robnptests – An R package for robust two-sample location and variability tests

09 January 2021

Summary

The package `robnptests` is a compilation of two-sample tests based on two criteria: The tests are (i) robust against outliers and (ii) (approximately) distribution free. Regarding the latter aspect, we implemented tests that keep an intended significance level and provide a reasonably high efficiency, both under a variety of continuous distributions. Robustness is achieved by using test statistics that are based on robust location and scale measures.

In what follows, we give a brief description of the package's contents. More details can be found in the introductory vignette of the package, which can be called by `vignette("robnptests-vignette")`, and the cited papers.

Data situation

We consider two samples of independent and identically distributed (i.i.d.) random variables $X_1, ..., X_m$ and $Y_1, ..., Y_n$, respectively. The underlying distributions are assumed to be continuous with cumulative distribution functions F_X and F_Y.

The tests are can be used for either one of the following cases:

1. Assuming that the parameters of both distributions are equal while the location parameters may be unequal, i.e. $F_X(x) = F_Y(x + \Delta)$, $\Delta \in \mathbb{R}$, the tests can be used to detect a location difference between the samples.
2. For equal location parameters with possibly unequal scale parameters, i.e. $F_X(x) = F_Y(x/\theta)$, $\theta > 0$, a transformation of the observations enables to identify differing scale parameters.

Statement of need

A popular test for the location setting is the two-sample t-test. It is considered to be robust against deviations from the normality assumption by keeping the specified significance level due to the central limit theorem. However, non-normality can result in a loss of power (Wilcox 2003). In addition, particularly for small samples, the t-test is prone to outliers (R. Fried and Dehling 2011). Distribution-free tests, like the two-sample Wilcoxon rank-sum test, can be nearly as powerful as the t-test under normality and may have higher power under non-normality. Still, they also can be vulnerable to outliers (Roland Fried and Gather 2007). The two-sample tests in `robnptests` combine (approximate) distribution independence and robustness against outliers. Thus, they are better suited for outlier-corrupted samples from unknown data-generating distributions. At the same time, such tests can be nearly as powerful as popular procedures like the aforementioned t-test or the Wilcoxon test on uncontaminated samples.

Figure 1 compares the power of the t-test, the Wilcoxon test and two robust tests: one based on the one-sample Hodges-Lehmann estimator (HL1 test) (Hodges and Lehmann 1963), and one based on Huber's M-estimator (Huber 1964) for a fixed location difference between the samples and a single outlier of increasing size. The power of the t-test decreases to zero, while the loss in power of the Wilcoxon test and both robust tests is bounded. The robust tests provide a higher power than the Wilcoxon test. The differences between the

Wilcoxon test and robust tests like those in the example can become more apparent when more outliers are involved (R. Fried and Dehling 2011).

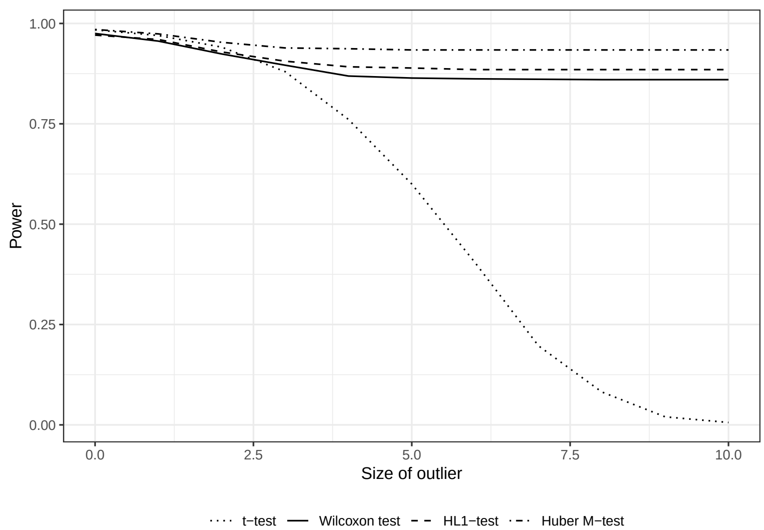

Figure 1: Power of the two-sample t-test, the Wilcoxon rank sum test, and a robust test based on the one-sample Hodges-Lehmann estimator on two samples of size $m = n = 10$ from two normal distributions with a location difference of $\Delta = 2$ and a single outlier of increasing size.

Like the t-test for the location problem, the performance of parametric tests for the location problem may deteriorate when their distributional assumption is not fulfilled. Moreover, nonparametric tests may be robust against violations of the distributional assumptions or outliers, but many of them cannot cope well with asymmetry (Roland Fried 2012). Applying the robust location tests to transformed observations as proposed by Roland Fried (2012), yields good results in terms of power and size under asymmetry and outlier corruption. However, the tests may be less efficient under symmetry.

Implemented two-sample tests

Each test statistic consists of a robust estimator for the location difference between the two populations that should be compared. This difference is divided by a robust estimator for the within-sample variance.

To obtain a distribution-free test decision, the p-value can be computed by using the permutation principle, the randomization principle, or a normal approximation. With the permutation principle, the tests hold the desired significance level exactly at the cost of large computing times even for quite small samples such as $m = n = 10$. The time can be reduced by using a randomization distribution and, even more, by taking advantage of the asymptotic normality of the location-difference estimators. The latter approach, however, is only advisable for large sample sizes $m, n > 30$.

The tests based on the following location estimators are described in R. Fried and Dehling (2011):

- The *difference of the sample medians* helps to achieve high robustness. However, this estimator is not very efficient under the normal distribution or distributions that do not deviate too much from it.
- To improve the efficiency, one can use the difference of the *one-sample Hodges-Lehmann estimators* (Hodges and Lehmann 1963) at the cost of loosing some robustness due to the lower breakdown point.
- Similarly, the *two-sample Hodges-Lehmann estimator* leads to a robust test with a higher power under normality than the tests based on the sample median.

For scaling, we use different estimators based on medians and pairwise differences (R. Fried and Dehling 2011).

In addition, we implemented tests based on *M-estimators*. This approach to robust location estimation allows for flexibility in how outliers are treated through the specification of the parameters of the corresponding ρ-function. We focus on Huber's ρ-function, the bisquare function and the Hampel ρ-function in a similar manner as described in Aboukalam (1992). Moreover, the package contains Yuen's t-test which uses the difference of *trimmed means* to estimate the location difference (Yuen and Dixon 1973).

Some of the robust scale estimators may become zero when ties occur in the data. This can happen in real-world applications when the measurements are rounded or stem from discrete distributions. Discretization can also lead to a loss in power or conservative tests. To cope with this, we add random noise from a uniform distribution with a small variance to each observation. This procedure is called `wobbling`. The objective to enable the computation of the test statistic without distorting the observations too much.

A more detailed overview of the implemented tests and corresponding test statistics can be found in the vignettes of the package.

Applications

Besides conventional two-sample problems, the tests can be used for the online detection of structural breaks in outlier-contaminated time series. Abbas, Fried, and Gather (2016) describe how intensity changes in image sequences generated by a virus sensor are automatically detected by applying the tests to the individual pixel time series of the sequence. Moreover, the test statistics can be used as control statistics for robust, (approximately) distribution-free control charts for time-series with a time-varying signal (Abbas and Fried 2017, 2020). In Abbas et al. (2019), the tests were applied to detect unusual sequences in time series of crack widths in concrete beams by searching for sudden scale changes.

Other packages with robust two-sample tests

The CRAN Task View for robust statistical methods currently lists the three packages `WRS2` (Mair and Wilcox 2019), `walrus` (Love and Mair 2018), and `robeth` (Marazzi 2020) that explicitly deal with robust hypothesis tests for the two-sample problem. `WRS2` contains a large collection of robust procedures which are presented in the book *Introduction to Robust Estimation and Hypothesis Testing* by (Wilcox 2012). `walrus` provides a different user interface for the functions in `WRS2`. The package `robeth` (Marazzi 2020) contains some robust tests for linear hypotheses.

The functions in `WRS2` concentrate on the heteroscedastic setting, whereas our focus lies on the homoscedastic case. The reason is that especially for small samples, estimating the within-sample variance separately for both samples, as is the case under heteroscedasticity, may lead to unreliable estimates. Moreover, choosing equal sample sizes $m = n$, can protect against a deteriorating test performance of our implemented tests under heteroscedasticity in terms of size and power (Staudte and Sheather 1990, 179).

The R package `nptest` (Helwig 2021) contains nonparametric versions of the two-sample t-test, which is realized by using the permutation and randomization principles described above on the t-statistic. This principle has also been studied in Abbas and Fried (2017), and, while achieving distribution independence, the test statistic lacks robustness against outliers.

References

Abbas, Sermad, and Roland Fried. 2017. "Control Charts for the Mean Based on Robust Two-Sample Tests." *Journal of Statistical Computation and Simulation* 87 (1): 138–55. https://doi.org/10.1080/00949655.2016.1194839.

———. 2020. "Robust Control Charts for the Mean of a Locally Linear Time Series." *Journal of Statistical Computation and Simulation* 90 (15): 2741–65. https://doi.org/10.1080/00949655.2020.1788562.

Abbas, Sermad, Roland Fried, and Ursula Gather. 2016. "Detection of Local Intensity Changes in Grayscale Images with Robust Methods for Time-Series Analysis." In *Solving Large Scale Learning Tasks: Challenges and Algorithms: Essays Dedicated to Katharina Morik on the Occasion of Her 60th Birthday*, edited by Stefan Michaelis, Nico Piatkowski, and Marco Stolpe, 251–71. Cham: Springer.

Abbas, Sermad, Roland Fried, Jens Heinrich, Melanie Horn, Mirko Jakubzik, Johanna Kohlenbach, Reinhard Maurer, Anne Michels, and Christine H. Müller. 2019. "Detection of Anomalous Sequences in Crack Data of a Bridge Monitoring." In *Applications in Statistical Computing: From Music Data Analysis to Industrial Quality Improvement*, edited by Nadja Bauer, Katja Ickstadt, Karsten Lübke, Gero Szepannek, Heike Trautmann, and Maurizio Vichi, 251–69. "Cham": Springer. https://doi.org/10.1007/978-3-030-25147-5_16.

Aboukalam, M. A. F. 1992. "Some Robust Two-Sample Test Statistics Based on M-Estimators of Location." *Communications in Statistics - Simulation and Computation* 21 (1): 133–48.

Fried, R., and H. Dehling. 2011. "Robust Nonparametric Tests for the Two Sample Location Problem." *Statistical Methods & Applications* 20 (4): 409–22. https://doi.org/10.1007/s10260-011-0164-1.

Fried, Roland. 2012. "On the Online Estimation of Piecewise Constant Volatilities." *Computational Statistics & Data Analysis* 56 (11): 3080–90. https://doi.org/10.1016/j.csda.2011.02.012.

Fried, Roland, and Ursula Gather. 2007. "On Rank Tests for Shift Detection in Time Series." *Computational Statistics & Data Analysis* 52: 221–33. https://doi.org/10.1016/j.csda.2006.12.017.

Helwig, Nathaniel E. 2021. *Nptest: Nonparametric Bootstrap and Permutation Tests*. https://CRAN.R-project.org/package=nptest.

Hodges, Joseph L., and Erich L. Lehmann. 1963. "Estimates of Location Based on Rank Tests." *The Annals of Mathematical Statistics* 34 (2): 598–611.

Huber, Peter J. 1964. "Robust Estimation of a Location Parameter." *The Annals of Mathematical Statistics* 35 (1): 73–101. https://doi.org/10.1214/aoms/1177703732.

Love, Jonathon, and Patrick Mair. 2018. *Walrus: Robust Statistical Methods*. https://CRAN.R-project.org/package=walrus.

Mair, Patrick, and Rand Wilcox. 2019. "Robust Statistical Methods in R Using the WRS2 Package." *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01246-w.

Marazzi, Alfio. 2020. *Robeth: R Functions for Robust Statistics*. https://CRAN.R-project.org/package=robeth.

Staudte, Robert G., and Simon J. Sheather. 1990. *Robust Estimation and Testing*. New York: Wiley.

Wilcox, Rand R. 2003. *Applying Contemporary Statistical Techniques*. Amsterdam: Academic Press.

———. 2012. *Introduction to Robust Estimation and Hypothesis Testing*. Fourth. Amsterdam: Elsevier.

Yuen, Karen K., and W. T. Dixon. 1973. "The Approximate Behaviour and Performance of the Two-Sample Trimmed t." *Biometrika* 60 (2): 369–74.