pROC: display and analyze ROC curves

Tools for visualizing, smoothing and comparing receiver operating characteristic (ROC curves). (Partial) area under the curve (AUC) can be compared with statistical tests based on U-statistics or bootstrap. Confidence intervals can be computed for (p)AUC or ROC curves. Sample size / power computation for one or two ROC curves are available.

In S+, the package comes with a graphical user interface.

Authors
> Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller.

Contact
> Xavier Robin

URL
> expasy.org/tools/pROC

License
> GPLv3

Acknowledgements
> The authors would like to thank E. S. Venkatraman and Colin B. Begg for their support in the implementation of their test, and Christophe Combescure and Anne-Sophie Jannot for their help with the implementation of the two ROC curves sample size / power test.

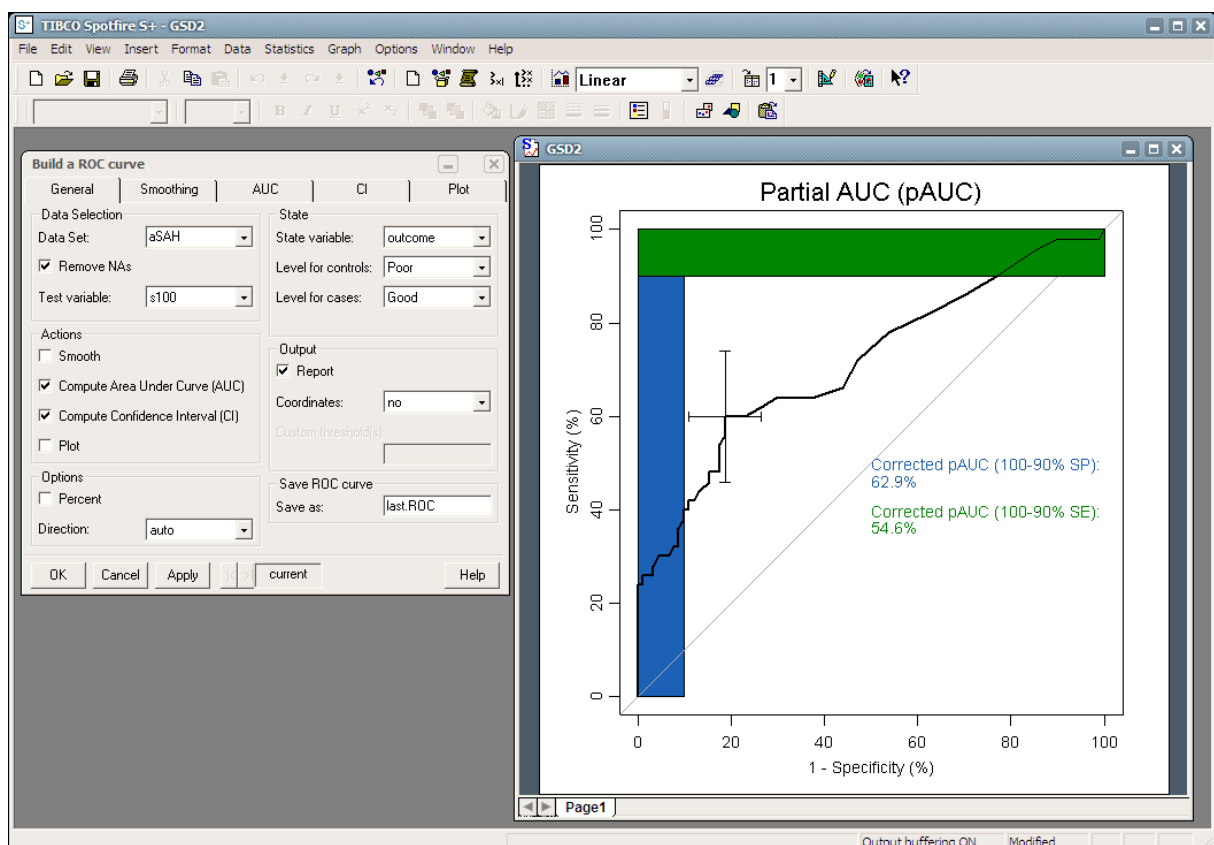**Version of this document**
> 1.5

Date of this document
> November 9, 2011

Table of contents

1. Installing and loading pROC

In the S+ command prompt, type:
```
install.pkgutils()
library(pkgutils)
```

Install pROC from the File menu > Find packages.... After a successful installation, the package must be loaded through the File > Load Library.... More details about the automatic and manual installation and about the update procedure can be found on http://expasy.org/pROC/installS.html.

2. Citation

If you use pROC in published research, please cite the following paper:

> Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). "pROC: an open-source package for R and S+ to analyze and compare ROC curves". *BMC Bioinformatics*, 12, p. 77. DOI: 10.1186/1471-2105-12-77.

The authors would be glad to hear how pROC is employed. You are kindly encouraged to notify Xavier Robin about any work you publish.

3. Using pROC

pROC is available from the Statistics menu of the main menu bar of S+. The options are:

- ROC curve...: to create and plot a ROC curve, and compute confidence intervals and (partial) AUC.
- Paired ROC curve comparison...: statistical tests for two ROC curves on the same data set.
- Paired ROC curve comparison...: statistical tests for two ROC curves on different data set.
- Power tests / sample size...: compute the power or the required sample size of a ROC curve, or of a paired ROC curve comparison test.
- Help: this help file

3.1. Abbreviations

The following abbreviations are employed extensively in this package:

- ROC: receiver operating characteristic
- AUC: area under the ROC curve
- pAUC: partial area under the ROC curve
- CI: confidence interval
- SP: specificity
- SE: sensitivity

4. ROC curve

The ROC curve dialog allows to build a ROC curve, and to compute its (partial) area under the curve, and confidence intervals.

4.1. General tab

- Data Selection
  - Data Set: the name of the data frame to use. The dataset must be loaded into S+ (for example with the File > Import data menu.
  - Remove NAs: if the Test or State variable contain missing values, remove the whole row of the data.
  - Test variable: the name of the test (continuous) variable to test within the data set. It must be a column of the selected data set.
- Actions
  - Smooth: check to apply smoothing to the ROC curve. More details can be set in the Smoothing tab.
  - Compute Area Under Curve (AUC): uncheck if you do not want to compute the AUC. More details can be set in the AUC tab.
  - Compute Confidence Interval (CI): check if you want to compute the CI. More details can be set in the CI tab.
  - Plot: check this box to create a plot. More details can be set in the Plot tab.
- Options
  - Percent: compute and report sensitivity and specificity in percent. If unchecked, fractions (between 0 and 1) are used instead.
  - Direction: the direction of the comparisons made to assess sensitivity and specificity. The comparison is made as "controls <> cases". "auto" determines the direction using the median. To interpret the

threshold, ">" means controls > t ≥ cases, and "<" means controls < t ≤ cases.

- State
  - State variable: this is the variable defining the two groups that the Test variable discriminates.
  - Level for controls: the value of State variable for the control observations.
  - Level for cases: the value of State variable for the case observations.
- Output
  - Report: uncheck if you do not want the report to display.
  - Coordinates: how to report the coordinates of the ROC curve. "no" reports nothing, "local maximas" reports only thresholds corresponding to a local maxima of the ROC curve and "all" reports all the thresholds.
- Save ROC curve
  - Save as: the ROC curve will be saved under this name. You can later access it from the command window, or from the Power tests / sample size window (see section 6. Power tests / sample size).

4.2. Smoothing tab
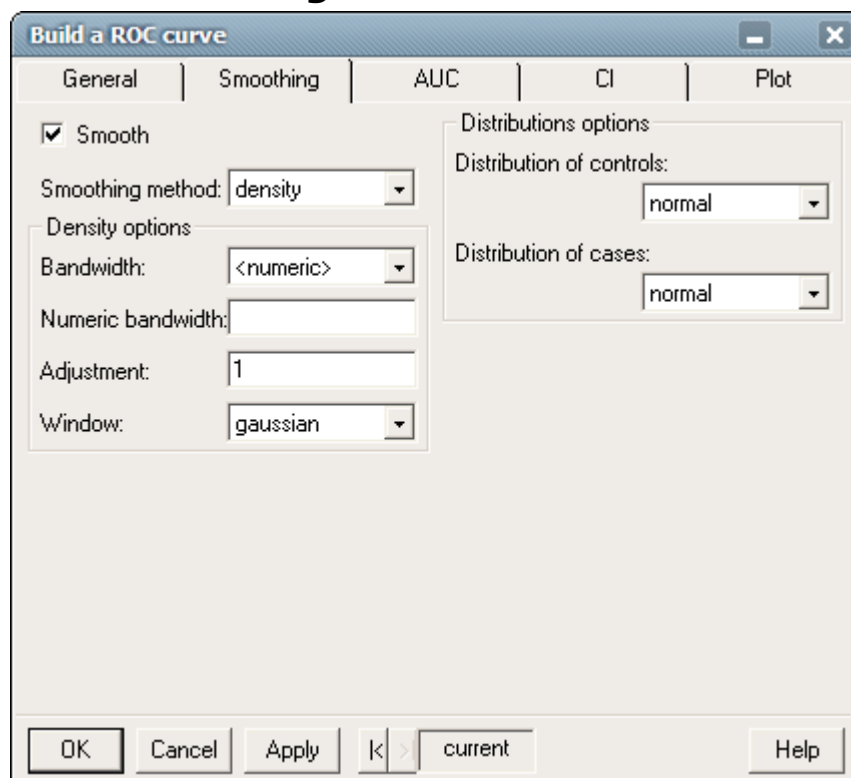
- Smooth: check if you want to perform smoothing.
- Smoothing method: the type of fit that will be employed to smooth the curve. "binormal" is a simple linear fit that will perform well in most cases. "density" will estimate a density distribution for cases and controls separately and deduce a smoothed ROC curve. With "fitdistr" you can specify a distribution for controls and cases.
- Density options: for "density" smoothing.

- Bandwidth: the width of the window. "bcv", "ucv", "sj", "hb" and "nrd" are documented in TIBCO Spotfire S+ Language Reference as bandwidth.* functions. "<custom>" permits to specify a numeric value in the following field.
- Numeric bandwidth: if "bandwidth" is "<custom>", a numeric value specifying the width of the window.
- Adjustment: the bandwidth will be multiplied by Adjustment. 1 will have no effect.
- Window: the kind of window. Note that all the windows are symmetric.
- Distribution options:
  - Distribution of controls, distribution of cases: density distributions for control and cases observations.

4.3. AUC tab
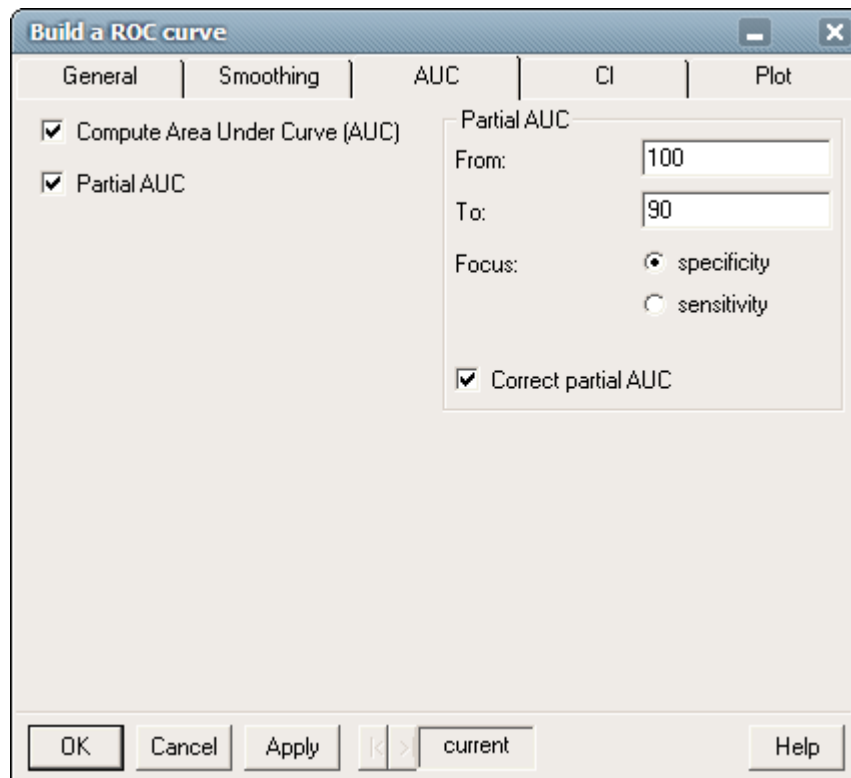
- Compute Area Under Curve (AUC): uncheck if you do not want to compute the AUC.
- Partial AUC: compute only a portion of the AUC, as defined in the Partial AUC group:
  - From, to: the bounds of the partial AUC. If percent was checked on the General tab, enter the numbers as percent (0-100), otherwise as proportion (0-1).
  - Focus: if the bounds are expressed in terms of sensitivity or specificity.
  - Correct partial AUC: correct the partial AUC in order to always have an area of 1 (or 100%) for a perfect partial discrimination and 0.5 (50%) for no discrimination, whatever the region specified. The formula by McClish (1989) is used.

4.4. CI tab
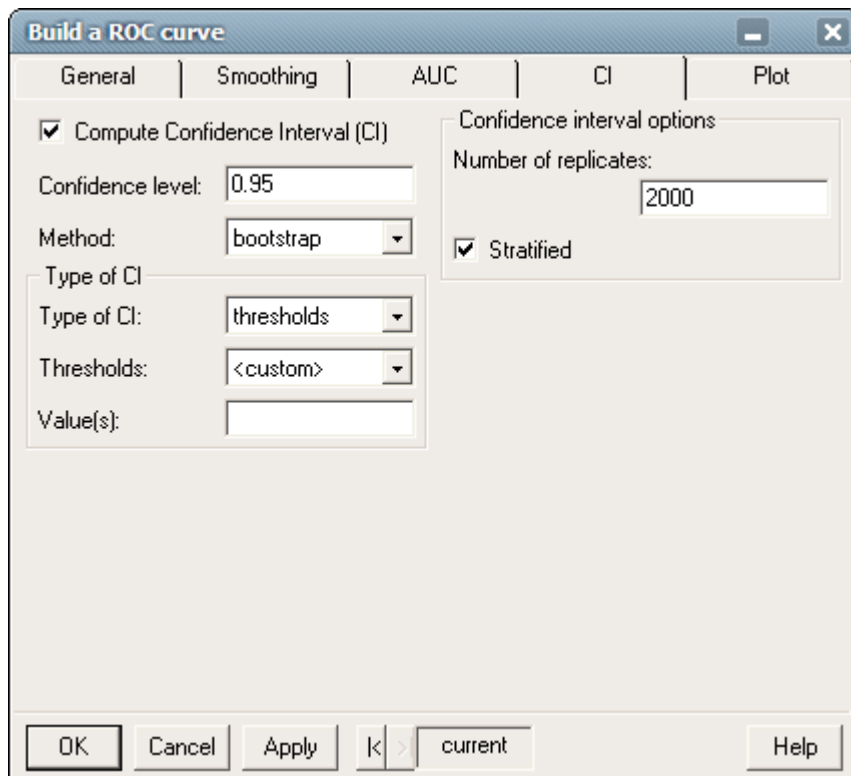
- Compute Confidence Interval (CI): check if you want to compute the CI.
- Confidence level: always between 0 and 1 (even if percent was checked on the General tab).
- Method: "delong" is faster and give exact results but is only available for full AUC with no smoothing. If only "bootstrap" is available, make sure "smooth" (in the Smoothing tab) and "partial AUC" (in the AUC tab) are not checked, and that "auc" is selected in Type of CI. "auto" will select the best method automatically.
- Type of CI:
  Note: modifications in these fields will be reflected in the General tab / output coordinates field and in the Plot tab / Plot CI, only if they have not been modified before.
  - Type of CI: "auc" to compute the CI of the AUC (as specified in the AUC tab), "thresholds" for the CI of thresholds (as specified in the Thresholds field just below), "se" to compute the CI of the sensitivity at fixed specificities, or "sp" to compute the CI of the specificity at fixed sensitivities.
  - Thresholds: if the type of CI is "thresholds", which thresholds must be tested? "all", "local maximas", or "<custom>" to define the threshold(s) in the Value field.
  - Value(s): if type of CI is "se" or "sp", or if "thresholds" is "<custom>", enter the values on which the CI will be assessed. For "se" or "sp", make sure to be consistent with the Percent checkbox in the General tab, or leave blank so the program will guess default values. You cannot leave the field blank for a custom threshold.
    Note: you can enter several values with an S+ expression, for example c(0.95, 0.9, 0.85) or seq(0, 1, 0.1).

– Confidence interval options:
  – Number of replicates: how many replicates for the CI? Higher numbers give a better approximation but take more time to compute
  – Stratified: controls and cases are sampled separately, so that each replicate contains exactly the same number of them.
  If unchecked, a non-stratified bootstrap is performed, and some replicates may not contain any case or control if the data is small or unbalanced. In such a case, a warning will be displayed and the resulting missing values will be ignored.

4.5. Plot tab
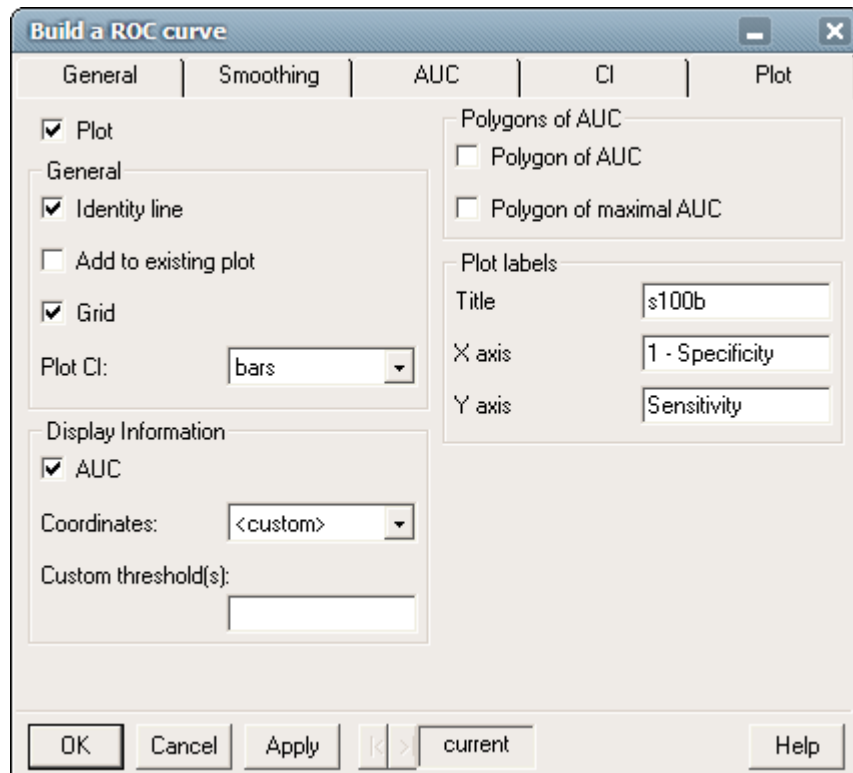
– Plot: check this box to create a plot.
– General:
  – Identity line: check to display the diagonal line showing the identity (no discrimiation power)
  – Add to existing plot: if unchecked, a new plot will be created. Otherwise, the ROC curve will be added to an existing one. Don't check if there is no current plot, or if the current plot is not a ROC curve.
  – Grid: display a grid behind the ROC curve
  – Plot CI: "no": don't plot the confidence interval (only value if you didn't select confidence intervals in the General or CI tabls). "bars": display the intervals as bars. "shape": only if the CI is of type "se" or "sp": show a confidence shape rather than bars.
– Display Information
  – AUC: print the value of the AUC in the center of the plot.
  – Coordinates: display the thresholds with specificity and sensitivity for a selected set of ROC points. "no" inhibits the printing, "best" shows

only the points corresponding to the maximal sum of sensitivity and specificity, "all" prints all the thresholds and "local maximas" shows a point with the numeric coordinates for each local maxima of the ROC curve. If "custom", fill-in the following field with the threshold(s) you want to display on the plot. Only "no" and "best" are available for smoothed ROC curves, and only sensitivity and specificity will be printed.

- Custom threshold(s): if "custom" is selected in "Coordinates", this field selects the threshold(s) that must be plotted. You can enter several thresholds with the following syntax: c(5, 8.5) to display the thresholds at 5 and 8.5.

- Polygons of AUC:
  - Polygon of AUC: highlight the AUC.
  - Polygon of maximal AUC: highlight the maximal region the AUC could take.
- Plot labels:
  - Title: the title above the graph, by default the name of the Test variable
  - X axis, Y axis: the labels of the X and Y axis.

5. Comparing ROC curves

The "ROC curves comparison" dialogs serve to make a statistical comparison of two ROC curves. The "Paired ROC curves comparison" is for ROC curves of variables measured on the same sample. If different samples are involved, select the "Unpaired ROC curves comparison" menu.

5.1. General tab (unpaired)

This tab is similar to the General tab to build ROC curves (4.1.). You have to select both data sets and the test variables, state variables and levels for each data set separately.

5.2. General tab (paired)

- Data Selection
  - Data Set: the name of the data frame to use.
  - Remove NAs: if one of the Test or State variable contains missing values, remove the whole row of the data.
  - Test variable1, Test variable 2: the names of the two test (continuous) variables defining the two ROC curves to compare.
- Options
  - Method: "delong" (only for comparing complete AUCs), "bootstrap" or "venkatraman" (only for comparing complete AUCs). With "auto", the appropriate method will be determined automatically (between bootstrap and delong. See the section 5 (Algorithms) of this document for more details about the algorithms implemented.
  - Percent: compute and report sensitivity and specificity in percent. If unchecked, fractions (between 0 and 1) are used instead.
  - Direction: the direction of the comparisons made to assess sensitivity and specificity. The comparison is made as "controls <> cases". "auto" determines the direction using the median.
- State
  - State variable: this is the variable defining the two groups that the Test variables discriminates.
  - Level for controls: the value of State variable in the control observations.
  - Level for cases: the value of State variable in the case observations.
- Output
  - Report: uncheck if you do not want the report to display.
- Save ROC curve

– Save as: the comparison object (htest) will be saved under this name. You can later access it from the command window.

5.3. Smoothing tab (paired and unpaired)
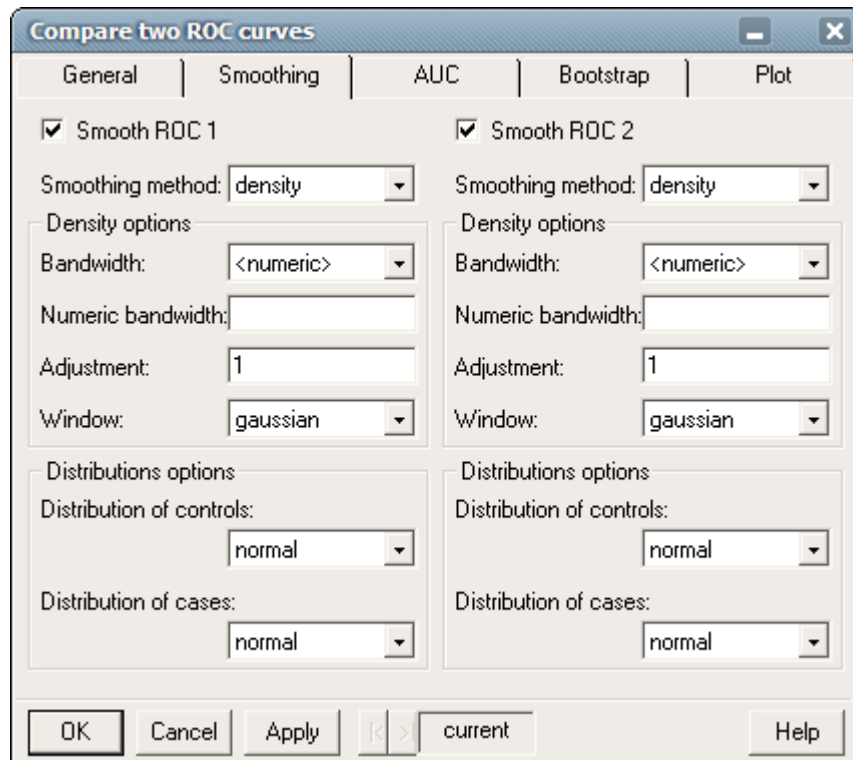
– Smooth ROC 1, Smooth ROC 2: smooth the ROC curves corresponding to Test variable1 and Test variable 2.
– Smoothing method: the type of fit that will be employed to smooth the curve. "binormal" is a simple linear fit that will perform well in most cases. "density" will estimate a density distribution for cases and controls separately and deduce a smoothed ROC curve. With "fitdistr" you can specify a distribution for controls and cases.
– Density options: for "density" smoothing.
  – Bandwidth: the width of the window. "bcv", "ucv", "sj", "hb" and "nrd" are documented in TIBCO Spotfire S+ Language Reference as bandwidth.* functions. "<custom>" permits to specify a numeric value in the following field.
  – Numeric bandwidth: if "bandwidth" is "<custom>", a numeric value specifying the width of the window.
  – Adjustment: the bandwidth will be multiplied by Adjustment. 1 will have no effect.
  – Window: the kind of window. Note that all the windows are symmetric.
– Distribution options:
– Distribution of controls, distribution of cases: density distributions for control and cases observations.

5.4. AUC tab (paired and unpaired)
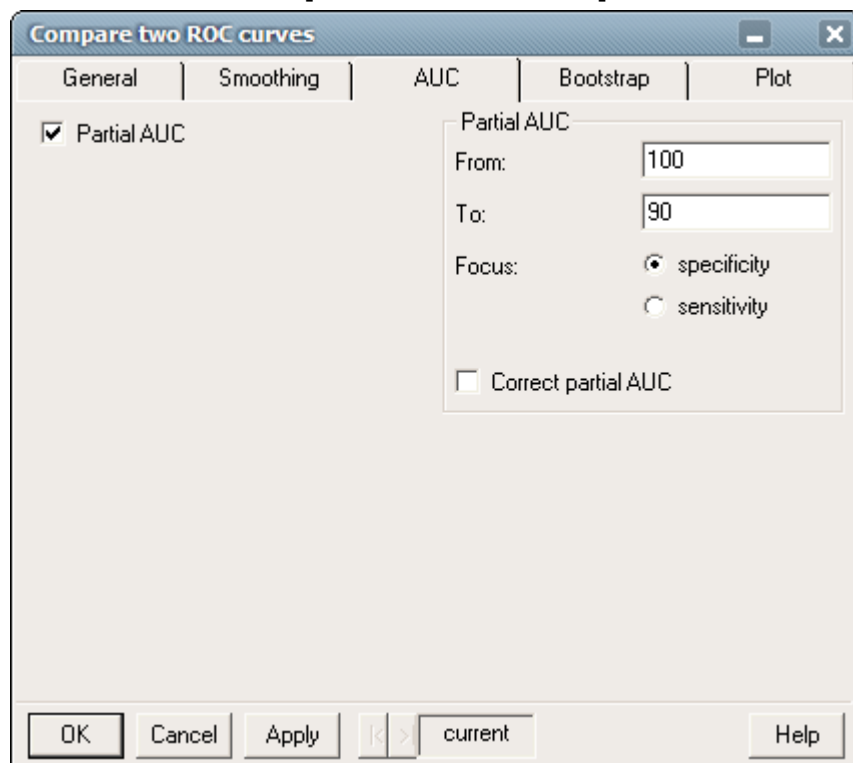
– Partial AUC: compute only a portion of the AUC, as defined in the Partial AUC group (not available if "method" (in "General" tab) is "delong" or "venkatraman"; enabling this option will also disable "delong" and "venkatraman" in the "General" tab "method" field):
  – From, To: the bounds of the AUC. If "percent" was checked on the "General" tab, enter the numbers as percent (0-100), otherwise as proportion (0-1).
  – Focus: if the bounds are expressed in terms of sensitivity or specificity.
  – Correct partial AUC: correct the partial AUC in order to always have an area of 1 (or 100%) for a perfect partial discrimination and 0.5 (50%) for no discrimination, whatever the region specified. The formula by McClish (1989) is used.

5.5. Bootstrap tab (paired and unpaired)
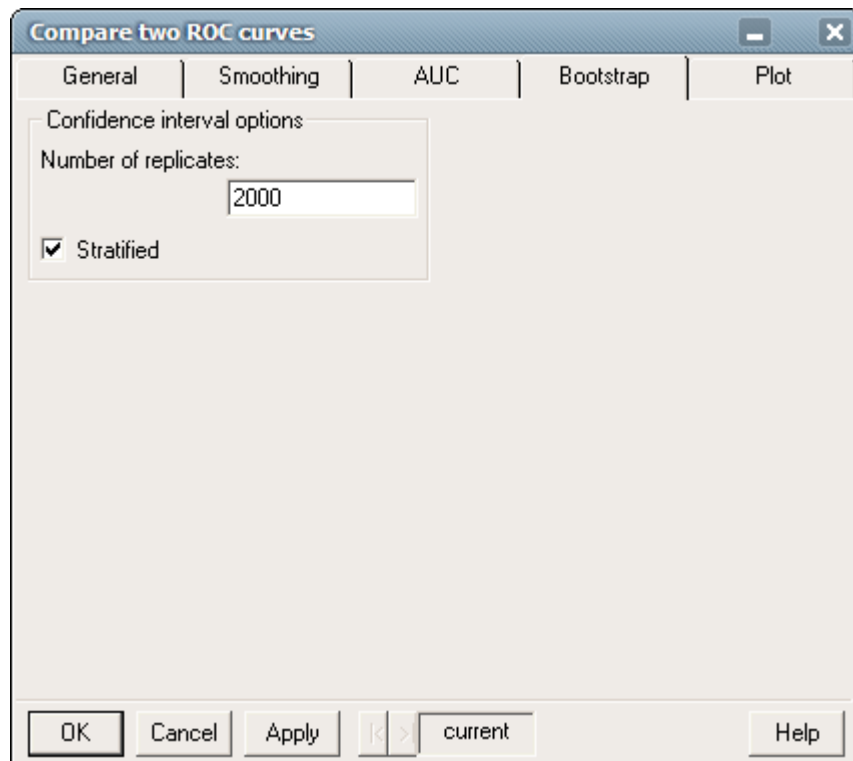
- – Bootstrap options (not available if "method" (in "General" tab) is "delong") for "bootstrap" test, and for the permutation of "venkatraman" test:
  - – Number of replicates: how many bootstrap replicates for the boostrap test, or how many permutations for the venkatraman test? Higher numbers give a better approximation but take more time to compute
  - – Stratified: (ignored unless "method" (in "General" tab) is set on "bootstrap"): controls and cases are sampled separately, so that each replicate contains exactly the same number of them. If unchecked, a non-stratified bootstrap is performed, and some replicates may not contain any case or control if the data is small or unbalanced. In such a case, a warning will be displayed and the resulting missing values will be ignored.

5.6. Plot tab (paired and unpaired)
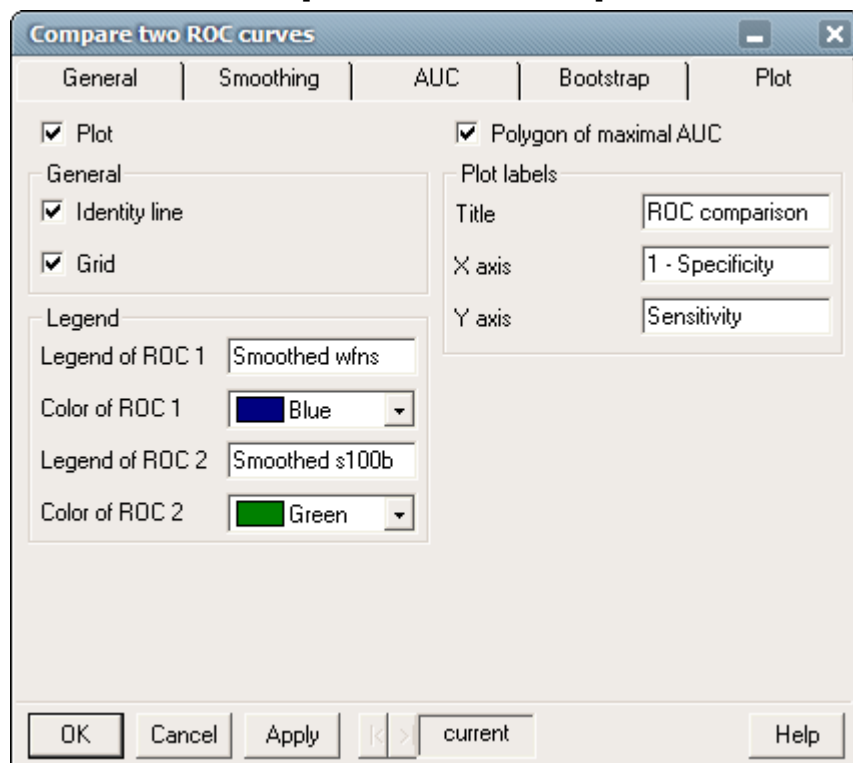
- – Plot: check this box to plot the two ROC curves that are compared.
- – General:
  - – Identity line: check to display the diagonal line showing the identity (no discrimiation power)
  - – Grid: display a grid behind the ROC curves
- – Legend: the legend that will be displayed in the plot.
  - – Legend of ROC 1, Legend of ROC 2: legend annotation.
  - – Color of ROC 1, Color of ROC 2: the colors of the two ROC curves in the plot, corresponding to the legend.
- – Polygon of maximal AUC: highlight the maximal region the AUC could take.
- – Plot labels:
  - – Title: the title above the graph.
  - – X axis, Y axis: the labels of the X and Y axis.

6. Power tests / sample size

The "Power tests and Sample size" dialogs serve allows making power and sample size test in relation with a two paired ROC curves test, or a one ROC curve test (Wilcoxon). It is available from the "Statistics" menu, either in pROC's "ROC curves" sub-menu with the "Power tests / sample size..." button, or from the "Power and sample size" sub-menu with button "ROC curves...".

See section 7.4 Power test / sample size for implementation details.
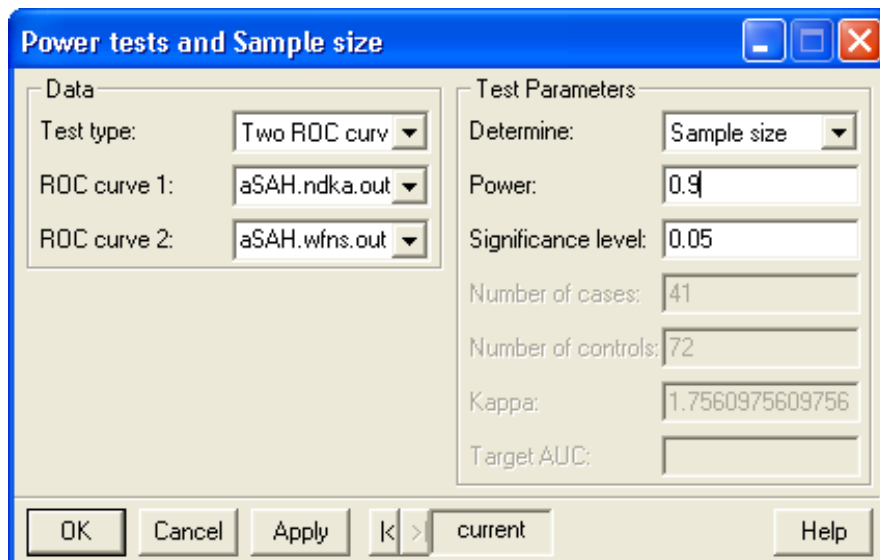
- Test type: which test to perform.
  - No ROC curve: enter estimated values for the sample size / power test of a single ROC curve
  - One ROC curve: sample size / power test for the single given ROC curve
  - Two ROC curves: sample size / power test for a paired comparison test between two ROC curves. ROC curve 1 (see below) is taken as the reference ROC curve.
- ROC curve 1, ROC curve 2: the ROC curves to test. For a Two ROC curves test, ROC curve 1 is taken as the reference ROC curve.
- Determine: states which parameter must be determined. As it is selected, the fields to fill in below should become white. All white fields must be filled for the test to succeed.
- Power: the target power of the test (1 – β error)
- Significance level: the target significance level of the test (α error)
- Number of cases, Number of controls: the given sample size
- Kappa: the ratio of the number of controls to the number of cases (κ = #controls / # cases).
- Target AUC: for a one-ROC curve test, the AUC of the ROC curve.

7. Algorithms

7.1. Area Under the Curve

The AUC is computed with the trapezoidal rule as described by Fawcett (2006). The partial AUC has been theoretized by McClish (1989). It is computed with the trapezoidal rule with interpolation if necessary. The partial AUC correction is done as described by McClish (formula 6).

7.2. Confidence intervals

The CI are computed with bootstrap. In stratified bootstrap, control and case observations are sampled independently, otherwise all the observations are sampled at the same time.

After "Number of replicates" repetitions, the quantiles of the resulting statistics are computed at (1 - "confidence level")/2, 0.5 and 1 - (1- "confidence level")/2.

7.3. ROC curve comparison

7.3.1. Bootstrap

With the "bootstrap" method, the processing is done as follow:

1. "Number of replicates" bootstrap replicates are drawn from the data. If "stratified" is checked, each replicate contains exactly the same number of controls and cases than the original sample, otherwise the numbers can vary.
2. for each bootstrap replicate, the AUC of the two ROC curves are computed and the difference is stored.
3. The following formula is used:

$$D=(AUC1\text{-}AUC2)/s$$

   where s is the standard deviation of the bootstrap differences and AUC1 and AUC2 the AUC of the two (original) ROC curves.

4. *D* is then compared to the normal distribution to compute the p-value.

7.3.2. DeLong

With the "delong" method, the processing is done as described in DeLong *et al.* (1988). Only comparison of two ROC curves is implemented. The method has been extended for unpaired ROC curves where the p-value is computed with an unpaired t-test with unequal sample size and unequal variance.

7.3.3. Venkatraman

The "venkatraman" method is implemented following the algorithm described in Venkatraman and Begg (1996) (for paired ROC curves) and Venkatraman (2000) (for unpaired ROC curves). The number of permutations is defined by the number of bootstrap replicates (section 5.5.) even though permutations are performed instead of bootstrap.

7.4. Power test / sample size

7.4.1. One ROC curve power calculation

When no ROC curve is given, the sample size computation is performed with the formulas 2 and 3 given in Obuchowski *et al.* (2004), p. 1123. For the sample size, the number of cases is computed directly from formulas 2 and the number of controls is deduced with kappa. AUC is optimized by S+ uniroot while power and significance levels are solved as quadratic equations.

When one ROC curve is given, the parameters (AUC, number of patients, kappa) are extracted and the computation is performed as with no ROC curve.

7.4.2. Two paired ROC curves power calculation

If two ROC curves are given, the computations is done with the formula 2 given in Obuchowski and McClish (1997), p. 1530. The variances and covariance are computed with DeLong (DeLong *et al.,* (1988), for full AUC) or bootstrap (for partial AUC). The null hypothesis is that the AUC of ROC curve 1 is the same than the AUC of ROC curve 2, with ROC curve 1 taken as the reference ROC curve.

For the sample size, the number of cases is computed directly from formulas 2 and the number of controls is deduced with kappa. Power and significance levels are solved as quadratic equations.

Power calculation for two unpaired ROC curves is not implemented.

8. References

James Carpenter and John Bithell (2000) "Bootstrap condence intervals: when, which, what? A practical guide for medical statisticians". *Statistics in Medicine* **19**, p. 1141–1164.

Elisabeth R. DeLong, David M. DeLong and Daniel L. Clarke-Pearson (1988) "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". *Biometrics* **44**, p. 837–845.

Tom Fawcett (2006) "An introduction to ROC analysis". *Pattern Recognition Letters* **27**, p. 861–874. DOI: 10.1016/j.patrec.2005.10.010.

James A. Hanley and Barbara J. McNeil (1982) "The meaning and use of the area under a receiver operating characteristic (ROC) curve". *Radiology* **143**, p. 29–36.

Donna Katzman McClish (1989) "Analyzing a Portion of the ROC Curve". *Medical Decision Making* **9**, p. 190–195. DOI: 10.1177/ 0272989X8900900307.

Nancy A. Obuchowski and Donna K. McClish (1997) "Sample size determination for diagnostic accurary studies involving binormal ROC curve indices". *Statistics in Medicine*, **16**, p. 1529–1542. DOI: 10.1002/(SICI)1097-0258(19970715)16:13<1529::AID-SIM565>3.0.CO;2-H.

Nancy A. Obuchowski, Micharl L. Lieber and Frank H. Wians Jr. (2004) "ROC Curves in Clinical Chemistry: Uses, Misuses, and Possible Solutions". *Clinical Chemistry*, **50**, 1118–1125. DOI: 10.1373/clinchem.2004.031823.

Margaret Pepe, Gary Longton and Holly Janes (2009) "Estimation and Comparison of Receiver Operating Characteristic Curves". *The Stata journal* **9**, 1. PMID: 20161343.

Xavier Robin, Natacha Turck, Jean-Charles Sanchez and Markus Müller (2009)" Combination of protein biomarkers". *useR! 2009*, Rennes.

Xavier Robin, Natacha Turck, Alexandre Hainard, *et al.* (2011) "pROC: an open-source package for R and S+ to analyze and compare ROC curves". *BMC Bioinformatics*, **7**, p. 77. DOI: 10.1186/1471-2105-12-77.

E. S. Venkatraman and Colin B. Begg (1996) "A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment". Biometrika **83**, p. 835–848. DOI: 10.1093/biomet/83.4.835.

E. S. Venkatraman (2000) "A Permutation Test to Compare Receiver Operating Characteristic Curves". *Biometrics* **56**, p. 1134–1138. DOI: 10.1111/j.0006-341X.2000.01134.x.

Kelly H. Zou, W. J. Hall and David E. Shapiro (1997) "Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests". *Statistics in Medicine* **18**, p. 2143–2156. DOI: 10.1002/(SICI)1097-0258(19971015)16:19<2143::AID-SIM655>3.0.CO;2-3.