A Lepage-type test for the GSHD

O.Y.Kravchuk, P.K.Pollett

University of Queensland, Brisbane, Australia
o.kravchuk@uq.edu.au, pkp@maths.uq.edu.au

Abstract

We introduce a new location-scale rank test efficient for the generalised secant hyperbolic distribution (GSHD). The GSHD consists of symmetric unimodal distributions of various tails, and is interesting in applications where the lack of normality is explained by the tail behaviour of the data distributions. The new test is a family of Lepage-type tests, each of which combines the standardised location and scale rank statistics efficient under their alternative hypotheses for a specific distribution.

Introduction

The GSHD was introduced in 2002 in [1]. The GSHD is a location-scale family of unimodal symmetric distributions that includes the Cauchy and the uniform distributions as its limiting heavy-tail and light-tail cases. A member of the distribution is completely specified by the location, scale, and shape (tail) parameters. The two-sample linear rank procedures of location and scale alternatives efficient for the GSHD were introduced in 2006 in [2, 3]. Both the location and scale rank procedures are robust to distributional misspecifications. However, the scale procedures are extremely sensitive to the presence of outliers. Moreover, the location estimators are regular almost for the whole family, while the scale ratio rank estimator is only regular conditioned on no difference in location. The optimal location, φ, and scale, φ_1, score functions are defined at $u \in [0,1]$ and given below for the range of the tail parameter, t:

$$(\varphi(u,t), \varphi_1(u,t)) = \begin{cases} \left(\sin(t(2u-1)), \ln\left(\frac{\sin(tu)}{\sin(t(1-u))}\right)\sin(t(2u-1))\right) & -\pi < t < 0, \text{ heavy-tailed}, \\ \left(2u-1, (2u-1)\ln\left(\frac{u}{1-u}\right)\right) & t = 0, \text{ logistic}, \\ \left(\sinh(t(2u-1)), \ln\left(\frac{\sinh(tu)}{\sinh(t(1-u))}\right)\sinh(t(2u-1))\right) & t > 0, \text{ normal and light-tailed}. \end{cases}$$

In what follows, we consider the two-sample location-scale problem with two independent samples of sizes m and n, $m+n=N$; let c denotes the following coefficients: $c(i) = \frac{1}{m}\sqrt{\frac{mn}{N}},\ i < m,\ c(i) = -\frac{1}{n}\sqrt{\frac{mn}{N}},\ i > m$; let R denote the rank of an observation in the pooled data set.

Location-scale rank test

We have studied and compared the properties of the following location-scale rank tests:

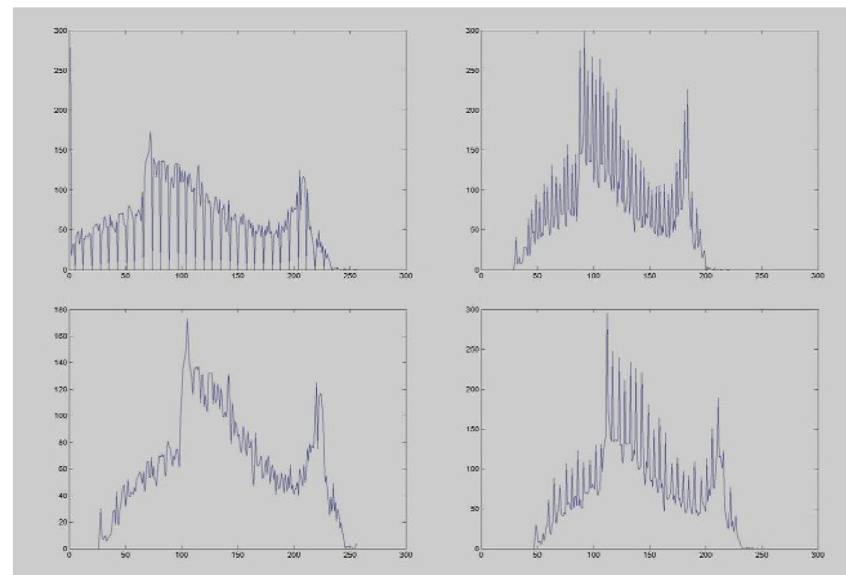
$$S(t=-\pi/2) = \left(\frac{\sqrt{2}\pi/(2N)}{\sin(\pi/(2N))}\right)^2 \left[\sum_{i=1}^{N} c(i)\cos\left(\frac{\pi}{2}\frac{2R_i-1}{N}\right)\right]^2 + \frac{8}{\pi^2}\left[\sum_{i=1}^{N} c(i)\ln\left(\tan\left(\frac{\pi}{2}\frac{R_i}{N+1}\right)\right)\cos\left(\pi\frac{R_i}{N+1}\right)\right]^2,$$

$$S(t=-\pi) = \frac{16\pi^2}{3N^2}\left[\left(\frac{2}{N}\sum_{i=[N/4]}^{3[N/4]}\cos\left(2\pi\frac{i}{N}\right)\sum_{j=1}^{i} c(R_i)\right)\right]^2 + \frac{8\pi^2}{N^2}\left[\sum_{i=1}^{N}\sin\left(2\pi\frac{i}{N}\right)\sum_{j=1}^{i} c(R_i)\right]^2,$$

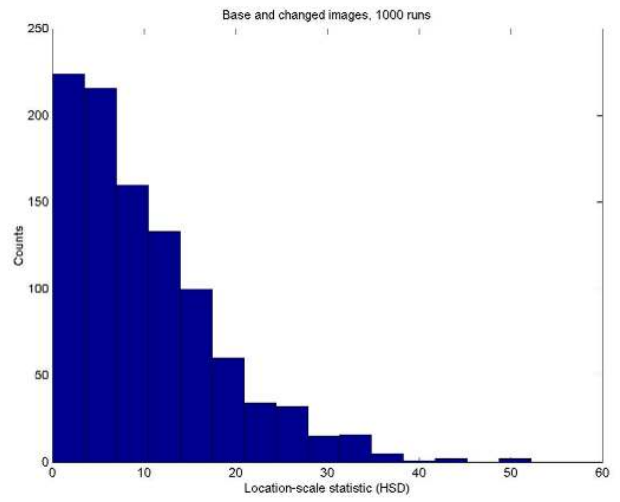
$$S_{CM} = \frac{2\pi^2}{N^2}\left[\sum_{i=1}^{N}\sin\left(\pi\frac{i}{N}\right)\sum_{j=1}^{i} c(R_i)\right]^2 + \frac{8\pi^2}{N^2}\left[\sum_{i=1}^{N}\sin\left(2\pi\frac{i}{N}\right)\sum_{j=1}^{i} c(R_i)\right]^2.$$

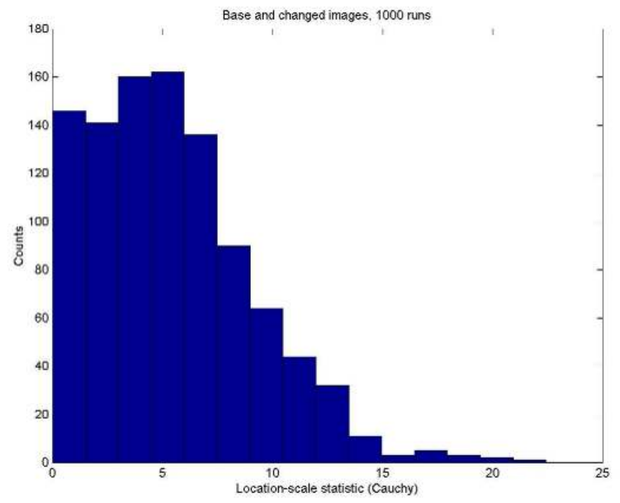
The poster is presented at IBC'06, Montreal, Canada.

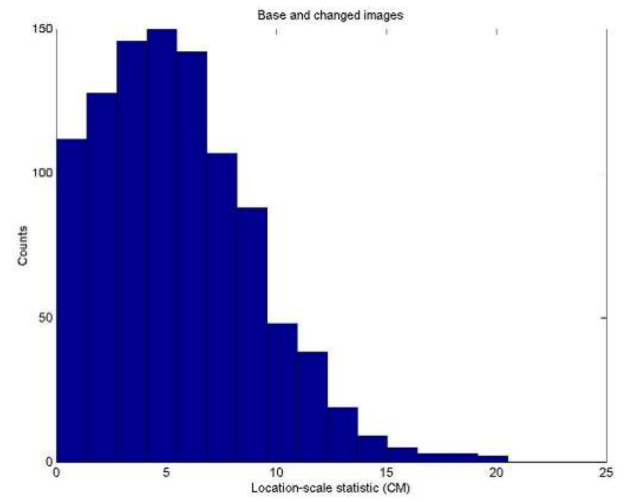
Numerical example



In this example, all images are 360×360 pixels in size and contain the same scene. The only difference in the images is the distribution of the intensity of the red channel. We want to compare the intensity of the red channel of these images by selecting small samples of pixels from the base (top left) and location-scale changed (bottom right) images. The exact distributions of the rank statistics for comparison of these images basing on 20 pixels from each of the images are shown below, for $S(t=-\pi/2)$. $S(t=-\pi)$ and S_{CM} correspondingly. Although all three tests work well, S_{CM} is the best option. The test based on $S(t=-\pi/2, HSD)$ performs almost as well. The test based on $S(t=-\pi/2, Cauchy)$ loses its power due to the presence of the second mode.







Conclusions

The new location-scale test is fairly robust to distributional misspecification, as long as the order of tail is correct. However, the test is sensitive to the presence of outliers under the scale-only alternative. To overcome this lack of robustness in practice, we suggest applying the location-scale test built on the first two components of the Cramer-von Mises statistic, S_{CM}, to heavy-tailed distributions, and truncating the scale score function at certain points for normal-like and light-tailed distributions.

References

1. Vaughanm D.C., The generalized secant hyperbolic distribution and its properties. *Comm. Statist. Theory Methods*, 31(2):219-238, 2002.
2. Kravchuk, O.Y., R-estimator of location of the GSHD. *Comm. Statist. Simulation Computation*, 35(1):1-18, 2006.
3. Kravchuk, O.Y. Two-sample scale rank procedures optimal for the GSHD, *Proceed. of ICSMRF*, Hawaii, January, 2006.