On the use of aggregation functions within tests of symmetry*

Marina Iturrate-Bobes[0009-0005-2616-5772] and
Raúl Pérez-Fernández[0000-0001-6966-1681]

Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Spain
{iturratemarina,perezfernandez}@uniovi.es

Abstract. Aggregation functions are a common tool in Statistics for constructing data summaries. One of the most prominent types of data summaries are the so-called skewnesss coefficients, which measure the degree of asymmetry of a probability distribution. In most cases, these skewness coefficients may be used for defining a test of symmetry. In the present work, some experiments are carried out comparing the performance of the tests of symmetry associated with a popular family of skewness coefficients introduced by Hinkley.

Keywords: Aggregation function · Symmetry · Skewness coefficient · Bootstrap

1 Introduction

Aggregation functions [1] have been widely used in the field of Statistics for constructing different data summaries. Aside of direct applications of an (averaging) aggregation function as an estimator of location, estimators of other population parameters such as scale, skewness and kurtosis are usually constructed by means of one or more aggregation functions. In this work, we deal with the use of aggregation functions for the construction of skewness coefficients allowing to measure the degree of asymmetry of the probability distribution of a real-valued random variable [4]. In particular, the median and other Ordered Weighted Averaging (OWA) operators are here used for constructing skewness coefficients.

For most choices of aggregation functions, the probability distribution of the sample version of the associated skewness coefficient is proven to be asymptotically normal and centered around zero for symmetric probability distributions [3]. This implies that in such case one could easily define a natural test of symmetry, assuming the asymptotic variance can be estimated. Here, we explore the use of the bootstrap for such purpose and define several tests of symmetry associated with different skewness coefficients constructed from popular aggregation functions. In particular, we consider the family of skewness coefficients introduced by Hinkley [2]. These tests of symmetry are compared in terms of

* This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "FEDER/UE".

preservation of the significance level under symmetric distributions, power under asymmetric distributions and robustness in the presence of outliers.

2 Tests of symmetry

If X is a continuous random variable with a density function f, then X is said to be symmetric about x_0 if it holds that $f(x_0 - x) = f(x_0 + x)$ for any $x \in \mathbb{R}$. Note that the median (if it is unique) and the mean (if it exists) of a symmetric random variable coincide and are necessarily the point at which the symmetry occurs.

Oja [4] introduced some desirable properties for a coefficient measuring the degree of asymmetry of a random variable, thus resulting in the formalization of the so-called *skewness coefficients*. More specifically, a skewness coefficient is any function $\gamma : \mathcal{F} \to \mathbb{R}$ (where $\mathcal{F}$ denotes a set of random variables) that satisfies the following properties:

(i) $\gamma(X) = 0$, if X is symmetric;
(ii) $\gamma(a\,X + b) = \gamma(X)$, for any $a > 0$ and $b \in \mathbb{R}$;
(iii) $\gamma(-X) = -\gamma(X)$.

Due to property (i), any skewness coefficient can be used for constructing a symmetry test by establishing some hypothesis on the value of the chosen skewness coefficient γ by considering $H_0 : \gamma(X) = 0$ versus $H_1 : \gamma(X) \neq 0$. It is important to remark that the null hypothesis actually is more general than that of symmetry since there may exist distributions for which the skewness coefficient takes the value zero but are not symmetric.

In order to construct the rejection region of the test of symmetry, it is necessary to know the distribution of the chosen skewness coefficient under the null hypothesis of symmetry. Luckily, for most skewness coefficients it holds that

$$\sqrt{n}\,\frac{\hat{\gamma} - \gamma_F}{\sqrt{V(\gamma, F)}} \underset{\mathcal{L}}{\rightsquigarrow} N(0,1)\,, \tag{1}$$

where $\hat{\gamma}$ is the sample version of the skewness coefficient γ, γ_F is the population value for the skewness coefficient γ at the distribution F, and $V(\gamma, F)$ denotes the asymptotic variance of γ at the distribution F. Note that $\gamma_F = 0$ for all symmetric distributions, however the asymptotic variance $V(\gamma, F)$ is dependent on the underlying distribution and may vary from one symmetric distribution to another one.

If the asymptotic distribution in Equation (1) holds for the chosen skewness coefficient, a rejection region for a test of symmetry associated with γ may be defined as follows:

$$RR = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; |\gamma(\mathbf{x}) - \gamma_0| > \frac{\sqrt{V^*(\gamma, \mathbf{x})}}{\sqrt{n}}\, z_{1-\alpha/2} \right\},$$

where $z_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of a normal distribution and $V^*(\gamma, \mathbf{x})$ is the bootstrap estimation of $V(\gamma, F)$, as proposed in [3].

In the following, we compare the performance of the tests of symmetry associated with the skewness coefficients introduced by Hinkley [2], parameterized by a given $p \in]0, 0.5[$, as follows:

$$\gamma_p(X) = \frac{\big(C_{1-p}(X) - C_{0.5}(X)\big) - \big(C_{0.5}(X) - C_p(X)\big)}{C_{1-p}(X) - C_p(X)}\,,$$

where $C_q(X)$ represents the quantile of order q of X for any $q \in]0, 1[$. Note that the sample version of the skewness coefficients in this family is obtained by substituting population quantiles by sample quantiles.

3 Experimental analysis

In the following, we provide an experimental analysis of the power of the proposed symmetry tests under different distributions. We consider the significance level $\alpha = 0.05$ and sample sizes $n = 25$ and $n = 200$. The power of the different tests is estimated by using Monte Carlo simulation (10^4 replications), setting the number of bootstrap replications to 10^3.

Table 1 presents the power of the symmetry tests associated with different members of the family of skewness coefficients introduced by Hinkley at four symmetric distributions, eight asymmetric distributions and four contaminated distributions of the form $(1-\varepsilon)N(0,1)+\varepsilon N(5,1)$. The experimental setup follows that of [3], so we refer to said paper for more details. Note that for $n = 25$ there are empty cells for each distribution because bootstrap resampling leads to samples in which many central values are repeated and the skewness coefficients are not properly defined in such a case when considering values of p close to 0.5.

It should be pointed out that the higher the considered value of p is, the lower the obtained powers are. This is because the focal point when considering a value of p close to 0.5 is the central part of the distribution instead of the tails. In summary, lower values of p result in more powerful but less robust tests in the presence of outliers (contaminated distributions).

4 Conclusion

We have compared the power of different tests of symmetry based on the use of a skewness coefficient constructed by using aggregation functions. In the future, we will explore the use of different skewness coefficients constructed from different aggregation functions.

References

1. Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E.: Aggregation Functions. Cambridge University Press, Cambridge (2009)
2. Hinkley, D.V.: On power transformations to symmetry. Biometrika **62**, 101–111 (1975)

3. Iturrate-Bobes, M., Pérez-Fernández, R.: Tests of symmetry based on the bootstrap estimation of the asymptotic variance of a skewness coefficient. Information Sciences **646**, 119378 (2023)
4. Oja, H.: On location, scale, skewness and kurtosis of univariate distributions. Scandinavian Journal of Statistics **8**, 154–168 (1981)

Distribution	n	$\gamma_{0.05}$	$\gamma_{0.10}$	$\gamma_{0.15}$	$\gamma_{0.20}$	$\gamma_{0.25}$	$\gamma_{0.30}$	$\gamma_{0.35}$	$\gamma_{0.45}$	$\gamma_{0.45}$
Normal	25	0.0303	0.0226	0.0165	0.0095	0.0036	0.0003			
	200	0.0419	0.0435	0.0379	0.0371	0.0338	0.0294	0.0208	0.0073	0.0000
Cauchy	25	0.1188	0.0721	0.0494	0.0318	0.0105	0.0020			
	200	0.0682	0.0587	0.0549	0.0501	0.0408	0.0318	0.0203	0.0086	0.0000
Laplace	25	0.0501	0.0359	0.0256	0.0164	0.0045	0.0004			
	200	0.0471	0.0472	0.0427	0.0416	0.0374	0.0334	0.0212	0.0067	0.0000
Uniform	25	0.0447	0.0356	0.0259	0.0131	0.0038	0.0006			
	200	0.0584	0.0529	0.0500	0.0425	0.0390	0.0326	0.0223	0.0064	0.0001
Distribution	n	$\gamma_{0.05}$	$\gamma_{0.10}$	$\gamma_{0.15}$	$\gamma_{0.20}$	$\gamma_{0.25}$	$\gamma_{0.30}$	$\gamma_{0.35}$	$\gamma_{0.45}$	$\gamma_{0.45}$
GLD7	25	0.2187	0.1363	0.0798	0.0382	0.0106	0.0023			
	200	0.9581	0.8092	0.6025	0.4000	0.2303	0.1211	0.0565	0.0164	0.0002
GLD8	25	0.6244	0.4109	0.2445	0.1184	0.0351	0.0059			
	200	1.0000	0.9994	0.9804	0.8657	0.6180	0.3350	0.1326	0.0304	0.0002
GLD9	25	0.2021	0.1209	0.0692	0.0308	0.0067	0.0007			
	200	0.9437	0.8190	0.6177	0.3986	0.2258	0.1144	0.0494	0.0122	0.0000
GLD10	25	0.3273	0.1998	0.1112	0.0494	0.0133	0.0017			
	200	0.9933	0.9530	0.8231	0.5929	0.3488	0.1770	0.0738	0.0156	0.0001
GLD11	25	0.0589	0.0358	0.0221	0.0143	0.0038	0.0006			
	200	0.1506	0.1195	0.0928	0.0681	0.0492	0.0356	0.0241	0.0075	0.0001
GLD12	25	0.1474	0.0889	0.0527	0.0266	0.0078	0.0012			
	200	0.7222	0.6029	0.4268	0.2720	0.1587	0.0876	0.0373	0.0096	0.0000
GLD13	25	0.7845	0.5790	0.3739	0.1866	0.0537	0.0084			
	200	1.0000	1.0000	0.9978	0.9626	0.7959	0.4835	0.2024	0.0418	0.0000
GLD14	25	0.8201	0.6229	0.4116	0.2141	0.0702	0.0112			
	200	1.0000	1.0000	0.9990	0.9743	0.8322	0.5259	0.2207	0.0456	0.0005
ε	n	$\gamma_{0.05}$	$\gamma_{0.10}$	$\gamma_{0.15}$	$\gamma_{0.20}$	$\gamma_{0.25}$	$\gamma_{0.30}$	$\gamma_{0.35}$	$\gamma_{0.45}$	$\gamma_{0.45}$
0.01	25	0.0370	0.0237	0.0207	0.0104	0.0032	0.0008			
	200	0.0489	0.0454	0.0426	0.0367	0.0308	0.0285	0.0194	0.0077	0.0000
0.05	25	0.1569	0.0626	0.0307	0.0148	0.0047	0.0004			
	200	0.4221	0.1070	0.0793	0.0556	0.0430	0.0323	0.0224	0.0080	0.0000
0.10	25	0.4121	0.2307	0.1094	0.0414	0.0097	0.0016			
	200	0.9873	0.5615	0.2132	0.1351	0.0772	0.0435	0.0254	0.0070	0.0000
0.20	25	0.7327	0.6015	0.4218	0.2222	0.0619	0.0080			
	200	1.0000	0.9998	0.9746	0.7000	0.3170	0.1610	0.0586	0.0120	0.0002

Table 1. Power of the the symmetry tests associated with different skewness coefficients at different distributions.