[CV.G.Poisson](#): for calculating critical values for group sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva, Ned Lewis, Martin Kulldorff.

References

Fireman B, et al. (2013). Exact sequential analysis for Poisson data with time varying probabilities. Manuscript in preparation.

Jennison C, Turnbull B. (2000). Group Sequential Methods with Applications to Clinical Trials. London: Chapman and Hall/CRC.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, 30, 58–78.

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Silva IR, Kulldorff M. (2015). Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Examples

```
### Example. Four chunks of data.
```

```
### Firstly, it is necessary to set up the input parameters.
```

```
## Here we use the Wald type alpha spending.
```

```
## Note: cut off the "#" symbol before running the two lines below, and,  
## very important, choose an actual "address" to save your set up information.
```

```
# AnalyzeSetUp.Poisson(name="VaccineA", SampleSize=100, alpha=0.05,  
# M=1,AlphaSpendType="power-type",rho=0.5,title="n",  
# address="C:/Users/Ivair/Documents")
```

```
### Now we can sequentially apply a test for each one of three chunks of data.
```

```
# -----
```

```
## Test 1 - Situation where the expected number of events under H0 is equal to 2.
```

```
## This first test uses the default Wald type alpha spending (AlphaSpend="n").
```

```
## Note: cut off the "#" symbol before running the line below.
```

```
# Analyze.Poisson(test=1,mu0=2,events=1,AlphaSpend="n")
```

```
## Test 2 - Situation where the expected number of events under H0 is equal to 0.8.
```

```
## Observe that here we use an arbitrary alpha spending of 0.02.
```

```
## Note: cut off the "#" symbol before running the line below.
```

```
# Analyze.Poisson(test=2,mu0=0.8,events=2, AlphaSpend=0.02)
```

```
## Test 3 - Situation of elevated number of events, but now the
```

```
## arbitrary alpha spending is of 0.04.
```

```
## Note: cut off the "#" symbol before running the line below.
```

```
# Analyze.Poisson(test=3,mu0=9,events=10, AlphaSpend=0.04)
```

AnalyzeSetUp.Binomial *Function to Set Up the Input Parameters Before Using the Analyze.Binomial Function for the First Time.*

Description

The function `AnalyzeSetUp.Binomial` must be run ahead of `Analyze.Binomial` in order to set up the sequential analysis before the first group of data is analyzed. The function obtains the main parameter settings and performs basic calculations that are necessary for the subsequent sequential analysis.

Usage

```
AnalyzeSetUp.Binomial(name,N,alpha=0.05,zp="n",pp="n",M=1,
AlphaSpendType="Wald",rho="n", title="n",address="n")
```

Arguments

name	The name of the sequential analysis. Must be identical for all looks at the data, and the same as the name given in the subsequent calls to the <code>Analyze.Binomial</code> function. It cannot be the same as for another sequential analysis that is run simultaneously on the same computer. There is no default.
N	The maximum sample size, at which the sequential analysis stops without rejecting the null hypothesis. There is no default value.
alpha	The overall significance level. Must be in the range (0,0.5]. The default is "alpha=0.05".
zp	The prediction for z, the expected ratio between cases and controls under the null hypothesis that will be specified in the <code>Analyze.Binomial</code> function. This variable is only needed when <code>AlphaSpendType="Wald"</code> , and it is used to calculate the appropriate rejection boundary. If the z used in <code>Analyze.Binomial</code> during the actual sequential analysis is different from zp, that is okay, and the sequential analysis will still maintain the correct alpha level. The default value is "z=1".
pp	The prediction for p, the expected probability under the null hypothesis that will be specified in the <code>Analyze.Binomial</code> function. This variable is only needed when <code>AlphaSpendType="Wald"</code> , and it is used to calculate the appropriate rejection boundary. If the p used in <code>Analyze.Binomial</code> during the actual sequential analysis is different from pp, that is okay, and the sequential analysis will still maintain the correct alpha level. There is no default value.
M	The minimum number of events required before the null hypothesis can be rejected. It must be a positive integer. The default value is "M=1".
AlphaSpendType	The type of alpha spending function to be used. The options are <code>AlphaSpendType="Wald"</code> and <code>AlphaSpendType="power-type"</code> . With the 'Wald' option, the Wald type upper rejection boundary is used, which is flat with respect to the

likelihood ratio. With the power-type option, the alpha spending uses a power function with parameter rho, with rho defined by the user. This alpha spending setting is automatically used when the `Analyze.Binomial` function is run, but, during the sequential analysis, and before each test, the user can always specify an arbitrary amount of alpha spending to be used up until and including that test. See below for details.

rho	The parameter rho is used to build the target alpha spending function according to a power-type function. See below for details. It is not used for other alpha spending options. The variable rho must be a positive number. The default value is "rho=1".
title	Title for the results shown in the output tables and the illustrative graphics. It can be any text string. The default is that there is no title.
address	The address of the directory where the settings information of this sequential analysis is saved.

Details

The function `AnalyzeSetUp.Binomial` has to be executed once, but just once, to set up the general statistical characteristics of the intended sequential analysis, which is performed using the companion `Analyze.Binomial` function.

Sequential analysis methods are devoted to analyze data sets that accrue cumulatively over time, by conducting multiple statistical tests sequentially as more data accrues. In such a setting, it is important to carefully plan the sequential analysis before the first data arrives. For example, it is important to maintain certain analysis parameter values over time to avoid counting the same data twice, and to make sure that there are no changes in the past data that has already been included in a prior test. To avoid these kinds of problems, the `AnalyzeSetUp.Binomial` function is used to set the analysis parameters a priori and to create a place to save the data as it accumulates over time. At the time of each sequential test, this information is then automatically imported by the `Analyze.Binomial` function, to ensure the correct concatenation of old and new information.

At each test, the function `Analyze.Binomial` makes this concatenation automatically, but it will only work if the function `AnalyzeSetUp.Binomial` is executed before performing the very first test.

When running `AnalyzeSetUp.Binomial`, the user has the opportunity to choose the directory where the file with the general setup information and the historical data are to be saved. Alternatively, the user can click in the "Cancel" button, in which case this file will be saved in the temporary directory. Important: The location of this parameter and data file is saved in the temporary directory, so that directory cannot be cleaned until the sequential analysis has been completed. Each sequential analysis needs a different identifier, which is set using the "name" parameter. Once a name is chosen, it has to be written exactly the same way when running the function `Analyze.Binomial`.

`AnalyzeSetUp.Binomial` and `Analyze.Binomial` works for different types of alpha spending plans ($F(t)$). One option is to use the classical Wald type upper rejection boundary, which is flat with respect to the likelihood function. This is the same boundary used by the `CV.Binomial` and `CV.G.Binomial` functions. In order to use this boundary, one should pre-specify the binomial probability p under the null hypothesis, or, equivalently, the ratio $z = 1/p - 1$, which is the number of controls matched to each case in a matched analysis. For example, if the probability of having a case (instead of a control) is $p = 1/(1+z) = 0.5$, then we have "z=1" (1:1 matching ratio), and, if $p = 0.25$, we have "z=3" (1:3 matching ratio).

In `AnalyzeSetUp.Binomial`, the predicted z is specified (the input z_p), but if it turns out that the actual z is different, that is okay, since the actual z that is specified in `Analyze.Binomial` does not have to be the same as the predicted z_p that is specified in `AnalyzeSetUp.Binomial`. The latter is only used to set the alpha spending plan. The former, the actual z , is used to calculate the likelihood function which in turn determines whether the null hypothesis should be rejected or not. If the actual z is variable, so that it is different for different observations, we recommend setting the predicted z to be our best guess about the average of the actual z s. Alternatively, instead of z_p the user can specify pp , the best guess about the average of the actual ps . Note that only one of these parameters has to be specified, but if both are entered the code will only work if z_p and pp are such that $pp=1/(1+z_p)$. Otherwise, an error message will appear to remind that such condition must be complied.

Another alpha spending option is the power-type alpha spending plan (Kim and DeMetz 1987, p150; Jennison and Turnbull 2000, p148), with parameter ρ : $F(t) = \alpha * t^{\rho}$, where α is the overall significance level and t is a fraction of N , the maximum length of sequential analysis.

In addition to selecting the alpha spending plan, it is necessary to specify the overall alpha, or maximum Type I error probability, for the sequential analysis as a whole. It is also necessary to specify the maximum length of the sequential analysis, N , so that the sequential analysis stops without rejecting the null hypothesis when N total observations have been obtained.

Value

`inputSetUp` The `AnalyzeSetUp.Binomial` function creates a data.frame with the main information concerning the tuning parameterization for the planned surveillance and the historical information about the performed tests. The 'inputSetUp' data.frame is used by `Analyze.Binomial`, then it must be available when running `Analyze.Binomial`, but there is no need to manually look at it.

Acknowledgements

Development of the `AnalyzeSetUp.Binomial` function was funded by:

- Food and Drug Administration, Center for Drug Evaluation and Research, through the Mini-Sentinel Project (base version, documentation);
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (user defined alpha spending functions, improved documentation);

We thank Claudia Coronel-Moreno for valuable editorial support.

See also

[Analyze.Binomial](#): for running the sequential analysis that was set up using the `AnalyzeSetUp.Binomial` function.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

Jennison C, Turnbull B. (2000), Group Sequential Methods with Applications to Clinical Trials, *no. ISBN 0-8493-0316-8, London: Chapman and Hall/CRC.*

Kim K, DeMets DL. (1987), Design and Analysis of Group Sequential Tests Based on the Type I Error Spending Rate Function. *Biometrika*, **74**, n.1: 149–154.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Examples

```
# See example in the description of the Analyze.Binomial function.
```

AnalyzeSetUp.CondPoisson

Function to Set Up the Input Parameters Before Using the Analyze.CondPoisson Function for the First Time.

Description

The function `AnalyzeSetUp.CondPoisson` must be run ahead of `Analyze.CondPoisson` in order to set up the sequential analysis before the first group of data is analyzed. The function obtains the main parameter settings and performs basic calculations that are necessary for the subsequent sequential analysis.

Usage

```
AnalyzeSetUp.CondPoisson(name, SampleSizeType="Events", T="n", K="n", cc,
alpha=0.05, M=1, AlphaSpendType="Wald", rho="n", title="n", address="n")
```

Arguments

name	The name of the sequential analysis. Must be identical for all looks at the data, and the same as the name given in the subsequent calls to the <code>Analyze.CondPoisson</code> function. It cannot be the same as for another sequential analysis that is run simultaneously on the same computer. There is no default.
SampleSizeType	It is a string specifying the scale of the maximum sample size at which the sequential analysis stops without rejecting the null hypothesis. The only two possibilities are "SampleSizeType=PersonTimeRatio" or "SampleSizeType=Events". The default is "SampleSizeType=Events". See details.
T	Maximum sample size defined in the scale of the ratio between surveillance and historical person-time. This only produces effects when "SampleSizeType=PersonTimeRatio".

K	Maximum sample size defined in the scale of the number of events observed in the surveillance period. This only produces effects when "SampleSizeType=Events".
cc	Number of events observed in the historical period.
alpha	The overall significance level. Must be in the range (0,0.5]. The default is "alpha=0.05".
M	The minimum number of events required before the null hypothesis can be rejected. It must be a positive integer. The default value is "M=1".
AlphaSpendType	The type of alpha spending function to be used. The options are AlphaSpendType="Wald" and AlphaSpendType="power-type". With the 'Wald' option, the Wald type upper rejection boundary is used, which is flat with respect to the likelihood ratio. With the power-type option, the alpha spending uses a power function with parameter rho, with rho defined by the user. This alpha spending setting is automatically used when the Analyze.CondPoisson function is run, but, during the sequential analysis, and before each test, the user can always specify an arbitrary amount of alpha spending to be used up until and including that test. See below for details.
rho	The parameter rho is used to build the target alpha spending function according to a power-type function. See below for details. It is not used for other alpha spending options. The variable rho must be a positive number. The default value is "rho=1".
title	Title for the results shown in the output tables and the illustrative graphics. It can be any text string. The default is that there is no title.
address	The address of the directory where the settings information of this sequential analysis is saved.

Details

The function `AnalyzeSetUp.CondPoisson` has to be executed once, but just once, to set up the general statistical characteristics of the intended sequential analysis, which is performed using the companion `Analyze.CondPoisson` function.

Sequential analysis methods are devoted to analyze data sets that accrue cumulatively over time, by conducting multiple statistical tests sequentially as more data accrues. In such a setting, it is important to carefully plan the sequential analysis before the first data arrives. For example, it is important to maintain certain analysis parameter values over time to avoid counting the same data twice, and to make sure that there are no changes in the past data that has already been included in a prior test. To avoid these kinds of problems, the `AnalyzeSetUp.CondPoisson` function is used to set the analysis parameters a priori and to create a place to save the data as it accumulates over time. At the time of each sequential test, this information is then automatically imported by the `Analyze.CondPoisson` function, to ensure the correct concatenation of old and new information.

At each test, the function `Analyze.CondPoisson` makes this concatenation automatically, but it will only work if the function `AnalyzeSetUp.CondPoisson` is executed before performing the very first test.

When running `AnalyzeSetUp.CondPoisson`, the user has to choose the directory where the file with the general setup information and the historical data are to be saved. This step is mandatory and error messages are reported if a non-valid address is informed. Important: The location of this parameter and data file is saved in the temporary directory, so that directory cannot be cleaned until

the sequential analysis has been completed. Each sequential analysis needs a different identifier, which is set using the "name" parameter. Once a name is chosen, it has to be written exactly in the same way when running the function `Analyze.CondPoisson`.

`AnalyzeSetUp.CondPoisson` and `Analyze.CondPoisson` work for different types of alpha spending plans ($F(t)$). One option is to use the classical Wald type upper rejection boundary, which is flat with respect to the likelihood function. This is the same boundary used by the `CV.CondPoisson` function.

Another alpha spending option is the power-type alpha spending plan (Kim and DeMetz 1987, p150; Jennison and Turnbull 2000, p148), with parameter rho: $F(t) = \alpha * t^{\rho}$, where α is the overall significance level and t is a fraction of `SampleSize`, the maximum length of sequential analysis.

Attention is required for the input parameter "SampleSizeType". With this parameter, the user can choose the scale of the maximum sample size at which the surveillance is stopped without rejecting the null hypothesis. The idea of having two options for defining the scale of the maximum sample size, "SampleSizeType=PersonTimeRatio" or "SampleSizeType=Events", was introduced by Silva et al. (2016). With `SampleSizeType="PersonTimeRatio"`, the upper limit on the time of surveillance is given in the scale of the ratio between the cumulative person-time from the surveillance data up to the k th event, P_k in the notation of Silva et al. (2016), by the person-time from the historical data, which was denoted with V by Silva et al. (2016).

If `SampleSizeType="PersonTimeRatio"`, then the user has to inform a positive value for the input parameter "T". Usually, choices between 2 and 5 are adequate. With `SampleSizeType="Events"`, the upper limit is given in the scale of the number of events observed during the surveillance, and hence the user must specify a positive integer for the input parameter "K", such as e.g. "K=50" or "K=150". For more details on the exact calculations of critical values implemented in this package, and of all the other statistical performance measures also available, see the work of Silva et al. (2016).

Value

`inputSetUp` The `AnalyzeSetUp.CondPoisson` function creates a data.frame with the main information concerning the tuning parameterization for the planned surveillance and the historical information about the performed tests. The 'inputSetUp' data.frame is used by `Analyze.CondPoisson`, then it must be available when running `Analyze.CondPoisson`, but there is no need to manually look at it.

Acknowledgements

Development of the `AnalyzeSetUp.CondPoisson` function was funded by:
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999.
 - Foundation for Research Support of Minas Gerais State (FAPEMIG), MG, Brazil, through the grant Demanda Universal.

See also

[Analyze.CondPoisson](#): for running the sequential analysis that was set up using the `AnalyzeSetUp.CondPoisson` function.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

- Jennison C, Turnbull B. (2000), Group Sequential Methods with Applications to Clinical Trials, *no. ISBN 0-8493-0316-8, London: Chapman and Hall/CRC.*
- Kim K, DeMets DL. (1987), Design and Analysis of Group Sequential Tests Based on the Type I Error Spending Rate Function. *Biometrika*, **74**, n.1: 149–154.
- Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.
- KKulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.
- Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.
- Silva IR, Li L, Kulldorff M. (2016). Exact Conditional Sequential Testing for Poisson Data. Working paper.

Examples

```
# See example in the description of the Analyze.CondPoisson function.
```

AnalyzeSetUp.Poisson *Function to Set Up the Input Parameters Before Using the Analyze.Poisson Function for the First Time.*

Description

The function `AnalyzeSetUp.Poisson` must be run ahead of `Analyze.Poisson` in order to set up the sequential analysis before the first group of data is analyzed. The function obtains the main parameter settings and performs basic calculations that are necessary for the subsequent sequential analysis.

Usage

```
AnalyzeSetUp.Poisson(name, SampleSize, alpha=0.05, M=1,  
AlphaSpendType="Wald", rho="n", title="n", address="n")
```

Arguments

name	The name of the sequential analysis. Must be identical for all looks at the data, and the same as the name given in the subsequent calls to the <code>Analyze.Poisson</code> function. It cannot be the same as for another sequential analysis that is run simultaneously on the same computer. There is no default.
SampleSize	The maximum length of surveillance at which the sequential analysis stops without rejecting the null hypothesis. It is defined in terms of the expected sample size under the null hypothesis. There is no default value.
alpha	The overall significance level. Must be in the range (0,0.5]. The default is "alpha=0.05".
M	The minimum number of events required before the null hypothesis can be rejected. It must be a positive integer. The default value is "M=1".
AlphaSpendType	The type of alpha spending function to be used. The options are <code>AlphaSpendType="Wald"</code> and <code>AlphaSpendType="power-type"</code> . With the 'Wald' option, the Wald type upper rejection boundary is used, which is flat with respect to the likelihood ratio. With the power-type option, the alpha spending uses a power function with parameter rho, with rho defined by the user. This alpha spending setting is automatically used when the <code>Analyze.Poisson</code> function is run, but, during the sequential analysis, and before each test, the user can always specify an arbitrary amount of alpha spending to be used up until and including that test. See below for details.
rho	The parameter rho is used to build the target alpha spending function according to a power-type function. See below for details. It is not used for other alpha spending options. The variable rho must be a positive number. The default value is "rho=1".
title	Title for the results shown in the output tables and the illustrative graphics. It can be any text string. The default is that there is no title.
address	The address of the directory where the settings information of this sequential analysis is saved.

Details

The function `AnalyzeSetUp.Poisson` has to be executed once, but just once, to set up the general statistical characteristics of the intended sequential analysis, which is performed using the companion `Analyze.Poisson` function.

Sequential analysis methods are devoted to analyze data sets that accrue cumulatively over time, by conducting multiple statistical tests sequentially as more data accrues. In such a setting, it is important to carefully plan the sequential analysis before the first data arrives. For example, it is important to maintain certain analysis parameter values over time to avoid counting the same data twice, and to make sure that there are no changes in the past data that has already been included in a prior test. To avoid these kinds of problems, the `AnalyzeSetUp.Poisson` function is used to set the analysis parameters a priori and to create a place to save the data as it accumulates over time. At the time of each sequential test, this information is then automatically imported by the `Analyze.Poisson` function, to ensure the correct concatenation of old and new information.

At each test, the function `Analyze.Poisson` makes this concatenation automatically, but it will only work if the function `AnalyzeSetUp.Poisson` is executed before performing the very first test.

When running `AnalyzeSetUp.Poisson`, the user has to choose the directory where the file with the general setup information and the historical data are to be saved. This step is mandatory and error messages are reported if non-valid address is informed. Important: The location of this parameter and data file is saved in the temporary directory, so that directory cannot be cleaned until the sequential analysis has been completed. Each sequential analysis needs a different identifier, which is set using the "name" parameter. Once a name is chosen, it has to be written exactly the same way when running the function `Analyze.Poisson`.

`AnalyzeSetUp.Poisson` and `Analyze.Poisson` work for different types of alpha spending plans ($F(t)$). One option is to use the classical Wald type upper rejection boundary, which is flat with respect to the likelihood function. This is the same boundary used by the `CV.Poisson` and `CV.G.Poisson` functions.

Another alpha spending option is the power-type alpha spending plan (Kim and DeMetz 1987, p150; Jennison and Turnbull 2000, p148), with parameter rho: $F(t) = \alpha * t^{\rho}$, where α is the overall significance level and t is a fraction of `SampleSize`, the maximum length of sequential analysis.

Value

`inputSetUp` The `AnalyzeSetUp.Poisson` function creates a data.frame with the main information concerning the tuning parameterization for the planned surveillance and the historical information about the performed tests. The 'inputSetUp' data.frame is used by `Analyze.Poisson`, then it must be available when running `Analyze.Poisson`, but there is no need to manually look at it.

Acknowledgements

Development of the `AnalyzeSetUp.Poisson` function was funded by:
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999.

See also

[Analyze.Poisson](#): for running the sequential analysis that was set up using the `AnalyzeSetUp.Poisson` function.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

- Jennison C, Turnbull B. (2000), Group Sequential Methods with Applications to Clinical Trials, *no. ISBN 0-8493-0316-8, London: Chapman and Hall/CRC.*
- Kim K, DeMets DL. (1987), Design and Analysis of Group Sequential Tests Based on the Type I Error Spending Rate Function. *Biometrika*, **74**, n.1: 149–154.
- Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.
- Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Examples

See example in the description of the Analyze.Poisson function.

CV.Binomial	<i>Calculates Exact Critical Values for Continuous Sequential Analysis with Binomial Data.</i>
-------------	--

Description

The function CV.Binomial obtains critical values for the continuous sequential MaxSPRT test with binomial data, using a Wald-type upper boundary, which is flat with respect to the likelihood ratio function, and an pre-specified upper limit on the sample size.

Usage

```
CV.Binomial(N,alpha=0.05,M=1,z="n",p="n")
```

Arguments

N	The upper limit on the sample size (length of surveillance) expressed in terms of the total number of events (cases plus controls). "N" must be a positive integer. To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for "RR=1.1", "alpha=0.01" and "z=1", the statistical power is approximately 1 for "N>500". There is no default value.
alpha	The significance level. The "alpha" level must be in the range (0,0.5]. The default value is "alpha=0.05".
M	The minimum number of events needed before the null hypothesis can be rejected. "M" must be a positive integer, and the default value is "M=1".
z	For a matched case-control analysis, z is the number of controls matched to each case under the null hypothesis. There is no default value.
p	The probability of having a case under the null hypothesis. There is no default value.

Details

For the continuous binomial MaxSPRT, `CV.Binomial` calculates the upper boundary used to determine if the null hypothesis is to be rejected at each analysis. This is done for pre-specified values of the statistical significance level (α) and an upper limit on the sample size equal to N .

The input z represents the number of controls matched to each case. For example, if there are 3 controls matched to each case, " $z=3$ ". In a self-control analysis, z is the ratio of the control interval to the risk interval. For example, if the risk interval is 2 days long and the control interval is 7 days long, $z=7/2$. In terms of p , the binomial probability under the null hypothesis, $p=1/(1+z)$, or equivalently, $z=1/p-1$. The parameter z must be a positive number.

Alternatively, instead of z the user can specify p directly. Note that only one of these inputs, z or p , has to be specified, but if both are entered the code will only work if z and p are such that $p=1/(1+z)$. Otherwise, an error message will appear to remind that such condition must be complied.

For details about the algorithm used to calculate the critical value, see the paper by Kulldorff et al. (2011).

For some configurations of N and α , there is no critical value that gives a Type I error probability that is exactly equal to the requested " α ". This is because of the discrete nature of binomial data. In such situations, `CV.Binomial` returns the greatest critical value that guarantees a Type I error probability smaller than " α ". Thus the critical value for the binomial sequential analysis is conservative in this sense.

Value

<code>cv</code>	The critical value for a significance level equal to α . The largest conservative value is provided when it is not possible to have an Type I error exactly equal to α .
<code>Type_I_Error</code>	The exact Type I error probability given <code>cv</code> . Always less than or equal to α .

Acknowledgements

Development of the `CV.Binomial` function was funded by:

- Food and Drug Administration, Center for Drug Evaluation and Research, through the Mini-Sentinel Project; base version, documentation;
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999; code revisions, increased computational speed, improved documentation.

We thank Ron Berman, University of California, Berkeley, for a key suggestion to speed up the calculations, and Bruce Fireman for helpful discussions.

See also

[Analyze.Binomial](#): for performing sequential analysis with group, continuous or unpredictable sequential fashion.

Author(s)

Ivair Ramos Silva, Ned Lewis, Ron Berman, Martin Kulldorff.

References

- Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.
- Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Examples

```
## Critical value for continuous binomial sequential analysis with
# a maximum sample size of 20 events, requiring at
# least 3 events to reject the null, and with a significance level of 0.05:

CV.Binomial(N=20,alpha=0.05,M=3,z=1.1)
```

CV.CondPoisson	<i>Critical Values for continuous sequential CMaxSPRT for Poisson data with limited information from historical cohort.</i>
----------------	---

Description

The function `CV.CondPoisson` calculates critical values for the continuous sequential CMaxSPRT, using a Wald-type upper boundary, which is flat with respect to the likelihood ratio function, and a pre-specified upper limit on surveillance length.

Usage

```
CV.CondPoisson(Inference="exact", StopType="Cases", T="n", K="n", cc,
D=0, M=1, alpha=0.05)
```

Arguments

Inference	Inference='liberal', 'exact', or 'conservative' for the computation approach. Inference='liberal' for the liberal approach with possibly underestimated critical values and higher-than-nominal Type I error rate, Inference='exact' for the exact approach with exact critical values and nominal Type I error rates, Inference='conservative' for the conservative approach with possibly overestimated critical values and lower-than-nominal Type I error rates. The default is Inference="exact".
StopType	StopType='Tal' or 'Cases' for the type of surveillance length definition. With StopType='Tal', the maximum surveillance length (i.e., the upper limit) is defined in terms of the ratio of the cumulative person-time in the surveillance population divided by the total cumulative person-time in historical data, i.e., $P_k/V \leq T$; with StopType='Cases', the maximum surveillance length is defined in terms of the observed number of events in the surveillance population, i.e., $k \leq K$. The default is StopType="Cases"

cc	The total number of observed adverse events in the historical data. There is no default value.
K	The upper limit on length of surveillance expressed in terms of the observed number of events in the surveillance population, i.e., $k \leq K$. This argument K is used if and only if StopType='Cases'. There is no default value.
T	The upper limit on length of surveillance expressed in terms of the ratio of the cumulative person-time in the surveillance population divided by the total cumulative person-time in historical data, i.e., $P_k/V \leq T$. This argument T is used if and only if StopType='Tal'. There is no default value.
D	The minimum number for the ratio P_k/V before the null hypothesis can be rejected. This argument is used together with T . The default value is $D = 0$. A delayed start with $D > 0$ is recommended to avoid signaling very early on such that very little information would be available to judge whether the signal is more likely to be a true signal or chance finding.
M	The minimum number of events needed before the null hypothesis can be rejected. This argument is used together with K . A delayed start with $M > 1$ is recommended to avoid signaling very early on such that very little information would be available to judge whether the signal is more likely to be a true signal or chance finding. The default value is $M=1$.
alpha	The significance level, or the type 1 error probability, which is the probability of rejecting the null hypothesis when it is true. The alpha level must be in the range (0,0.5]. The default value is alpha=0.05.

Details

For continuous sequential analysis with CMaxSPRT for Poisson data and limited historical data, CV.CondPoisson calculates the critical value that constitutes the upper boundary used to determine if the null hypothesis should be rejected. This is done for pre-specified values of the statistical significance level (alpha) and an upper limit which can be defined based on either the observed number of events, "K", or the ratio "T" between the cumulative person-times in the surveillance population versus the historical data, as well as other parameter settings.

The test is one-sided, so that the null hypothesis is only rejected when there are more events than expected.

Following the results of Silva et al. (2016), the function offers three computation approaches which calculate liberal, exact, and conservative critical values respectively. When the upper limit is medium (e.g., $K = 50$) or large, the computational requirements for the exact approach can be high. The recommendation is to use the exact approach when the upper limit is small (e.g., $K = 10$), use the conservative approach when the upper limit is medium ($K = 50$) or large but cc is small, and use the liberal approach when cc is medium (e.g., 50) or large. Exact numerical results show that the three approaches yield very similar results when K and cc are reasonably large.

Value

Type_I_Error	The actual Type I error, for the exact approach. It equals the nominal level specified by the argument "alpha". For the liberal approach, the actual Type I error rate may be higher than the specified nominal level. For the conservative
--------------	---

approach, the actual Type I error rate may be lower than the specified nominal level.

cv The critical value for a significance level. For the exact approach, it is the exact critical value, for the liberal approach, it is the smallest liberal value, for the conservative approach, it is the largest conservative value.

Acknowledgements

Development of the CV.CondPoisson function was funded by:
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1, v2.0.2).

See also

[SampleSize.CondPoisson](#): calculating the upper limit with given Alpha, RR, and a desired power level for continuous CMaxSPRT.

[Performance.CondPoisson](#): calculating the statistical power, expected time to signal and expected time of analysis for continuous CMaxSPRT.

Author(s)

Ivair Ramos Silva, Lingling Li

References

Li L, Kulldorff M. (2010). A conditional maximized sequential probability ratio test for pharmacovigilance. *Statistics in Medicine*, 30;29(2):284-95.

Silva IR, Li L, Kulldorff M. (2016). Exact Conditional Sequential Testing for Poisson Data. Working paper.

Examples

```
# Calculates the exact critical value with upper limit of
# K=20 and a delayed start of a minimum of 2 cases, historical
# data has 20 events, and for a statistical significance level
# of 0.05.
# res<- CV.CondPoisson(Inference="exact", StopType="Cases",K=20,cc=20,
# M=2,alpha=0.05)

# which gives the results:
# res
# $Type_I_Error
# [1] 0.05
# $cv
# [1] 3.149115

# Calculates the liberal critical value with a upper limit of
```

```

# T=0.5 and a delayed start of D=0.1, i.e., the cumulative
# person-time in the surveillance population is at least
# one-tenth of the total cumulative person-time in historical
# data, historical data has 20 events, and for a statistical
# significance level of 0.05.
# res2<- CV.CondPoisson(Inference="liberal",StopType="Tal",T=0.5,cc=20,
# D=0.1,alpha=0.05)

# which gives the results:
# res2
# $Type_I_Error
# [1] 0.05
# $cv
# [1] 2.874993

```

CV.G.Binomial

Calculates Exact Critical Values for Group Sequential Analysis with Binomial Data.

Description

The function `CV.G.Binomial` calculates the critical values for the group sequential MaxSPRT test with Binomial data, using a Wald-type upper boundary, which is flat with respect to the likelihood ratio function, and with a pre-specified upper limit on the sample size (N).

Usage

```
CV.G.Binomial(N,alpha=0.05,M=1,z="n",p="n",GroupSizes)
```

Arguments

N	The upper limit on the sample size (length of surveillance) expressed in terms of the total number of events (cases plus controls). "N" must be a positive integer. To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for "RR=1.1" and "alpha=0.01" and "z=1", the statistical power is approximately 1 for "N>500". There is no default value.
alpha	The significance level. The "alpha" level must be in the range (0,0.5]. The default value is "alpha=0.05".
M	The minimum number of cases needed before the null hypothesis can be rejected. It must be a positive integer, and the default value is 'M=1'
z	For a matched case-control analysis, z is the number of controls matched to each case under the null hypothesis. There is no default value.
p	The probability of having a case under the null hypothesis. There is no default value.

GroupSizes Vector with the number of events (cases+controls) between two consecutive looks (tests) at the data, i.e, the group sizes. The length of this vector is equal to the maximum number of tests. The entries do not have to be the same, but they must sum up "N". If the group sizes is an integer instead of a vector, then that integer is the group size for all looks at the data, and the number of looks is "N/GroupSizes".

Details

For group sequential analysis with binomial data, the function `CV.G.Binomial` calculates critical value for the decision of accepting/rejecting the null hypothesis. This is done for a pre-defined significance level (α) and an upper limit on the sample size (N), as well as other parameter settings. When the length of 'GroupSizes' is equal to one, N must be a multiple of GroupSizes.

The input z represents the number of controls matched to each case. For example, if there are 3 controls matched to each case, " $z=3$ ". In a self-control analysis, z is the ratio of the control interval to the risk interval. For example, if the risk interval is 2 days long and the control interval is 7 days long, $z=7/2$. In terms of p , the binomial probability under the null hypothesis, $p=1/(1+z)$, or equivalently, $z=1/p-1$. The parameter z must be a positive number.

Alternatively, instead of z the user can specify p directly. Note that only one of these inputs, z or p , has to be specified, but if both are entered the code will only work if z and p are such that $p=1/(1+z)$. Otherwise, an error message will appear to remind that such condition must be complied.

For several configurations of N, GroupSizes and M, there are no critical values for a true Type I error probability strictly equal to α . This is because of the discrete nature of binomial data. For such cases, the function `CV.G.Binomial` returns the greatest critical value that will guarantee a type I error probability that is smaller than ' α ', so that the sequential analysis is conservative.

Value

cv The critical value for a significance level equal to α . The largest conservative value is provided when it is not possible to have an Type I error exactly equal to α .

Type_I_Error The exact Type I error probability given **cv**, and it is always less than or equal to α .

Acknowledgements

Development of the `CV.G.Binomial` function was funded by:

- Food and Drug Administration, Center for Drug Evaluation and Research, through the Mini-Sentinel Project (base version, documentation);
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (code revisions, increased computational speed, improved documentation).

See also

[SampleSize.Binomial](#): for calculating the minimum sample size given a target power in continuous sequential analysis with binomial data.

[Analyze.Binomial](#): for performing sequential analysis with a user defined alpha spending function.

[CV.Poisson](#): for calculating critical values for continuous sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva, Ned Lewis, Martin Kulldorff.

References

Fireman B, et al. (2013) Exact sequential analysis for binomial data with timevarying probabilities. Manuscript in Preparation.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
# Example 1:
result<- CV.G.Binomial(N=25,alpha=0.05,M=1,z=7/2,GroupSizes=5)
# if you type:
result
# then you will get the following output:
# [[1]]
# [1] 1.9852

# [[2]]
# [1] 0.04775995

# Example 2:

result<- CV.G.Binomial(N=50,alpha=0.05,M=1,z=7/2,GroupSizes=c(10,10,15,15))
cv<- as.numeric(result[1])
# if you type:
cv
# then you will get the following output:
# [1] 1.99202
```

Description

The function `CV.G.Poisson` calculates exact critical values for group sequential analysis with Poisson data, using a Wald type upper boundary, which is flat with respect to the likelihood ratio function, and with a pre-specified upper limit on the sample size.

Usage

```
CV.G.Poisson(SampleSize,alpha=0.05,GroupSizes,M=1)
```

Arguments

SampleSize	The upper limit on the sample size (length of surveillance) expressed in terms of the expected number of events under the null hypothesis. The "SampleSize" must be greater than 0. There is no default value.
M	The minimum number of events needed before the null hypothesis can be rejected. The default value is M=1. If there are frequent looks at the data, so that the group sizes are extremely small, a value of M=1 means that even a single event can reject the null hypothesis if it occurs sufficiently early. According to Kulldorff and Silva(2015), a reasonable choice is M=4.
alpha	The significance level, or the type 1 error probability, which is the probability of rejecting the null hypothesis when it is true. The alpha level must be in the range (0,0.5]. The default value is alpha=0.05.
GroupSizes	Vector containing the expected number of events under H0 for each test. The values must be positive numbers. The dimension of this vector must be equal to the maximum number of sequential tests. Thus, the sum of the entries in GroupSizes has to be equal to SampleSize. There is no default value.

Details

For group sequential analysis with Poisson data, CV.G.Poisson calculates the critical value that constitutes the upper boundary used to determine if the null hypothesis should be rejected. This is done for pre-specified values of the statistical significance level (alpha) and an upper limit on the sample size, determining the maximum length of surveillance, as well as other parameter settings. The test is one-sided, so that the null hypothesis is only rejected when there are more events than expected.

For several configurations of SampleSize, Looks and M there is no critical value that gives a probability of Type I error that is exactly equal to alpha. In such cases, the function CV.G.Poisson returns the largest critical value that will guarantee a type I error probability that is smaller than alpha, so that the sequential analysis is conservative.

For large values of the maximum SampleSize, such as 200 or more, the computational requirements can be high. To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for "RR=1.1" and "alpha=0.01", the statistical power is approximately 1 for a maximum sample size greater than 500.

Value

cv	The critical value for a significance level equal to alpha. The largest conservative value is provided when it is not possible to have a type I error that is exactly equal to alpha.
----	---

Acknowledgements

Development of the CV.G.Poisson function was funded by:

- Food and Drug Administration, Center for Biologics Evaluation and Research, through the Mini-Sentinel Post-Rapid Immunization Safety Monitoring (PRISM) program (v1.0);
- National Council of Scientific and Technological Development (CNPq), Brazil (v1.0);
- Bank for Development of the Minas Gerais State (BDMG), Brazil (v1.0);

- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1, v2.0.2).

See also

[Performance.G.Poisson](#): Calculates the statistical power, expected time to signal and expected sample size for group sequential analysis with Poisson data.

[CV.Poisson](#): Calculating critical values for continuous sequential analysis with Poisson data.

[CV.G.Binomial](#): Calculates critical values for group sequential analysis with binomial data.

Author(s)

Ivair Ramos Silva, Ned Lewis, Martin Kulldorff.

References

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
## Calculates the critical value for five group sequential looks, at 5, 11,  
## 17, 22 and 30 expected events under the null hypothesis, and for a  
## statistical signifi-  
## cance level of 0.05.
```

```
CV.G.Poisson(SampleSize=30,alpha=0.05,GroupSizes= c(5,6,6,5,8))
```

CV.Poisson

Critical Values for Continuous Sequential Analysis with Poisson Data.

Description

The function `CV.Poisson` obtains critical values for the continuous sequential MaxSPRT test with Poisson data, using a Wald type upper boundary, which is flat with respect to the likelihood ratio function, and with a pre-specified upper limit on the sample size.

Usage

```
CV.Poisson(SampleSize,D=0,M=1,alpha=0.05)
```

Arguments

SampleSize	The upper limit on the sample size (length of surveillance) expressed in terms of the expected number of events under the null hypothesis. The SampleSize must be greater than 0. To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for RR=1.1 and alpha=0.01, the statistical power is approximately 1 for a maximum sample size greater than 500. There is no default value.
D	The expected number of events under the null hypothesis before the first look at the data. The default is D=0, which is also the best choice. This means that there is no delay in the start of the sequential analyses. It is required that $D \leq \text{SampleSize}$.
M	The minimum number of events needed before the null hypothesis can be rejected. The default value is M=1, which means that even a single event can reject the null hypothesis if it occurs sufficiently early. A good rule of thumb is to set M=4 (Kulldorff and Silva, 2015).
alpha	The significance level, or the type 1 error probability, which is the probability of rejecting the null hypothesis when it is true. The alpha level must be in the range (0,0.5]. The default value is alpha=0.05.

Details

For the continuous sequential analysis with Poisson data, using the maximized sequential probability ratio test (MaxSPRT), CV.Poisson calculates the upper boundary used to determine if the null hypothesis should be rejected. This is done for pre-specified values on the statistical significance level (alpha) and the upper limit on the sample size, determining the maximum length of surveillance. The algorithm used to calculate the critical value is described by Kulldorff et al. (2011).

For some configurations of SampleSize, D and alpha, there is no critical value that gives a significance level that is exactly equal to the requested alpha. In such situations, CV.Poisson returns the greatest critical value that will guarantee an alpha level less than the alpha specified, so that sequential analysis is conservative.

For large values of SampleSize, such as 200 or more, the computational requirements can be high. To speed things up, the function will sometimes use one of two look-up tables that contain pre-calculated critical values for a pre-selected set of parameter values (TableCV.PoissonD and TableCV.PoissonM).

Value

cv	The critical value for a significance level equal to alpha. The largest conservative value is provided when it is not possible to have a Type I error exactly equal to alpha.
----	---

Acknowledgements

- Development of the CV.Poisson function was funded by:
- Food and Drug Administration, Center for Biologics Evaluation and Research, through the Mini-Sentinel Post-Rapid Immunization Safety Monitoring (PRISM) program (v1.0).
 - National Council of Scientific and Technological Development (CNPq), Brazil (v1.0).

- Bank for Development of the Minas Gerais State (BDMG), Brazil (v1.0).
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1, 2.0.2).

See also

[SampleSize.Poisson](#): for calculating the sample size needed for Continuous Sequential Analysis with Poisson Data.

[Performance.Poisson](#): for calculating the statistical power, expected time to signal and expected time of analysis for continuous sequential analysis with Poisson data.

[CV.G.Poisson](#): for calculating critical values for group sequential analysis with Poisson data.

[CV.Binomial](#): for calculating critical values in continuous sequential analysis with binomial data.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
## Calculates the critical value for continuous sequential analysis with
## a maximum sample size of ten expected cases under the null hypothesis,
## requiring at least 3 events to reject the null, and with a significance
## level of 0.05:
```

```
CV.Poisson(SampleSize=10,D=0,M=3,alpha=0.05)
```

```
## In this example, no critical value exist that will give the desired 0.05
## alpha level exactly. Instead, the function produces the critical value
## that makes the alpha as large as possible without exceeding 0.05.
```

```
CV.Poisson(SampleSize=3,D=1,M=1,alpha=0.05)
```

Description

The function `Performance.Binomial` calculates several performance metrics for the continuous binomial MaxSPRT for fixed upper limit on the sample size ("N"), minimum number of events required before rejecting the null hypothesis ("M"), critical value ("cv") and a relative risk ("RR"). The metrics calculated are the statistical power, the expected time to signal when the null hypothesis is rejected, and the expected sample size at the end of the analysis whether the null hypothesis was rejected or not.

Usage

```
Performance.Binomial(N,M=1,cv,z="n",p="n",RR=2)
```

Arguments

N	The upper limit on the sample size (length of surveillance) expressed in terms of the total number of events. "N" must be greater than 0. To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for "RR=1.1" and "alpha=0.01" and "z=1", the statistical power is approximately 1 for "N>500". There is no default value.
M	The minimum number of events needed before the null hypothesis can be rejected. The default value is 'M=1', which means that even a single event can reject the null hypothesis if it occurs sufficiently early. A good rule of thumb is to set 'M=4' (Kulldorff and Silva, 2015). It must be a positive integer.
cv	Critical value, defining the upper rejection boundary. The null hypothesis is rejected when the log-likelihood value is greater than "cv". The "cv" parameter is usually obtained by first running <code>CV.G.Binomial</code> . It must be a positive number, and there is no default.
z	For a matched case-control analysis, z is the number of controls matched to each case under the null hypothesis. There is no default value.
p	The probability of having a case under the null hypothesis. There is no default value.
RR	The relative risk (≥ 1) for which statistical power, expected signal time and expected length of surveillance are calculated. The default is $RR=2$.

Details

For the continuous Binomial MaxSPRT, the function `Performance.Binomial` calculates the statistical power, the expected time to signal when the null hypothesis is rejected, and the expected sample size until the analysis ends whether the null is rejected or not. When the null hypothesis is true, the probability of having a case, instead of a control, is $p = 1/(1 + z)$. But, if the null hypothesis is false, and the true relative risk is a value ' $RR > 1$ ', then the probability of having a case is $p = RR/(RR + z)$. If the user wants to calculate the exact Type I error probability for a given "cv", that can be done by setting "RR=1", in which case the power output value is the exact size of the test.

The input z represents the number of controls matched to each case. For example, if there are 3 controls matched to each case, "z=3". In a self-control analysis, z is the ratio of the control interval

to the risk interval. For example, if the risk interval is 2 days long and the control interval is 7 days long, $z=7/2$. In terms of p , the binomial probability under the null hypothesis, $p=1/(1+z)$, or equivalently, $z=1/p-1$. The parameter z must be a positive number.

Alternatively, instead of z the user can specify p directly. Note that only one of these inputs, z or p , has to be specified, but if both are entered the code will only work if z and p are such that $p=1/(1+z)$. Otherwise, an error message will appear to remind that such condition must be complied.

Value

Power	The statistical power.
ESignaltime	The expected time to signal given that the null hypothesis is rejected.
EsampleSize	The expected sample size when the analysis ends (length of surveillance) whether the null hypothesis was rejected or not.

Acknowledgements

Development of the Performance.Binomial function was funded by:
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999.

See also

[CV.Binomial](#): for calculating critical values in continuous sequential analysis with binomial data.
[SampleSize.Binomial](#): for calculating the minimum sample size given a target power in continuous sequential analysis with binomial data.
[Performance.G.Binomial](#): for calculating the statistical power, expected time to signal and expected sample size for group sequential analysis with binomial data.

Author(s)

Ivair Ramos Silva and Martin Kulldorff.

References

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, **30**: 58–78.
 Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
result<- Performance.Binomial(N=30,M=1,cv=2,z=1,RR=2)
# if you type:
result
# then you will get the following output:
# $power
# [1] 0.658732
```

```
# $signaltime
# [1] 10.7893

# $surveillancetime
# [1] 17.3453
```

Performance.CondPoisson

Statistical Power, Expected Signal Time and Sample Size for the continuous sequential CMaxSPRT for Poisson data with limited information from historical cohort.

Description

The function `Performance.CondPoisson` calculates several performance metrics for the continuous CMaxSPRT for selected computation approach, the type of upper limit definition and its value, critical value, number of historical data events, criteria of delayed start, the Type I error rate, and a relative risk. The metrics calculated are the statistical power, the expected time to signal when the null hypothesis is rejected, and the expected sample size at the end of the analysis whether the null hypothesis was rejected or not.

Usage

```
Performance.CondPoisson(Inference="exact", cv, StopType="Cases",
T="n", K="n", cc, D=0, M=1, RR=2)
```

Arguments

Inference	Inference='liberal', 'exact', or 'conservative' for the computation approach. Inference='liberal' for the liberal approach with possibly underestimated critical values and higher-than-nominal Type I error rate, Inference='exact' for the exact approach with exact critical values and nominal Type I error rates, Inference='conservative' for the conservative approach with possibly overestimated critical values and lower-than-nominal Type I error rates. The default is Inference='exact'.
cv	Critical value, defining the upper rejection boundary for the log-likelihood ratio test statistic. The null hypothesis is rejected when the log-likelihood value is greater than 'cv'. The 'cv' parameter is usually obtained by first running <code>CV.CondPoisson</code> . It must be a positive number, and there is no default.
StopType	StopType='Tal' or 'Cases' for the type of surveillance length definition. The default value is 'StopType=Cases'. See details.
T	The upper limit on length of surveillance expressed in terms of the ratio of the cumulative person-time in the surveillance population divided by the total cumulative person-time in historical data, i.e., $P_k/V \leq T$. This argument T is used if and only if StopType='Tal'. There is no default value.

K	The upper limit on length of surveillance expressed in terms of the observed number of events in the surveillance population, i.e., $k \leq K$. This argument K is used if and only if <code>StopType='Cases'</code> . There is no default value.
cc	The total number of observed adverse events in the historical data. There is no default value.
D	The minimum number for the ratio P_k/V before the null hypothesis can be rejected. This argument is used together with <code>StopType='Tal'</code> . The default value is $D = 0$.
M	The minimum number of events needed before the null hypothesis can be rejected. This argument is used together with <code>StopType='Cases'</code> . The default value is $M=1$.
RR	The target relative risk for which statistical power, expected signal time and expected length of surveillance are calculated. The default is ' <code>RR=2</code> '.

Details

For continuous sequential analysis with Poisson data with limited historical information, the `Performance.CondPoisson` function calculates the statistical power, the expected time to signal when the null hypothesis is rejected and the expected sample size until the analysis ends whether the null is rejected or not. The sample size (i.e., the upper limit) can be expressed either in terms of the ratio "T" of the cumulative person-time in the surveillance population divided by the total cumulative person-time in historical data (`StopType="Tal"`), i.e., $P_k/V \leq T$, or in terms of the observed number of events "K" in the surveillance population (`StopType="Cases"`), i.e., $k \leq K$. Large values of the `SampleSize`, greater than say 1000, may leads to long computing times. When the statistical power is close to 1, then the expected time to signal will be very close to the expected sample size.

For the parameter of delayed start, "D", a delayed start with $D > 0$ is recommended to avoid signaling very early on such that very little information would be available to judge whether the signal is more likely to be a true signal or chance finding. Similarly, if the delayed start is defined in terms of the number of events, "M", a setting such that $M > 1$ is recommended to avoid signaling very early on such that very little information would be available to judge whether the signal is more likely to be a true signal or chance finding.

Value

Power	The statistical power.
ESignalTime	The expected time to signal given that the null hypothesis is rejected.
ESampleSize	The expected sample size when the sequential analysis ends (length of surveillance) whether the null hypothesis was rejected or not.

Acknowledgements

Development of the `Performance.CondPoisson` function was funded by:
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1, v2.0.2). - Foundation for Research Support of Minas Gerais State (FAPEMIG), MG, Brazil, through the grant Demanda Universal.

See also

[CV.CondPoisson](#): calculating the critical value for continuous CMaxSPRT.

[SampleSize.CondPoisson](#): calculating the sample size for continuous CMaxSPRT.

Author(s)

Ivair Ramos Silva, Lingling Li

References

Li L, Kulldorff M. (2010). A conditional maximized sequential probability ratio test for pharmaco-vigilance. *Statistics in Medicine*, 30;29(2):284–95.

Silva IR, Li L, Kulldorff M. (2016). Exact conditional maximized sequential probability ratio test adjusting for covariates. Working paper.

Examples

```
## First calculate the critical value with upper limit defined in terms of
## the number of observed events in surveillance population (K=50), with 50
## events in historical data, no delayed start, and alpha=0.05:
# res<-CV.CondPoisson(Inference="exact",StopType="Cases",K=20,cc=50,M=1,
# alpha=0.05)

# cvt<- res[[2]]

## calculate the performance using the critical value 'cvt' from the previous
## step, under RR=1.5:
#Performance.CondPoisson(Inference="exact",cv=cvt,StopType="Cases",K=20,cc=50,
# M=1,RR=1.5)
```

Performance.G.Binomial

*Statistical Power, Expected Signal Time and Expected Length of
Surveillance for Group Sequential Analysis for Binomial Data.*

Description

The function `Performance.G.Binomial` calculates several performance metrics for the group binomial MaxSPRT for fixed upper limit on the sample size ("N"), critical value ("cv"), maximum number of tests to be performed, minimum number of events required before rejecting the null hypothesis (M), and true relative risk (RR). The metrics calculated are the statistical power, the expected time to signal when the null hypothesis is rejected and the expected sample size at the end of the analysis whether the null hypothesis was rejected or not.

Usage

```
Performance.G.Binomial(N,M=1,cv,z="n",p="n",RR=2,GroupSizes)
```

Arguments

N	The upper limit on the sample size (length of surveillance) expressed in terms of the total number of events. "N" must be a positive integer. To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for "RR=1.1" and "alpha=0.01" and "z=1", the statistical power is approximately 1 for "N>500". There is no default value.
M	The minimum number of events needed before the null hypothesis can be rejected. "M" must be a positive integer, and the default value is M=1.
cv	Critical value, defining the upper rejection boundary. The null hypothesis is rejected when the log-likelihood value is greater than "cv". The "cv" parameter is usually obtained by first running CV.Binomial. It must be a positive number, and there is no default.
z	For a matched case-control analysis, z is the number of controls matched to each case under the null hypothesis. There is no default value.
p	The probability of having a case under the null hypothesis. There is no default value.
RR	The relative risk under the alternative hypothesis. If p and q are binomial probabilities under the null and the alternative hypotheses, respectively, then "RR=(1/p-1)/(1/q-1)".
GroupSizes	Vector with the number of events (cases+controls) between two consecutive looks (tests) at the data, i.e, the group sizes. The length of this vector is equal to the maximum number of looks. The entries do not have the same, but they sum up to N. If the group sizes is an integer instead of a vector, then that integer is the group size for all looks at the data, and the number of looks is "N/GroupSizes".

Details

The function `Performance.G.Binomial` calculates the statistical power, the expected time to signal when the null hypothesis is rejected and the expected sample size by using the group sequential MaxSPRT for binomial data given a fixed relative risk (RR). When the null hypothesis is true, the probability of having a case, instead of a control, is $p = 1/(1+z)$. But, if the null hypothesis is false, and the true relative risk is a value $RR > 1$, then the probability of having a case is $p = RR/(RR+z)$.

If the user wants to calculate the actual overall size (or the maximum Type I error probability) of the test, that can be done by setting $RR=1$, in which case the power output value is the test size. This is made for a given flat critical value, an upper limit on the sample size (N), and a delayed start in the vigilance given by M.

The input z represents the number of controls matched to each case. For example, if there are 3 controls matched to each case, "z=3". In a self-control analysis, z is the ratio of the control interval to the risk interval. For example, if the risk interval is 2 days long and the control interval is 7 days long, $z=7/2$. In terms of p, the binomial probability under the null hypothesis, $p=1/(1+z)$, or equivalently, $z=1/p-1$. The parameter z must be a positive number.

Alternatively, instead of z the user can specify p directly. Note that only one of these inputs, z or p , has to be specified, but if both are entered the code will only work if z and p are such that $p=1/(1+z)$. Otherwise, an error message will appear to remind that such condition must be complied.

Value

Power	Statistical power.
ESignalTime	The expected time to signal given that the null hypothesis was rejected.
ESampleSize	The expected sample size when the null was rejected or not.

Acknowledgements

Development of the Performance.G.Binomial function was funded by:
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999.

See also

[CV.G.Binomial](#): for calculating critical values in group sequential analysis with binomial data.
[Performance.Binomial](#): for calculating the statistical power, expected time to signal and expected time of analysis for continuous sequential analysis with binomial data.
[Analyze.Binomial](#): for performing sequential analysis with group, continuous or unpredictable sequential fashion.
[CV.Poisson](#): for calculating critical values for continuous sequential analysis with Poisson data.
[Performance.G.Poisson](#): for calculating statistical power, expected time to signal and expected length of surveillance for group sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva and Martin Kulldorff

References

Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Examples

```
result<- Performance.G.Binomial(N=40,M=1,cv=2.5,z=1,RR=2,GroupSizes=c(2))
# if you type:
result
# then you will get the following output:
# $Power
# [1] 0.6594118

# $`ESignalTime`
# [1] 17.18626

# $`ESampleSize`
# [1] 24.95635
```

Performance.G.Poisson *Calculates Statistical Power, Expected Time to Signal and Expected Sample Size for Group Sequential Analysis with Poisson Data.*

Description

The Performance.G.Poisson function calculates statistical power, expected time to signal and the expected sample size for group sequential analysis with Poisson data. It is required that the number of looks at the data is pre-specified, and equally spaced in terms of the sample size. This function can also be used to calculate the statistical significance level alpha, by setting RR=1, in which case the power output value is the alpha level.

Usage

```
Performance.G.Poisson(SampleSize,cv,GroupSizes,M=1,RR=2)
```

Arguments

SampleSize	The upper limit on the sample size (length of surveillance) expressed in terms of the expected number of events under the null hypothesis. SampleSize must be greater than 0. There is no default value.
cv	The critical value for a group sequential analysis with Poisson data. This is calculated using the CV.G.Poisson . There is no default value.
GroupSizes	Vector containing the expected number of events under H0 for each test. The values must be positive numbers. The dimension of this vector must be equal to the maximum number of sequential tests. Thus, the sum of the entries in GroupSizes has to be equal to SampleSize. There is no default value.
M	The minimum number of events needed before the null hypothesis can be rejected. M must be a positive integer, and the default value is M=1. According to Kulldorff and Silva (2015) , a reasonable choice is M=4.
RR	The relative risk under the alternative hypothesis. Can be any number greater or equal to 1. The default value is RR=2.

Details

The function Performance.G.Poisson calculates the statistical power, the expected time to signal when the null hypothesis is rejected, and the expected sample size at the end of the sequential analysis, whether the null hypothesis was rejected or not. The user specifies the upper limit on the sample size (SampleSize), the upper boundary critical value (cv), a vector with the expected number of cases under H0 at each test (GroupSizes), the minimum number of events required before rejecting the null hypothesis (M) and the relative risk under the alternative hypothesis (RR).

The calculations are done using an upper rejection boundary that is flat with respect to the log likelihood ratio, as expressed using the critical value. The different looks at the data, or the different tests, are assumed to be equally spaced with respect to the accrual of the sample size.

To avoid very large computation times, we suggest not using `SampleSize` values greater than 1000. Typically, this is not a major restriction. For example, for $RR=1.1$ and $\alpha=0.01$, the statistical power is approximately 1 for a maximum sample size greater than 500.

Value

<code>Power</code>	The statistical power.
<code>ESignalTime</code>	The expected time to signal given that the null hypothesis is rejected.
<code>ESampleSize</code>	The expected sample size when the sequential analysis ends (length of surveillance), whether the null hypothesis is rejected or not.

Acknowledgements

Development of the `Performance.G.Poisson` function was funded by:

- Food and Drug Administration, Center for Biologics Evaluation and Research, through the Mini-Sentinel Post-Rapid Immunization Safety Monitoring (PRISM) program (v1.0).
- National Council of Scientific and Technological Development (CNPq), Brazil (v1.0).
- Bank for Development of the Minas Gerais State (BDMG), Brazil (v1.0).
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1,2.0.2).

See also

[CV.G.Poisson](#): Calculates critical values for group sequential analysis with Poisson data.

[Performance.Poisson](#): Calculates the statistical power, expected time to signal and expected sample size for continuous sequential analysis with Poisson data.

[Performance.G.Binomial](#): Calculates the statistical power, expected time to signal and expected sample size for group sequential analysis with binomial data.

Author(s)

Ivair Ramos Silva and Martin Kulldorff

References

Silva IR, Kulldorff M. (2015), Continuous versus Group Sequential Analysis for Vaccine and Drug Safety Surveillance. *Biometrics*, 71 (3), 851–858.

Examples

```
## First use the CV.G.Poisson function to calculate the critical value for
## 5 sequential looks at the data, spaced six units apart, and with a
## statistical significance level of 0.05:
```

```
cvt<- CV.G.Poisson(SampleSize=30,alpha=0.05,GroupSizes=c(6,6,6,6,6))
```

```
## For an alternative hypothesis of a relative risk of RR=1.5, calculates the
## statistical power, the expected time to signal, and the expected sample size
## at the end of the sequential analysis.
```

```
(Performance.G.Poisson(SampleSize=30,cv=cvt,GroupSizes=c(6,6,6,6,6),RR=1.5))
```

Performance.Poisson	<i>Calculates Statistical Power, Expected Time to Signal and Expected Sample Size for Continuous Sequential Analysis with Poisson Data.</i>
---------------------	---

Description

The Performance.Poisson function calculates three different performance metrics for the continuous sequential analysis with Poisson data: the statistical power, the expected time to signal when the null hypothesis is rejected and the expected sample size at the end of the analysis whether the null hypothesis was rejected or not. The user specifies the relative risk under the alternative hypothesis (RR), as well as the sequential analysis parameters. To calculate the statistical significance level alpha, $RR=1$, in which case the power output value is the alpha level.

Usage

```
Performance.Poisson(SampleSize, D = 0, M = 1, cv, RR = 2)
```

Arguments

SampleSize	The upper limit on the sample size (length of surveillance) expressed in terms of the expected number of events under the null hypothesis. The SampleSize must be greater than 0. There is no default value.
D	The expected number of events under the null hypothesis at the first look at the data. The default is $D=0$, which is also the best choice. This means that there is no delay in the start of the sequential analysis. It is required that $D \leq \text{SampleSize}$.
M	The minimum number of events needed to be observed before the null hypothesis can be rejected. The default is $M=1$, which means that even a single event can reject the null hypothesis if it occurs sufficiently early. A good rule of thumb is to set $M=4$ (Kulldorff and Silva, 2015).
cv	The critical value constituting the upper rejection boundary. This can be calculated using the CV.Poisson function.
RR	The relative risk under the alternative hypothesis. It is required that $RR \geq 1$. The default value is $RR=2$.

Details

For continuous sequential analysis with Poisson data, the Performance.Poisson function calculates the statistical power, the expected time to signal when the null hypothesis is rejected and the expected sample size until the analysis ends whether the null is rejected or not. The sample size is expressed in terms of the expected number of events under the null hypothesis. Large values of the SampleSize, greater than say 1000, may leads to long computing times. When the statistical power is close to 1, then the expected time to signal will be very close to the expected sample size, since both are measured in information time, using the expected events under the null hypothesis as the unit.

To avoid very large computation times, we suggest not using values greater than 1000. Typically, this is not a major restriction. For example, for $RR=1.1$ and $\alpha=0.01$, the statistical power is approximately 1 for a maximum sample size greater than 500.

Value

Power	The statistical power.
ESignalTime	The expected time to signal given that the null hypothesis is rejected.
ESampleSize	The expected sample size when the sequential analysis ends (length of surveillance) whether the null hypothesis was rejected or not.

Acknowledgements

Development of the Performance.Poisson function was funded by:

- Food and Drug Administration, Center for Biologics Evaluation and Research, through the Mini-Sentinel Post-Rapid Immunization Safety Monitoring (PRISM) program (v1.0).
- National Council of Scientific and Technological Development (CNPq), Brazil (v1.0).
- Bank for Development of the Minas Gerais State (BDMG), Brazil (v1.0).
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1,2.0.2).

See also

[CV.Poisson](#): Calculates critical values for continuous sequential analysis with Poisson data.

[SampleSize.Poisson](#): Sample size calculations for continuous sequential analysis with Poisson data.

[Performance.G.Poisson](#): Calculates the statistical power, expected time to signal and expected sample size for group sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva and Martin Kulldorff

References

- Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance. *Sequential Analysis*, 30: 58–78. Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
## Suppose we want to find the statistical power to detect a relative risk
## of 2 when doing up to at most 20 months of surveillance, as well as the
## expected time to signal when the null hypothesis is rejected. During
## each month, we expected to see 0.5 events if the null hypothesis is true.
## This means that the upper limit on the sample size is 20*0.5=10 expected
## events under the null hypothesis. We will then first calculate the critical
## value for an upper limit on the sample size equal to 10 and a significance
## level of alpha=0.05:

# cvt<- CV.Poisson(SampleSize=10,alpha=0.05)
# cvt
# [1] 3.467952
```



```

## After that, we use the Performance.Poisson function to calculate the
## power and the expected time to signal when the null hypothesis is
## rejected for the alternative hypothesis with a relative risk equal to 2:

## Power, expected signal time and expected sample size for a relative risk
## equal to 2:
# Performance.Poisson(SampleSize=10,cv=cvt,RR=2)
#      Power      ESignalTime      ESampleSize
# [1,] 0.6850634      4.130985      5.979353)

## From the results, we see that the statistical power is 68.5%. When the null
## is rejected, the expected time to signal is 4.13 in the unit of events
## expected under the null. If data is collected uniformly over time at the
## rate of 0.5 expected counts per month, the expected time to signal is
## 4.13/0.5= 8.26 months.

## The above calculations can also be accomplished using one single command line:

# Performance.Poisson(SampleSize=10,cv=CV.Poisson(SampleSize=10,alpha=0.05),RR=2)
#      Power      ESignalTime      ESampleSize
# [1,] 0.6850634      4.130985      5.979353

```

SampleSize.Binomial *Sample Size Calculation for Continuous Sequential Analysis with Binomial Data.*

Description

The function `SampleSize.Binomial` obtains the sample size needed to guarantee a desired statistical power, for a fixed true relative risk, when doing continuous sequential analysis for binomial data with a Wald-type upper boundary, which is flat with respect to the log-likelihood ratio. It can also be used to approximate the sample size needed when doing group sequential analysis for binomial data.

Usage

```
SampleSize.Binomial(RR,alpha=0.05,power=0.9,M=1,z="n",p="n")
```

Arguments

RR	A target vector of relative risks to be detected with the requested statistical powers.
alpha	The significance level. The default value is "alpha=0.05". Must be in the range (0, 0.5].
power	The target vector of overall statistical powers to detect an increased risk of the relative risk (RR). The default value is "power=0.90".

M	The minimum number of events needed before the null hypothesis can be rejected. It must be a positive integer. The default value is "M=1".
z	For a matched case-control analysis, z is the number of controls matched to each case under the null hypothesis. There is no default value.
p	The probability of having a case under the null hypothesis. There is no default value.

Details

The function `SampleSize.Binomial` calculates the sample size N , to be used for the continuous binomial MaxSPRT in order to provide the desired statistical power for a user-specified relative risk RR . The required sample size (`Required_N`) is expressed in terms of the total number of observations, and it is the number of observations by which the sequential analysis will end without rejected the null hypothesis. The solution is exact using iterative numerical calculations.

The required sample size, N , increases for increasing values of power, while N decreases for increasing values of α , the relative risk RR and the minimum number of events needed to signal M . For increasing values of z , the required sample size N can either decrease or increase.

While this function calculates the required sample size for continuous sequential analysis, it can also be used as an approximation for group sequential analyses. With the same `Required_N`, and all other parameters being the same, a group sequential analysis will always give higher statistical power than a continuous sequential analysis, so `SampleSize.Binomial` can be use to ensure the required statistical power for group sequential analyses.

The input z represents the number of controls matched to each case. For example, if there are 3 controls matched to each case, " $z=3$ ". In a self-control analysis, z is the ratio of the control interval to the risk interval. For example, if the risk interval is 2 days long and the control interval is 7 days long, $z=7/2$. In terms of p , the binomial probability under the null hypothesis, $p=1/(1+z)$, or equivalently, $z=1/p-1$. The parameter z must be a positive number.

Alternatively, instead of z the user can specify p directly. Note that only one of these inputs, z or p , has to be specified, but if both are entered the code will only work if z and p are such that $p=1/(1+z)$. Otherwise, an error message will appear to remind that such condition must be complied.

Value

`SampleSize_by_RR_Power`

A table containing the main performance measures associated to the required samples sizes for each combination of RR and power. The sample size N is provided in terms of the total number of observations. In the case-control setting, this is equal to the total number of cases and controls. In the self-control setting, it is equal to the total number of events in either the risk or the control interval.

Acknowledgements

Development of the `SampleSize.Binomial` function was funded by:

- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999.

See also

[CV.Binomial](#): for calculating critical values in continuous sequential analysis with binomial data.
[Performance.Binomial](#): for calculating the statistical power, expected time to signal and expected time of analysis for continuous sequential analysis with binomial data.
[SampleSize.Poisson](#): sample size calculation for continuous sequential analysis with Poisson data.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
result<- SampleSize.Binomial(RR=5,alpha=0.01,power=0.88,M=1,z=2)
# if you type:
result
# then you will get the following output:
# $Required_N
# [1] 25

# $cv
# [1] 4.59581

# $Type_I_Error
# [1] 0.009755004

# $Actual_power
# [1] 0.8855869
```

SampleSize.CondPoisson

Sample size calculation for the continuous sequential CMaxSPRT for Poisson data with limited information from historical cohort.

Description

The function `SampleSize.CondPoisson` obtains the required sample size (length of surveillance) needed to guarantee a desired statistical power for a pre-specified relative risk, when doing continuous sequential CMaxSPRT, using a Wald-type upper boundary, which is flat with respect to the likelihood ratio function.

Usage

```
SampleSize.CondPoisson(cc,D=0,M=1,alpha=0.05,power=0.9,RR=2)
```

Arguments

cc	The total number of observed adverse events in the historical data. There is no default.
D	The minimum number for the ratio P_k/V before the null hypothesis can be rejected. The default value is $D = 0$. A delayed start with $D > 0$ is recommended to avoid signaling very early on such that very little information would be available to judge whether the signal is more likely to be a true signal or chance finding.
M	The minimum number of events needed before the null hypothesis can be rejected. The default value is $M=1$. A delayed start with $M > 1$ is recommended to avoid signaling very early on such that very little information would be available to judge whether the signal is more likely to be a true signal or chance finding.
alpha	The significance level. It must be in the range (0,0.5]. The default value is $\alpha=0.05$.
power	The target vector of overall statistical powers to detect an increased relative risk (RR). The default value is $\text{power}=0.90$.
RR	The target vector of relative risks to be detected with the requested vector of statistical powers. The default value is $\text{RR}=2$.

Details

When using the CMaxSPRT and the CV.CondPoisson function to conduct continuous sequential analysis for Poisson data and limited historical data, the null hypothesis is rejected when the log likelihood ratio exceeds the predetermined critical value calculated by CV.CondPoisson. The sequential analysis ends without rejecting the null hypothesis when a predetermined upper limit on the sample size is reached, expressed either in terms of the ratio of the cumulative person-time in the surveillance population divided by the total cumulative person-time in historical data (Stop-Type="Tal"), or in terms of the observed number of events in the surveillance population (Stop-Type="Cases"). For example, the sequential analysis may end as soon as the sample size is such that the cumulative person-time in the surveillance population is twice the cumulative person-time in historical data, or there are 50 observed events in the surveillance population. The function SampleSize.CondPoisson calculates what the upper limit on the sample size (length of surveillance) that is required for the continuous CMaxSPRT to achieve the desired statistical power for a pre-specified relative risk RR. It outputs the upper limit on sample size for both definitions of the surveillance length, one expressed in terms of the ratio of the cumulative person-time in the surveillance population divided by the total cumulative person-time in historical data (T), and the other one expressed in terms of the observed number of events in the surveillance population (K). To save computing time, the liberal computation approach proposed by Silva et al. (2016) is used in SampleSize.CondPoisson to find the solution.

Value

SampleSize_by_RR_Power

A table containing the main performance measures associated to the required samples sizes, expressed in the scale of the number of events in the surveillance period, for each combination of RR and power.

Acknowledgements

Development of the SampleSize.CondPoisson function was funded by:
 - National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1, v2.0.2). - Foundation for Research Support of Minas Gerais State (FAPEMIG), MG, Brazil, through the grant Demanda Universal.

See also

[CV.CondPoisson](#): calculating the critical value for continuous CMaxSPRT.

[Performance.CondPoisson](#): calculating the statistical power, expected time to signal and expected time of analysis for continuous CMaxSPRT.

Author(s)

Ivair Ramos Silva, Lingling Li

References

Li L, Kulldorff M. (2010). A conditional maximized sequential probability ratio test for pharmacovigilance. *Statistics in Medicine*, 30;29(2):284-95.

Silva IR, Li L, Kulldorff M. (2016). Exact Conditional Sequential Testing for Poisson Data. Working paper.

Examples

```
# Sample size required to obtain a power of 90%, for a relative risk of 1.5,
# no delay for starting the surveillance (D=0), under an alpha level of 5%,
# with 5 events in the historical data.

# res<- SampleSize.CondPoisson(cc=50,D=0,M=1,alpha=0.05,power=0.9,RR=2)

# If we write:
# res
# then we get:

# $K
# [1] 66

# $Tal
# [1] 0.7890625
```

```
# $cv
# [1] 3.626436
```

SampleSize.Poisson *Sample Size Calculation for Continuous Sequential Analysis with Poisson Data.*

Description

The function `SampleSize.Poisson` obtains the required sample size (length of surveillance) needed to guarantee a desired statistical power for a pre-specified relative risk, when doing continuous sequential analysis for Poisson data with a Wald type upper boundary, which is flat with respect to the log-likelihood ratio. It can also be used to approximate the sample size needed when doing group sequential analysis for Poisson data.

Usage

```
SampleSize.Poisson(alpha=0.05,power=0.9,M=1,D=0,RR=2,
precision=0.000001)
```

Arguments

alpha	The significance level. The default value is $\alpha=0.05$. Must be in the range (0,0.5].
RR	The target vector of relative risks to be detected with the requested statistical vector of powers. The default value is $RR=2$.
power	The target vector of overall statistical powers to detect an increased relative risk (RR). The default value is $power=0.90$.
M	The minimum number of events needed before the null hypothesis can be rejected. It must be a positive integer. A good rule of thumb is to set $M=4$ (Kull-dorff and Silva, 2015). The default value is $M=1$, which means that even a single event can reject the null hypothesis if it occurs sufficiently early.
D	The expected number of events under the null hypothesis at the first look at the data. This is used when there is an initial large chunk of data arriving, followed by continuous sequential analysis. The default value is $D=0$, which is also the best choice. This means that there is no delay in the start of the sequential analyses. If D is very large, the maximum sample size will be set equal to D if a non-sequential analysis provides the desired power.
precision	The tolerance for the difference between the requested and actual statistical power. Should be very small. The default value is $precision=0.000001$.

Details

When using the `MaxSPRT` and the `CV.Poisson` function to conduct continuous sequential analysis for Poisson data, the null hypothesis is rejected when the log likelihood ratio exceeds the pre-determined critical value calculated by `CV.Poisson`. The sequential analysis ends without rejecting the null hypothesis when a predetermined upper limit on the sample size is reached, expressed in terms of the expected number of events under the null hypothesis. For example, the sequential analysis may end as soon as the sample size is such that there are 50 expected events under the null.

The function `SampleSize.Poisson` calculates what the upper limit on the sample size (length of surveillance) that is required for the continuous Poisson based `MaxSPRT` to achieve the desired statistical power for a pre-specified relative risk `RR`. The solution is exact using iterative numerical calculations (Kulldorff et al., (2011).

While designed for continuous sequential analysis, the `SampleSize.Poisson` function can also be used to approximate the required upper limit on the sample size that is needed when doing group sequential analysis for Poisson data, using the `CV.G.Poisson` function.

Value

`SampleSize_by_RR_Power`

A table containing the main performance measures associated to the required samples sizes, expressed in terms of the expected number of events under the null hypothesis, for each combination of `RR` and power.

Acknowledgements

Development of the `SampleSize.Poisson` function was funded by:

- National Council of Scientific and Technological Development (CNPq), Brazil (v1.0).
- Bank for Development of the Minas Gerais State (BDMG), Brazil (v1.0).
- National Institute of General Medical Sciences, NIH, USA, through grant number R01GM108999 (v2.0.1,2.0.2).

See also

[CV.Poisson](#): for calculating critical values for continuous sequential analysis with Poisson data.

[CV.G.Poisson](#): for calculating critical values for group sequential analysis with Poisson data.

[Performance.Poisson](#): for calculating the statistical power, expected time to signal and expected sample size for continuous sequential analysis with Poisson data

[SampleSize.Binomial](#): for calculating the minimum sample size given a target power in continuous sequential analysis with binomial data.

Author(s)

Ivair Ramos Silva, Martin Kulldorff.

References

- Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. (2011). A Maximized Sequential Probability Ratio Test for Drug and Safety Surveillance. *Sequential Analysis*, 30: 58–78. Kulldorff M, Silva IR. (2015). Continuous Post-market Sequential Safety Surveillance with Minimum Events to Signal. *REVSTAT Statistical Journal*, 15(3): 373–394.

Examples

```
### Example 1:
## Sample size required to obtain a power of 80%, for a relative risk of 3, no delay for starting the surveillance (D=0),
## and when the null hypothesis can be rejected with one event (M=1) under an alpha level of 5%.

# result1<- SampleSize.Poisson(alpha=0.05,power=0.8,M=1,D=0,RR=3)
# result1

## Example 2:
## Sample size required to obtain a power of 90%, for a relative risk of 2, no delay for starting the surveillance (D=0),
## and when the null hypothesis can be rejected only after 2 events (M=2) under an alpha level of 10%.
##
# result2<- SampleSize.Poisson(alpha=0.1,power=0.9,M=2,D=0,RR=2)
# result2
```


Index

- *Topic **Binomial sequential analysis**
 - Analyze.Binomial, 11
 - AnalyzeSetUp.Binomial, 25
 - *Topic **Continuous CmaxSPRT CV**
 - CV.CondPoisson, 36
 - *Topic **Continuous CmaxSPRT performance**
 - Performance.CondPoisson, 48
 - *Topic **Continuous CmaxSPRT sample size**
 - SampleSize.CondPoisson, 59
 - *Topic **Continuous MaxSPRT analysis**
 - CV.Binomial, 34
 - CV.Poisson, 43
 - SampleSize.Binomial, 57
 - SampleSize.Poisson, 62
 - *Topic **Group MaxSPRT analysis**
 - CV.G.Binomial, 39
 - CV.G.Poisson, 41
 - *Topic **Poisson sequential analysis**
 - Analyze.CondPoisson, 16
 - Analyze.Poisson, 21
 - AnalyzeSetUp.CondPoisson, 28
 - AnalyzeSetUp.Poisson, 31
 - *Topic **Sequential analysis**
 - Sequential-package, 2
- Analyze.Binomial, 3, 11, 27, 35, 40, 52
Analyze.CondPoisson, 3, 16, 30
Analyze.Poisson, 3, 21, 33
AnalyzeSetUp.Binomial, 3, 11, 13–15, 25
AnalyzeSetUp.CondPoisson, 3, 17–19, 28
AnalyzeSetUp.Poisson, 3, 19, 21–23, 31
- CV.Binomial, 3, 34, 45, 47, 59
CV.CondPoisson, 3, 36, 50, 61
CV.G.Binomial, 3, 15, 39, 43, 52
CV.G.Poisson, 3, 15, 19, 24, 41, 45, 53, 54, 63
CV.Poisson, 3, 40, 43, 43, 52, 55, 56, 63
- Performance.Binomial, 3, 45, 52, 59
Performance.CondPoisson, 3, 38, 48, 61
Performance.G.Binomial, 3, 15, 47, 50, 54
Performance.G.Poisson, 3, 19, 23, 43, 52, 53, 56
Performance.Poisson, 3, 45, 54, 55, 63
- SampleSize.Binomial, 3, 15, 40, 47, 57, 63
SampleSize.CondPoisson, 3, 38, 50, 59
SampleSize.Poisson, 3, 19, 23, 45, 56, 59, 62
Sequential (Sequential-package), 2
Sequential-package, 2