Discussion Papers in Economics

No. 03/2018

Tests on Asymmetry for Ordered Categorical Variables

Ingo Klein
University of Erlangen-Nürnberg

Monika Doll
University of Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/

Tests on Asymmetry for Ordered Categorical Variables

Ingo Klein[a,], Monika Doll[a,]

[a] *University of Erlangen-Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany*

Februar 7, 2018

Abstract

Skewness is a well-established statistical concept for continuous and to a lesser extent for discrete quantitative statistical variables. However, for ordered categorical variables almost no literature concerning skewness exists, although this type of variables is common for behavioral, educational, and social sciences. Suitable measures of skewness for ordered categorical variables have to be invariant with respect to the group of strictly increasing, continuous transformations. Therefore, they have to depend on the corresponding maximal-invariants. Based on these maximal-invariants we propose a new class of skewness functionals, show that members of this class preserve a suitable ordering of skewness and derive the asymptotic distribution of the corresponding skewness statistic. Finally, we show the good power behavior of the corresponding skewness tests and illustrated these tests by applying real data examples.

Keywords: Ordered Categorical Variables, Skewness Analysis, Skewness Ordering, Maximalinvariants

---

Email addresses: `ingo.klein@fau.de` (Ingo Klein), `monika.doll@fau.de` (Monika Doll)

1. Introduction

Ordered categorical data become more and more important for various fields of research. Thus, for political and medical research non-income dimensions such as patient life quality, life satisfaction, political philosophy, or level of education are of increasing interest. In the business sector, customers are often asked to rate a product's quality, their customer satisfaction, or the quality of an order process, with companies adjusting their processes and strategies according to those answers. In online trading, recommender systems rely on customers' evaluation of a product to be able to offer and advertise on a more individualized level. Most of the time, all of these variables are measured using k-point rating scales. Answers to those scales lead to rating data, thus ordered categorical data, which are characterized by a fixed number of ordered categories whose differences have no meaningful interpretation. Moreover, ties occure frequently for this kind of data as the amount of categories is much smaller than the amount of observations. However, analyzing rating data reveals that the median is often the middle category choosable, due to social desirability. In general, the median, besides being a location measure provides also information on skewness. Therefore, the median increases the more skewed to the left a distribution is, showing that the concept of location is not completely independent from the concept of skewness. However, for ordered categorical data the same median can occure for totally differently skewed datasets. Thus, this paper claims that for ordered categorical variables skewness is more informative than location, as the concept of skewness unfolds more insights of the whole distribution of this kind of data. However, certain skewness properties are demanded for measuring skewness and a formal ordering of distributions with respect to skewness is needed for comparing ordered categorical datasets by skewness.

Several aspects of ordered categorical variables like dispersion, bipolarization, inequality, or peakedness are discussed in the literature, while with few exceptions only location measures are regarded. Moreover, to our knowledge, no discussion exists on what kind of skewness orderings ordered categorical data should preserve. As differences of ordered categorical variables' values cannot be meaningfully interpreted, orderings and measures for these variables can only depend on cumulated frequencies but not on the measured values themselves. Thus, from the very beginning, this important fact excludes the most well-known and prominent measures of skewness like the third standardized moment, measures based on the differences of quantiles, and L-moments. Moreover, orderings like the convex ordering and weaker alternatives proposed by Oja (1981) are also not approriate for ordered categorical variables. Therefore, this paper proposes new measures of symmetry and new tests on asymmetry, which will be based on a new formal skewness ordering for ordered categorical variables. This new formal skewness ordering is based on the previous work of Klein (1999).

The paper is structered as follows: First, a literature review on measurement and testing of skewness will be given which shows that all formal definitions of skewness are based on an ordering of skewness that depends on quantile functions, which however is only unique for continuous distribution functions. Second, a new skewness ordering for ordered categorical variables will be introduced, which is restricted to a fixed number of categories. Third, new skewness functionals defined as weighted sums of cumulated probabilities will be proposed that provide new linear skewness statistics by applying them to observed cumulated frequencies. Fourth, distribution properties of theses statistics will be shown and hypotheses of symmetry and skewness will be derived from the new skewness ordering. Fifth, alternative skewness tests for one-sided alternatives and their properties on discrete data will be presented. Sixth, the power of tests on symmetry will be compared on discrete variables in the one and two sample cases, showing that the newly introduced test can be optimal in the sense of Neyman and Pearson for simple hypotheses.

2. Literature Review

This section provides a short literature review on measures of and tests on skewness for continuous, discrete, and ordered categorical types of data. It will be shown that for continuous variables major research on various measures and concepts of skewness exist, that for discrete data rarely now literature on skewness can be found, and that for ordered categorical data the few existing literature mostly describes descriptive properties of skewness.

First, we focus on the concept of skewness regarding continuous data. In this case, continuity, by a support on an infinite number of possible values, prevents the occurence of ties. So far, literature considered generation of skewed distributions, modelling using skewed distribution, the measurement of skewness, skewness orderings, and statistical inference on continuous data. Thus, these skewness considerations can be taken as being the basis of skewness discussions for other types of data. Generating skewed distributions by a half split of the scale parameter was proposed by Fechner (1897). As an alternative to this Fechner approach, Tukey (1960), Azzalini (1985), Fernández and Steel (1998), and Arellano-Valle et al. (2005) used certain transformations to generate skewness. Theodossiou (1998) and Grottke (2002) applied the Fechner approach to the t distribution, which provided generalized skewed t distributions. Ferreira and Steel (2006) proposed a general probability transformation for generating univariate skewed distributions. Modelling data that follows a skewed distribution was achieved by Pearson (1895) by applying the Gamma distribution. Thus, Pareto (1897) used skewed distributions to model economic variables like income. Regarding the measurement of skewness, Pearson (1895) introduced the standardized distance between mean and modus, while Yule (1912) used the standardized distance between mean and median. Charlier (1905) and Edgeworth (1904) independently from each other proposed the standardized third moment as a measure of skewness and Bowley (1920) defined a quantile based measure, which is more robust with respect to outliers. All of these proposals of skewness measures follow a pragmatic way of introducing a statistical concept by defining skewness as the value of skewness measurement. Considerations on the ordering of skewness of absolutely continuous variables were made by van Zwet (1964) proposing a so-called convex ordering that reasonable skewness measures have to preserve. Oja (1981) showed that the third standardized moment preserves this convex ordering and discussed weaker orderings. Moreover, he demonstrated in which way the theory of measurement can be applied to define several statistical concepts, including skewness. MacGillivray (1986) and Arnold and Groeneveld (1995) considered further orderings and axiomatizations. Klein and Fischer (2006) present that splitting the scale parameter provides a skewness functional preserving the convex ordering of van Zwet (1964). Concerned with inferencial aspects of skewness were Rayner et al. (1995), Tabor (2010), and Doane and Seward (2011) at comparing the power of various tests on skewness. Premaratne and Bera (2005) proposed a score test on asymmetry. Inferential aspects of skewness for continuous distribution functions were also considered in nonparametric statistics. As for the parametric statistics, the assumption of continuity prevents the occurrence of ties and insures that the test statistics are distributional free under the null hypothesis of symmetry. Apart from this assumption nonparametric statistics does not require quantitative data. Proposed tests on symmetry mainly differ in the regarded alternative. Lehmann (1953) considered a special alternative to symmetry and shows that the Wilcoxon signed rank test is locally most powerful for this alternative. Hájek and Sidák (1967) discussed signed rank tests for the null hypothesis of symmetry considering location shifts as alternative. Yanagimoto and Sibuya (1972) introduced the concept of 'positively biasedness' as an alternative to symmetry, while their hypothesis is based on a skewness ordering that is not relying on the use of the quantile function. Ley and Paindaveine (2009) considered an optimal rank test for the alternative of Ferreira & Steel asymmetry, and Cassart et al. (2008) introduced a new rank test for the Fechner asymmetry. Thas et al. (2005) proposed new one sample tests for symmetry in a specifically constructed contingency table based on the second and third components of Pearson's χ^2 statistic.

Next, we focus on the concept of skewness regarding discrete quantitative data. In this case, variables can only take on a certain number of values. With a finite support they are a special case of ordered categorical variables. For discrete distributions, Hotelling and Solomons (1932) derived limits for the standardized difference of mean and median. Majindar (1962) and Groeneveld (1991) refined and generalized their result. Especially the discussion of Groeneveld (1991) concerned the skewness of Poisson, Binomial, Geometric, and Zipf distributions. von Hippel (2005) showed that the comparison of mean, median, and mode can lead to a wrong decision on what kind of skewness is underlying in the distribution. Rohatgi and Székely (1989) derived inequalities about skewness and kurtosis. Referring inference on skewness for discrete data, it is worth to note that almost all publications that discuss tests on skewness, consider locations shifts as alternative not asymmetry. Nonparametric tests on symmetry with only a small number of ties were discussed by Pratt (1959), Chanda (1963), Conover (1973), and Vorlicková (1972). To our knowledge, skewness properties and skewness orderings were not considered for discrete distributions, yet. One reason could be that for discrete data the concepts of location and dispersion are not easily separable from the concept of skewness as it is in location-scale families.

Finally, we focus on the concept of skewness regarding ordered categorical data. With a fixed number of categories, this kind of variables are discrete ones while the distances between the categories are not interpretable in a meaningful manner. For measuring skewness on ordered categorical variables, it was already mentioned that those measures rely on probabilities under mild invariance requirements. Therefore, corresponding skewness statistics can only depend on observed frequencies following a multinomial distribution. Klein (2001) considered skewness measures for ordered categorical variables without discussing their stochastic properties. For inferencial aspects, Bowker (1948) discussed skewness in the context of contingency tables with the help of the χ^2 statistic. Bhapkar (1961) and Bhapkar (1966) developed a generalized likelihood ratio test for linear restrictions on the probability vector of a multinomial distribution while special linear restrictions provide the concept of symmetry. Robertson and Wright (1981) tested against order restrictions of the probability vector. Dykstra et al. (1995), Dykstra et al. (1995b), and Bhattacharya (1997) provided likelihood ratio tests on skewness. Bhattacharya and Nandram (1996) proposed a Bayes procedure for inference concerning multinomial populations under stochastic ordering.

Thus, for ordered categorical variables, a discussion of skewness properties and skewness orderings as well as a skewness functional that only depends on cumulated frequencies is needed.

3. Skewness Properties and Skewness Ordering

The theory of measurement states a statistical measure to be a structure preserving map of a basic qualitative structure into the equivalent quantitative structure defined on the real numbers. By applying a structure preserving map it is ensured that assertions on the quantitative structure have a meaning in the context of the underlying qualitative structure. For the concept of skewness this rather fundamental way of measuring statistical concepts means to determine skewness properties for indicating a distribution to be symmetric or skewed to the left (right) and to determine an ordering of skewness for comparison of different distributions. In addition, it means to define a measure of skewness as a map from the basic relations into the real numbers such that the skewness properties and the skewness ordering are preserved.

3.1. Skewness Properties

Accordingly, an appropriate measure of skewness should take the value 0 if the distribution is symmetric, the value -1 if it is extremeley skewed to the left, and the value +1 if the distribution is extremely skewed to the right to preserve the skewness properties.

Regarding ordered categorical variables, skewness properties can only rely on a vector of probabilities, as no meaningful interpretation of the values' differences can be made. Thus, let $p = (p_i, ..., p_k)$ be a vector of probabilities, then symmetry can be described by

$$p_i = p_{k+1-i}, \quad i = 1, 2, \ldots, k.$$

Let P_i denote the cumulative probabilities with $P_i = \sum_{j=1}^{i} p_j$, then symmetry can alternatively be described by

$$P_i = 1 - P_{k-i} \text{ or } Q_i = \frac{P_i + P_{k-i}}{2} = \frac{1}{2}, \qquad i = 1, 2, \ldots, k-1.$$

Using P_i and Q_i, $i = 1, 2, \ldots, k-1$, a distribution can be defined as being skewed to the left or skewed to the right. In the following, let $l = [k/2]$ be the integer part of $k/2$ (see Klein (2001) and Dykstra et al. (1995)).

Definition 3.1. *Let $p = (p_1, p_2, \ldots, p_k)$ with $\sum_{i=1}^{k} p_i = 1$. Then p is called skewed to the left (right) if*

$$Q_i \leq \frac{1}{2} \;\; (Q_i \geq \frac{1}{2}), \;\; i = 1, 2, \ldots, l = [k/2]. \tag{1}$$

Skewness to the left (right) is as well called negative (positive) skewness. Extremely negative skewness is characterized by $p_i = 0$, $P_i = 0$ and $Q_i = 0$ for $i = 1, 2, \ldots, k-1$. Extremely positive skewness is characterized by $p_1 = 1$, $P_i = 1$ and $Q_i = 1$, $i = 1, 2, \ldots, k-1$. Moreover, defintion 3.1 is identical with the characterization of asymmetry by Dykstra et al. (1995) in case of k being odd and mass points being $-l, -l-1, \ldots, -1, 0, 1, \ldots, l$ instead of $1, 2, \ldots, k$.

In analogy to Dykstra et al. (1995) we can consider definitions of asymmetry that are stronger than (1).

Lemma 3.1. *Let $p = (p_1, ..., p_k)$ be a vector of probabilities. p is skewed to the left if*

$$p_1 \leq p_2 \leq \ldots \leq p_k, \tag{2}$$

or

$$p_i \leq p_{k+1-i}, \quad i = 1, 2, \ldots, l = [k/2]. \tag{3}$$

Obviously, (2) is stronger than (3), while both skewness definitions are stronger than (1) as examples 3.1 and 3.2 show.

Example 3.1. *We consider a discrete random variable following a binomial distribution with parameters $n = 6$ and success probability $p_0 = 2/3$. This random variable has 7 ordered categories illustrating a so called 7-point rating scale, which is often used in questionnaires. Here, $Q_i < 1/2$, $i = 1, 2, \ldots, 6$ characterizing the vector of probabilities p to be skewed to the left by (1). Additionally, $p_i < p_{8-i}$ for $i = 1, 2, 3 = l = [7/2]$ characterizing the vector of probabilities p to be skewed to the left by (3). However, $p_5 \nleq p_6$, such that p cannot be characterized to be skewed to the left by (2).*

Table 1: Probabilities from a binomial distribution with parameters $n = 6$ and success probability $2/3$

category i	p_i	P_i	Q_i
1	0.001	0.001	0.457
2	0.017	0.018	0.334
3	0.083	0.101	0.211
4	0.219	0.320	0.211
5	0.329	0.649	0.334
6	0.263	0.912	0.457
7	0.088	1.000	

Thus, example 3.1 showed that skewness property definitions (1) and (3) are preserved, while (2) is not.

Example 3.2. *For $p = (0.1, 0.3, 0.2, 0.1, 0.3)$ the vector of cumulated probabilities is $P = (0.1, 0.4, 0.6, 0.7, 1.0)$ and $Q = (0.1, 0.2, 0.2, 0.3, 0.2)$.*

Here, $Q_i < 1/2$ for $i = 1, 2 = l = [5/2]$, while $p_2 > p_{6-2=4}$. Therefore p is characterized to be skewed to the left with respect to (1), but p is not characterized to be skewed to the left if skewness is defined by (3).

Finally, example 3.3 shows that there are vectors of probabilities that can neither be classified as being skewed to the left nor skewed to the right.

Example 3.3. *We choose a vector of probabilities p such that categories 1 and 6 are both modi of this distribution. Here, p cannot be classified as skewed to the left or skewed to the right by either skewness definition. Regarding skewness definition (1), no characterization of skewness can be made as $Q_1 = (P_1 + P_6)/2 > 0.5$ while $Q_2 = (P_2 + P_5)/2 < 0.5$. Regarding skewness definition (2), no characterization of skewness can be made as $p_5 \leq p_6$, while $p_1 \nleq p_2$. Regarding skewness definition (3), no characterization of skewness can be made as $p_2 \leq p_6$, while $p_1 \nleq p_7$.*

Table 2: Probabilities p from a distribution such that p is neither skewed to left nor skewed to the right

category i	p_i	P_i	Q_i
1	0.214	0.214	0.572
2	0.071	0.286	0.393
3	0.071	0.357	0.393
4	0.071	0.429	0.393
5	0.071	0.500	0.393
6	0.429	0.929	0.572
7	0.071	1.000	

3.2. Skewness Orderings

Applying a structure preserving map for the concept of skewness means to not only determine skewness properties but also a well defined ordering of skewness so that statistical hypotheses can be formulated and distributions can be compared with respect to skewness (MacGillivray (1986)). No approach exists yet to define a skewness ordering by applying the theory of measurement in the sense of Oja (1981) for ordered categorical variables with a fixed number of categories. Several skewness orderings were proposed explicitly or implicitly based on well defined and unique quantile functions, which requires (absolute) continuous distribution functions (see e.g. van Zwet (1964), Oja (1981), MacGillivray (1986), Arnold and Groeneveld (1995)). Thus, even for the case of discrete quantitative random variables these orderings are not suitable. Therefore, an alternative ordering of skewness is needed in the case of discrete distributions and especially in the case of ordered categorical random variables.

Implicitly, Dykstra et al. (1995) consider a skewness ordering for discrete distributions by restricting their discussion of several alternative hypotheses of asymmetry to regarding a random variable X that is symmetrically distributed around 0. Here, symmetry can be characterized not considering the quantile function but the distribution or probability mass function by the fact that X and $-X$ are identically distributed. Thus, special kinds of asymmetry are given if X is stochastically greater than $-X$ or $P(X = j) \geq P(X = -j)$, $j > 0$. This definition of asymmetry can be generalized to a skewness ordering for ordered categorical variables as shown by Klein (1999). A supposed shortcoming of this definition is that only vectors of probabilities with equal length can be compared, which corresponds to the fact that only measurements on the same rating scale are comparable.

Definition 3.2. *Let $p_i, p'_i \geq 0$, $i = 1, 2, \dots, k$ with $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} p'_i = 1$. P_i, P'_i, $i = 1, 2 \dots, k$ are the corresponding cumulative probabilities and $Q_i = (P_i + P_{k-i})/2$, $Q'_i = (P'_i + P'_{k-i})/2$, $i = 1, 2, \dots, k-1$. Then, $p = (p_1, p_2, \dots, p_k)$ is called more skewed the left than $p' = (p'_1, p'_2, \dots, p'_k)$ (shortly: $p \preceq p'$) $\iff$*

$$Q_i \leq Q'_i, \quad i = 1, 2, \dots, l = [k/2]. \tag{4}$$

The following lemma states that several sufficient conditions for a vector of probabilities exist to preserve (4).

Lemma 3.2. *Let $p_i, p'_i \geq 0$, $i = 1, 2, \dots, k$ with $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} p'_i = 1$. P_i, P'_i, $i = 1, 2 \dots, k$ are the corresponding cumulative probabilities and $Q_i = (P_i + P_{k-i})/2$, $Q'_i = (P'_i + P'_{k-i})/2$, $i = 1, 2, \dots, l = [k/2]$. Then $Q_i \leq Q'_i$, $i = 1, 2, \dots, l$ if*

1. $P_i \leq P'_i, \quad i = 1, 2, ..., k-1$

or

2. $$q_i \leq q'_i, \quad i = 1, 2, ..., l \tag{5}$$

with $q_i = (p_i + p'_{k+1-i})/2, \quad q'_i = (p'_i + p_{k+1-i})/2.$

The second condition (5) is the well-known stochastic ordering of distributions. For two distributions with the same finite support the distribution that is stochastically larger automatically is more skewed to the left. Condition (5) obviously defines a stronger ordering than (4) because

$$Q_i \leq Q_i' \iff \sum_{j=1}^{i}(q_i - q_i') \leq 0$$

for $i = 1, 2, \ldots, l$.

Example 3.4. *Consider the simple example*

$$(P_1, P_2, P_3) = (0.1, 0.3, 0.7) \;\; and \;\; (P_1', P_2', P_3') = (0.15, 0.35, 0.66).$$

The corresponding Q_i and Q_i' are

$$(Q_1, Q_2, Q_3) = (0.4, 0.3, 0.4) \;\; and \;\; (Q_1', Q_2', Q_3') = (0.405, 0.35, 0, 405).$$

This means that $P_3' < P_3$, while $Q_i \leq Q_i'$, $i = 1, 2, \ldots, k-1$.

Klein (2001) showed that the third standardized moment does not preserve skewness ordering (4) for integer-valued random variables.

4. New Class of Skewness Measures and Functionals

Appropriate skewness measures and functionals for ordered categorical variables should preserve the skewness properties and skewness orderings presented in section 3.

4.1. Skewness Measures

Classical measures of skewness like Pearson's skewness measure as standardized distance betweeen mean and modus or the standardized third moment rely on a meaningful interpretation of the values' differences. However, for ordered categorical variables such a meaningful interpretation of the values' differences is missing. Therefore, these classical measures are not suitable for ordered categorical variables, because the corresponding ordinal scale can be transformed by an arbitrarily chosen strictly increasing transformation, which changes the values' differences.

Klein (1994) gives a formal proof that measures of skewness for ordered categorical variables can only depend on cumulated probabilities but not on actual values u_i, $i = 1, 2, \ldots, k$. The only necessary assumption is that the skewness measure is comparison invariant for all strictly increasing transformations and absolute invariant with respect to translations. This restricts an appropriate measure of skewness for ordered categorical variables to the class of measures that depend on cumulated probabilities.

Moreover, a measure of skewness's definition is required to include a suitable skewness ordering (Oja (1981)). Such a suitable skewness ordering '$\preceq$' was presented in section 3 as (4) or (5). Consequently, a suitable measure of skewness for ordered categorical variables can be defined as follows.

Definition 4.1. *$S : A \to \mathbf{R}$ is a measure of skewness with $A = \{p \in [0,1]^k | p'\iota = 1\}$ for $\iota'(1, 1, \ldots, 1)$ iff:*

1. *Let $p = (p_1, p_2, \ldots, p_k) \in A$ and $p' = (p_k, p_{k-1}, \ldots, p_1)$. Then $S(p) = -S(p')$.*

2. *Let $p, p' \in A$ with $p \preceq p'$. Then $S(p) \leq S(p')$.*

Klein (2001) proved that a special class of so-called linear skewness functionals are skewness measures in the sense of definition 4.1 if skewness ordering (4) is applied. Functionals of this special class of linear skewness measures have the form

$$S^b(P) = \frac{1}{k-1} \sum_{i=1}^{k-1} b(Q_i) \tag{6}$$

with a generating function b with properties

1. $b : [0,1] \to \mathbb{R}$,

2. b is continuous in $(0,1)$,

3. b is antisymmetric, this means $b(p) = -b(1-p)$ for $p \in (0,1)$,

4. b is constraint and strictly increasing in $[0,1]$.

Without loss of generality we assume that $b(0) = -1$ and $b(1) = 1$ to restrict the domain of S^b to the interval $[-1, 1]$. These limits will be attained by an extremely negatively skewed (-1) or extremely positively skewed $(+1)$ distribution.

Suitable generating functions are easy to find, as is shown in example 4.1.

Example 4.1. *Examples for generating functions b.*

- *Kiesl (2003) discusses a so called position-based measure of dispersion. Transferring the generating function of this measure of dispersion to the skewness measurement situation leads to*

$$b_1(p) = 2(p - 1/2), \quad p \in [0, 1].$$

- *Transferring the generating function that was applied for the measurement of dispersion by Gini (1955) leads to*

$$b_2(p) = \begin{cases} 4(p^2 - 1/4) & \text{for } p \leq 1/2 \\ 4(p - 1/2)(3/2 - p) & \text{for } p > 1/2 \end{cases}.$$

- *In the context of skewness measurement for quantitative variables Premaratne and Bera (2005) discuss the generating function*

$$b_3(p) = \frac{\tan^{-1}(p - 1/2)}{\tan^{-1}(1/2)}.$$

Generating functions $b_1(p)$, $b_2(p)$, and $b_3(p)$ are displayed in figure 4.1.

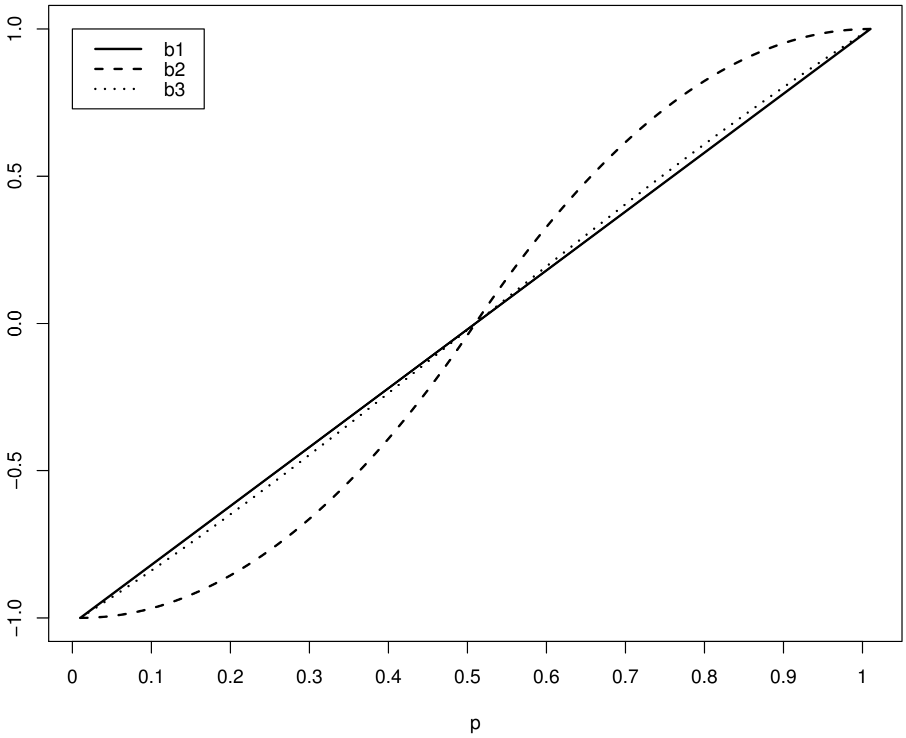

Figure 4.1: The generating functions of example 4.1.

Figure 4.1 shows that the generating functions $b_1(p)$ and $b_3(p)$ are very similar. Due to this, we will only discuss the generating functions $b_1(p)$ and $b_2(p)$ in the following. Moreover, S^{b_1} will be called S^K referring to the generating function applied by Kiesl (2003) and S^{b_2} will be called S^G referring to the generating function applied by Gini (1955).

As the concept of skewness is not completely independent from the concept of location, it is of interest to show that the position-based measure of skewness S^K can be represented as the expectation of an integer-valued random variable. If p_i will be considered as probabilities of an equally spaced integer-valued random variable and due to the fact that arithmetic means can be calculated by integrating the distribution function and the survival function, S^K can be identified as a function of the arithmetic mean.

Lemma 4.1. *Let U be a random variable with support $\{i_1, i_1+1, \ldots, i_1+(k-1)\}$, $i_1 \in \mathbb{Z}$. Let $p_i = P(U = i_1 + (i-1))$, $i = 1, 2, \ldots, k$. $\mu = E(U) = \sum_{i=1}^{k}(i_1 + (i-1))p_i$ is the* *expectation of U. Then it holds*

$$S^K(P) = \frac{2}{k-1}\left(ki_1 + \frac{k-1}{2} - \mu\right). \tag{7}$$

Thus, the skewness measure $S^K(P)$ shows a distribution to be more skewed to the left the bigger the distribution's mean.

4.2. Skewness Functionals

Up to now we defined measures of skewness for a vector of probabilities. For using these measures as test statistics, they have to be applied to the random vector of relative frequencies $f = (n_1/n, \dots, n_k/n)$. Consequenly, the sample measure of skewness itself is a random variable following a certain distribution. This skewness measure explicitly depends on $H_i = (F_i + F_{k-i})/2$, $i = 1, 2, \dots, k-1$ with cumulated frequencies $F_i = \sum_{j=1}^{i} f_j$, $i = 1, 2, \dots, k$.

The linear sample skewness measure has the form

$$S^b(F) = \frac{1}{k-1}\sum_{i=1}^{k-1} b(H_i)$$

with a generating function b. In the following, we call $S^b(F)$ the linear skewness statistic. This statistic as well as the difference $S^b(F) - S^b(F')$ can be used to test hypotheses of symmetry and identical asymmetry against alternatives of asymmetry. Therefore, the exact or limiting distribution of this test statistic under the null hypothesis is needed for the one sample case as well as for the two sample case.

One sample case

The exact distribution of the linear skewness statistic depends on the vector $p = (p_1, \dots, p_k)$ of the population's probabilities. If p is known, the distribution can be calculated by total enumeration or by simulation. For total enumeration all partitions of the number n in k non negative integers has to be generated. Simulation requires drawing samples from the multinomial distribution. For larger values of n and k the calculation of the exact distribution is rather tedious. In this case the limiting distribution of $S^b(F)$ is needed.

Theorem 4.1. *Let $N = (N_1 \ldots, N_k)$ be multinomially distributed with parameters n and $p = (p_1, \ldots, p_k)$ with $\sum_{i=1}^k N_i = n$ and $\sum_{i=1}^k p_i = 1$. Denote $F_i = \sum_{j=1}^i N_i/n$, $P_i = \sum_{j=1}^i p_j$ and $H_i = (F_i + F_{k-i})/2$, $Q_i = (P_i + P_{k-i})/2$, $i = 1, 2, \ldots, k-1$. If $b : [0,1] \to [-1,1]$ is differentiable on $(0,1)$ and $b'(Q_i) \neq 0$ for $i = 1, 2, \ldots, k-1$, then $S_k(N) = \sum_{i=1}^{k-1} b(H_i)$ is asymptotically normally distributed with mean $\sum_{i=1}^{k-1} b(P_i)$ and variance*

$$\sigma_n^2 = \frac{1}{n}\left(\sum_{i=1}^{k-1} P_i(1-P_i)b'(Q_i)^2 + 2\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1} P_i(1-P_j)b'(Q_i)b'(Q_j)\right).$$

This variance can consistently be estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n}\left(\sum_{i=1}^{k-1} F_i(1-F_i)b'(H_i)^2 + 2\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1} F_i(1-F_j)b'(H_i)b'(H_j)\right).$$

See appendix for proof of theorem 4.1. For the asymptotic distribution under the null hypothesis of symmetry $b'(1/2) \neq 0$ has to be required. Considering special cases of generating functions $b(H_i)$, the asymptotic distributions can be formulated as follows.

Example 4.2. *Asymptotic distributions of the linear skewness statistic for special cases of generating functions.*

- *For $b_1(p)$ it is $b_1'(p) = 2 \neq 0$ for $p \in (0,1)$. The asymptotic variance of $S^K(N) = \frac{1}{k-1}\sum_{i=1}^{k-1} 2(H_i - 1/2)$ for $Q_i = 1/2$, $i = 1, 2, \ldots, k-1$ is*

  $$\sigma_n^2 = \frac{4}{n(k-1)^2}\left(\sum_{i=1}^{k-1} P_i(1-P_i) + 2\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1} P_i(1-P_j)\right)$$

- *For $b_2(p)$ it is*

  $$b_2'(p) = \begin{cases} 8p & \text{for } p \in (0,1/2) \\ 8(1-p) & \text{for } p \in (1/2,1) \end{cases}$$

  This derivative does not vanish in $(0,1)$. The asymptotic variance of $S^G(N)$ for $Q_i = 1/2$, $i = 1, 2, \ldots, k-1$ is

  $$\sigma_n^2 = \frac{1}{n(k-1)^2}\left(\sum_{i=1}^{k-1} P_i(1-P_i)b_2'(Q_i)^2 + 2\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1} P_i(1-P_j)b_2'(Q_i)b_2'(Q_j)\right)$$

Two sample case

For the two sample case, a test can be based on the difference $S^b(f_1) - S^b(f_2)$ of the values of a linear skewness statistic S^b for two independent samples with observed frequencies $f_{i'} = (n_{i'1}/n, \ldots, n_{i'k}/n)$, $i = 1, 2, \ldots, k$ and $i' = 1, 2$. $S^b(f_{i'})$ is asymptotically normally distributed. Hence, the difference also asymptotically follows a normal distribution

$$S^b(f_1) - S^b(f_2) \overset{a}{\sim} N\left(0, \sigma^2_{n_1} + \sigma^2_{n_2}\right).$$

4.3. Statistical Hypotheses

Considering the skewness properties and the skewness orderings presented in section 3, statistical hypotheses for one and two sample tests can be derived. These hypotheses can be tested applying the linear rank statistic and its distribution presented in section 4.

For the one sample case, symmetry means that $p_i = p_{k+1-i}$, $i = 1, 2, \ldots, k-1$. Thus, the null hypothesis of symmetry can be stated as

$$H_0: \ Q_i = 1/2, \ \ i = 1, 2, \ldots, l = [k/2].$$

For the two sample case, the null hypothesis of identical symmetry of two populations $p = (p_1, \ldots, p_k)$ and $p' = (p'_1, \ldots, p'_k)$ can be stated as

$$H'_0: \ Q_i = Q'_i \ \ i = 1, 2, \ldots, l = [k/2]$$

or equivalently

$$H'_0: \ q_i = q_{k+1-i}, \ \ i = 1, 2, \ldots, l = [k/2].$$

This means that the null hypothesis of identical skewness of two independent samples can be reformulated as null hypothesis of symmetry for the synthetic vector of probabilities $q = (q_1, \ldots, q_k)$ (see Lemma 3.2 (5)).

Statistical testing distinguishes between so-called one-sided and two-sided alternatives to the null hypotheses of symmetry and identical asymmetry. As two-sided alternatives are of minor interest for a directed concept like asymmetry, we focus on one-sided alternatives. Two-sided alternatives can be discussed in a similar way.

For the one sample case, the alternative hypothesis that $p = (p_1, \ldots, p_k)$ is skewed to the left, or for the two sample case, the alternative hypothesis that p is more skewed to the left than $p' = (p'_1, \ldots, p'_k)$ leads to $H_1 \setminus H_0$ or $H'_1 \setminus H'_0$ with

$$H_1 : \; Q_i \leq 1/2 \text{ for } i = 1, 2, \ldots, l$$
$$H'_1 : \; Q_i \leq Q'_i \text{ for } i = 1, 2, \ldots, l.$$

5. Alternative Skewness Tests for One-Sided Alternatives

To evaluate the performance of tests based on the new class of skewness statistics introduced in section 4, which is so far the only class of skewness statistics suitable for ordered categorical data, size and power comparisons have to be performed applying discrete variables. Discrete quantitative random variables are also ordered categorical random variables. Therefore, alternative skewness tests based on the third standardized moment, linear rank tests, and a likelihood ratio test under order restrictions as well as their distributions in case of discrete quantitative variables are presented in the following.

5.1. Skewness Test Based on The Third Standardized Moment

The asymptotic distribution of the third standardized moment is well known in the case of continuous quantitative random variables. However, its distribution in case of discrete quantitative random variables is required.

One sample case

Definition 5.1. *Let $u_1 < u_2 < \ldots < u_k$ be the realizations of a discrete random variable U. Set $p_i = P(U = u_i)$, $i = 1, 2, \ldots, k$. Then the functional*

$$SK(u, p) = \frac{\sum_{i=1}^{k} (u_i - \mu)^3 p_i}{(\sigma^2)^{3/2}}$$

with

$$\mu = \sum_{i=1}^{k} u_i p_i, \quad \sigma^2 = \sum_{i=1}^{k} (u_i - \mu)^2 p_i$$

is the third standardized moment of the distribution of U.

Applying this functional to the relative frequencies $f = (f_1, \ldots, f_k)$ leads to the third standardized sample moment $SK(u, f)$,which can be used for a test of symmetry. In line with Premaratne and Bera (2005) we call this test $\sqrt{b_1}$ test. Its exact distribution can be derived either by total enumeration or simulation. The asymptotic distribution is given by the following theorem.

Theorem 5.1. *Let $u_1 < u_2 < \ldots < u_k$ be the realizations of a discrete random variable U. Set $p_i = P(U = u_i)$, $i = 1, 2, \ldots, k$. Let $N = (N_1 \ldots, N_k)$ be multinomially distributed with parameters n and $p = (p_1, \ldots, p_k)$ with $\sum_{i=1}^{k} N_i = n$. $f = (f_1, \ldots, f_k) = N/n$. The third standardized sample moment*

$$SK(u, f) = \frac{M_3(u, f)}{(S^2(u, f))^{3/2}}$$

with

$$M_3(u, f) = \sum_{i=1}^{k}(u_i - M_1(u, f))^3 f_i, \;\; S^2(u, f) = \sum_{i=1}^{k}(u_i - M_1(u, f))^2 f_i$$

and $M_1(u, f) = \sum_{i=1}^{k} u_i f_i$ is asymptotically normally distributed with mean $SK(u, p)$ and variance $\nabla' \Sigma \nabla$. ∇ is the gradient vector with elements

$$\frac{\partial SK(u, f)}{\partial f_i} = -3\frac{u_i}{\sqrt{S^2(u, f)}} + \frac{(u_i - M_1(u, f))^3}{S^2(u, f)^{3/2}} - \frac{3}{2} SK(u, f) \frac{u_i^2 - 2M_1(u, f)u_i}{S^2(u, f)},$$

$i = 1, 2, \ldots, k$, and Σ denotes the covariance matrix with elements

$$\sigma_{ij} = Cov(f_i, f_j) = \begin{cases} p_i(1 - p_i)/n & \text{for } i = j \\ -p_i p_j/n & \text{for } i \neq j \end{cases}$$

See appendix for the proof of this theorem.

Two sample case

The difference between two independent samples' third standardized sample moments can be used to test the equality of skewness if u is a k-dimensional vector of discrete values. Therefore, $SK(u, f_1) - SK(u, f_2)$ is asymptotically normally distributed with mean $SK(u, p_1) - SK(u, p_2)$ and variance-covariance matrix $\nabla_1' \Sigma_1 \nabla_1 + \nabla_2' \Sigma_2 \nabla_2$. The elements of $\nabla_{i'}$ are

$$\frac{\partial SK(u, f_{i'})}{\partial f_{i'i}} = -3\frac{u_i}{\sqrt{S^2(u, f_{i'})}} + \frac{(u_i - M_1(u, f_{i'}))^3}{S^2(u, f_{i'})^{3/2}} - \frac{3}{2} SK(u, f_{i'}) \frac{u_i^2 - 2M_1(u, f_{i'})u_i}{S^2(u, f_{i'})},$$

$i = 1, 2, \ldots, k$ and $i' = 1, 2$. The covariance matrix $\Sigma_{i'}$ is given by the elements

$$\sigma_{ij} = Cov(f_{i'i}, f_{i'j}) = \begin{cases} p_{i'i}(1 - p_{i'i})/n & \text{for } i = j \\ -p_{i'i}p_{i'j}/n & \text{for } i \neq j \end{cases}$$

for $i, j = 1, 2, \ldots, k$ and $i' = 1, 2$.

5.2. Skewness Tests Based on Linear Rank Statistics

Signed rank tests were developed for testing the location of symmetric distributions. Hájek and Sidák (1967) titled section III.5.1 of 'Theory of Rank Statistics' as 'Tests of Symmetry', but discussed in this section a test on location shift for families of symmetric distributions. Therefore, Yanagimoto and Sibuya (1972) p. 423, stated explicitly:

> *"The purpose of this paper is to make clear the notion of a 'positively biased' one-dimensional random variable as an alternative to 'symmetric about zero'. This notion is useful to make more precise statements on the test of symmetry than discussed in previous publications [...] ."*

To these previous publications belongs Hájek and Sidák (1967)'s 'Theory of Rank Statistics'. Thus, Yanagimoto and Sibuya (1972) considered 'positively biasedness' as alternative to 'symmetric about zero'.

Let $X_1, \ldots, X_n$ be a simple random sample from the distribution $F(.)$, $|X|^{(1)}, \ldots, |X|^{(n)}$ the ordered statistics of the absolute values of X_i, $i = 1, 2, \ldots, n$ and $Z_i = 1$ $(= -1)$ if $|X|^{(i)}$ corresponds to a positive (negative) X_i, $i = 1, 2, \ldots, n$. Then

$$S = \sum_{i=1}^{n} a(i) Z_i$$

defines a signed rank test statistic with scores $a(i)$, $i = 1, 2, \ldots, n$. Yanagimoto and Sibuya (1972) proved that S can be applied for an unbiased test of the null hypothesis of symmetry against the alternative $H_1 \setminus H_0$ if the scores are nondecreasing, i.e. $a(1) \leq \ldots \leq a(n)$. Examples of nondecreasing scores are Median scores ($a(i) = 1$), Wilcoxon scores ($a(i) = i$), normal scores ($a(i) = \Phi^{-1}(i/(n+1))$, and expected normal scores ($E(\Phi^{-1}(U^{(i)}))$.

An extremely skewed probability vector p contains elements p_i being 0 or 1. Due to this, the score function $a(.)$ should have a constrained range. This excludes e.g. normal scores or expected normal scores. Thus, power comparisons in section 6 between tests based on the new class of linear skewness functionals and tests based on linear rank statistics can be made based on Median and Wilcoxon scores. We restrict our analysis to those based on Wilcoxon scores as the Median test was shown to provide lower power than the Wilcoxon-Mann-Whitney test, even in the double exponential family at small sample sizes (see Freidlin and Gastwirth (2000), Ramsey (1971)). Applying $S^a(.)$ to the vector f of observed frequencies $n_1/n, \ldots, n_k/n$ gives the linear rank skewness statistic. The exact distribution can either be calculated by total enumeration or simulation.

A test based on linear rank statistics with Wilcoxon scores for the one sample case is identical to the Wilcoxon signed rank test. Its limiting distribution is given by a normal distribution with mean $0.25(1 + 1/n)$. In the presence of ties the variance of the limiting distribution can be corrected to achieve a better approximation for small sample sizes:

$$\sigma_W^2 = \frac{n(n+1)(2n+1)}{24n^4} - \frac{1}{48n^4}\sum_{j=1}^{k}((nb_j)^3 - nb_j)$$

with

$$b_j = p_j + p_{k+1-j}, \quad j = 1, 2, \ldots, l$$

(Büning and Trenkler (1994)). A test based on linear rank statistics with Wilcoxon scores for the two sample case is identical to the Wilcoxon-Mann-Whitney test. Adapted for ordered categorical data,

$$W = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi(pn_1, p'n_2)}{n_1 n_2},$$

with $\psi(x_1, x_2) = 1$ if $x_1 > x_2$, $\psi(x_1, x_2) = 0.5$ if $x_1 = x_2$, and $\psi(x_1, x_2) = 0$ otherwise.

This statistic is asymptotically normally distributed with mean 0.5 and variance

$$\sigma_W^2 = \frac{n(1 - \sum_{i=1}^{k}/(t_k^3/n^3))}{12(n-1)npp'},$$

with $n = n_1 + n_2$, $p_i = n_i/n$, $t_i = (pn_1)_i + (p'n_2)_i$, while k denotes the number of categories (see Zhao et al. (2008)).

5.3. *Likelihood Ratio Skewness Test under Order Restrictions*

The general linear hypotheses $H_0: \ p_i = p_{k+1-i}, \ \ i = 1, 2, \ldots, k.$ and $H_0': \ Q_i = Q_i', \ \ i = 1, 2, \ldots, k-1$ against an unspecified alternative can be tested using a generalized likelihood ratio test. To do so, we have to maximize the logarithmic likelihood function under the restriction of the null hypothesis. The Wald test is asymptotically equivalent to the generalized likelihood ratio test and avoids this optimization under restrictions. Tests on categorical data and especially the equivalence of the χ^2- and the Wald test was discussed by Bhapkar (1961) and Bhapkar (1966). For ordered categorical variables we have specified or directed alternatives. For these one-sided alternatives, Dykstra et al. (1995) discussed likelihood ratio tests for symmetry of discrete distributions for the one sample case.

One sample case

Dykstra et al. (1995) discussed discrete distributions with support $\{1, 2, \ldots, 2l+1\}$, hence an odd number $k = 2l + 1$ of mass points. For this special case they discuss an alternative that is equivalent to the skewness property (1):

$$H_1: \; P_i \leq 1 - P_{k-i}, \;\; i = 1, 2, \ldots, l$$

For the test of symmetry H_0 against $H_1 \setminus H_0$ the likelihood ratio statistic

$$T_1 = 2 \sum_{i=1}^{k} n_i (\ln \hat{p}_i^{(1)} - \ln \hat{p}_i^{(0)})$$

is used, with n_i being the sample frequency of i, $\hat{p}_i^0 = (n_i + n_{k+1-i})/(2n)$ being the Maximum Likelihood (ML) estimator of p_i under H_0, and $\hat{p}_i^1$ being the ML estimator of p_i under the restriction of the alternative H_1 for $i = 1, 2, \ldots, k$.

Using results from isotonic regression, the following theorem states states that $\hat{p}_i^1$, $i = 1, 2, \ldots, k$ is given by the least squares projection onto the cone of nondecreasing vectors.

Theorem 5.2. *(Dykstra et al. (1995), p. 722) Let $\hat{p}_i = n_i/n > 0$, $i = 1, 2, \ldots, k$. Set $\hat{p}_- = (\hat{p}_{k+1-i}, \hat{p}_{k-i}, \ldots, \hat{p}_1)$. Then the ML estimator $\hat{p}^1$ of $p = (p_1, p_2, \ldots, p_k)$ subject to the restriction H_1 is given by*

$$\hat{p}^{(1)} = \hat{p} E_{\hat{p}} \left(\frac{\hat{p} + \hat{p}_-}{2\hat{p}} | I \right).$$

$E_w(x|I)$ denotes the least squares projection with weights w of the vector x onto the cone I of nondecreasing vectors.

The least squares projection can be computed by the pool adjacent violators algorithm (PAVA) which is implemented in R (see De Leeuw et al. (2010)). The exact distribution can be derived by total enumeration or simulation. Dykstra et al. (1995) derive the limiting distribution of the likelihood ratio statistic T_1 under the null hypothesis of symmetry if the alternative is $H_1 \setminus H_0$ as a mixing of χ^2 distributions.

Theorem 5.3. *(Dykstra et al. (1995), p. 724) If p satisfies H_0 and $p_i > 0$, $i = 1, 2, \dots, k = 2l + 1$, then*

$$\lim_{n\to\infty} P(T_1 \geq t) = \sum_{i=l+1}^{k} p(i, k, p_r) P(\chi^2_{k-i} \geq t)$$

where $p(0, k, p_r)$ is the probability that $E_{p_r}(V_r|J)$ is identically 0, $p(i, k, p_r)$ for $i = 1, 2, \dots, l$ is the probability that $E_{p_r}(V_r|J)$ has i distinct non-zero values, and p_r is the restriction of p to $\{l + 2, ..., k\}$. Furthermore, the least favorable asymptotic distribution is

$$\sup_p \lim_{n\to\infty} P(T_1 \geq t) = \frac{1}{2} P(\chi^2_{(k-1)/2-1} \geq t) + \frac{1}{2} P(\chi^2_{(k-1)/2} \geq t)$$

Tests on similar alternatives to symmetry as discussed in Dykstra et al. (1995) are provided by Bhattacharya (1997).

Two sample case

Tests on asymmetry for the two sample case considering the stronger alternative hypothesis $p_i < p'_i$ were discussed in Dykstra et al. (1995b). However, we consider a less restrictive alternative and the null hypothesis that two vectors of probability $p = (p_1, \dots, p_k)$ and $p' = (p'_1, \dots, p'_k)$ have the same amount of skewness

$$H'_0: \quad Q_i = Q'_i, \quad i = 1, 2, \dots, l = [k/2].$$

This null hypothesis can be reformulated to a one sample hypothesis

$$q_i = q_{k+1-i}, \quad i = 1, 2, \dots, l$$

with $q_i = (p_i + p'_{k+1-i})/2$, $i = 1, 2, \dots, k$ (see (5)). Due to the skewness ordering (4) p is more skewed to the left if

$$H'_1: \quad Q_i \leq Q'_i, \quad i = 1, 2, \dots, l$$

or equivalently if

$$H_1' : \sum_{j=1}^{i} q_j \leq \sum_{j=1}^{i} q_{k+1-j} \quad i = 1, 2, \ldots, l$$

holds. Thus, the alternative hypothesis for the two sample case is $H_1' \setminus H_0'$. Reformulating the alternative hypothesis into a one sample problem gives the advantage that the one sample likelihood ratio test under order restrictions proposed by Dykstra et al. (1995) can be used for testing. Thus, this likelihood ratio test's procedure and its asymptotic distribution have to be applied to the probability vector q. Without any restriction the ML estimator $\hat{q}_i$ of q_i is $(n_i/n + n'_{k+1-i}/n')/2$ for $i = 1, 2, \ldots, k$. Under H_0', the ML estimator $\hat{q}_i^0$ of q_i is

$$\hat{q}_i^{(0)} = \frac{n_i/n + n'_{k+1-i}/n' + n'_i/n' + n_{k+1-i}/n}{4}$$

for $i = 1, 2, \ldots, k$. From theorem (5.2) we know that the ML estimator of q_i under $H_1' \setminus H_0'$ $\hat{q} = (\hat{q}_i^{(1)}, \ldots, \hat{q}_i^{(1)})$ is

$$\hat{q}^{(1)} = \hat{q} E \left(\frac{\hat{q} + \hat{q}_-}{2\hat{q}} |I \right).$$

The likelihood ratio test statistic for testing H_0' against $H_1' \setminus H_0'$ is given by

$$T_1 = 2 \sum_{i=1}^{k} (n_i + n'_{k+1-i})(\ln \hat{q}_i^{(1)} - \ln \hat{q}_i^{(0)})$$

Its asymptotic distribution is analogous to the one sample case.

6. Power Comparisons

Power comparisons of skewness tests based on the new class of linear skewness function with alternative skewness tests have to be performed on a parametric discrete distribution with suitable support and a well-defined skewness parameter which can be varied. Several such discrete distributions exist.

One possible distribution is the binomial distribution. For this distribution the skewness statistic S^K is uniformly most powerful for one sided alternatives. Let U be distributed according to

$$f(u; k-1, p_0) = \binom{k-1}{u} p_0^u (1 - p_0)^{(k-1)-u}, \quad u = 0, 1, \ldots, k-1$$

with $\mu_k = E(X) = (k-1)p_0$. This distribution is symmetric around the median if $p_0 = 1/2$. For $p_0 > 1/2$ ($p_0 < 1/2$) we get a distribution that is skewed to the left (right). Thus, we can manipulate the skewness of the distribution by manipulating p_0. The corresponding standardized third moment is

$$SK = \frac{1-2p_0}{\sqrt{(k-1)p_0(1-p_0)}}, \; p_0 \in (0,1).$$

The linear skewness function is given by

$$S^K = 1 - 2p_0,$$

due to $\mu = (k-1)p_0$ and the fact that S^K can be represented as the expectation of an integer valued variable (see (7)). Let $X_1, \ldots, X_n$ be a simple random sample from the binomial distribution. Then, the Maximum Likelihood estimator for p_0 is given by

$$\hat{p}_0 = \frac{1}{n(k-1)} \sum_{i=1}^{k} X_i.$$

Due to the invariance property of Maximum Likelihood estimators the ML estimators for SK and S^K are given by

$$\widehat{SK} = \frac{1-2\hat{p}_0}{\sqrt{(k-1)\hat{p}_0(1-\hat{p}_0)}}$$

and

$$\hat{S}^K = 1 - 2\hat{p}_0.$$

Denote N_i as the number of random draws resulting in the realization $i-1$ for $i = 1, 2, \ldots, k-1$. If $F_i = \sum_{j=1}^{i} N_j/n$, $i = 1, 2, \ldots, k-1$, then there is a simple relationship between the ML estimator $\hat{p}_0$ and the linear sample skewness measure $S^K(F)$:

$$\hat{p}_0 = 1 - \frac{1}{k-1} \sum_{i=1}^{k-1} F_i = \frac{1}{2} - \frac{S^K(F)}{2}.$$

$$S^K(p(.;p_0)) = 1 - \frac{2}{k-1}(k-1)p_0 = 1 - 2p_0.$$

This means that the linear sample skewness statistic $S^K(F)$ and the ML estimator $\hat{S}^K$ are identical. Thus, for the binomial distribution a test on symmetry is identical to a test on the probability parameter p_0. The Neyman-Pearson test for this parameter gives a most powerful test for

$$H_0: \; p = p_0 \;\text{ against }\; H_1: \; p = p_1.$$

It is easy to derive that this most powerful test can be based on the ML estimator $\hat{p}_0$. Because the third standardized sample moment $\widehat{SK}$ and the linear skewness statistic $S^K(F)$ are monotone functions of $\hat{p}_0$, the Neyman-Pearson test can be based on these statistics equivalently.

Another parametric discrete distribution with suitable support is the truncated negative binomial distribution, as every discrete distribution with support that is a subset of all integers can be truncated such that the support is restricted to $D = \{i_1, i_1+1, \ldots, i_1+k-1\}$ for $i_1 \in \mathbb{Z}$ and $k \in \mathbb{N}$. Consider the negative binomial distribution with probability mass function

$$f(x; r, p) = \binom{x+r-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \ldots, \quad r > 0, \; 0 < p \leq 1.$$

Let $i_1 = 0$ and

$$P(D) = \sum_{x=0}^{k-1} f(x; r, p)$$

then

$$f(x; r, p | X \leq k-1) = \frac{f(x; r, p)}{P(D)}, \quad x = 0, 1, 2, \ldots, k-1 \tag{8}$$

defines the truncated negative binomial distribution.

Lemma 6.1. *Let X be negative binomially distributed with parameters r, p and $X|X \leq k-1$ be a random variable with probability mass function (8).*

1. *If $\mu = E(X) = r(1-p)/p$ then*

$$E(X|X \leq k-1) = \mu - (k-1+r)\frac{1-p}{p}\frac{f(k-1; r, p)}{P(D)}$$

2. *$E(X|X \leq k-1)$ is monotonic decreasing in p, $p \in (0, 1]$.*

$E(X|X \leq k-1)$ is a skewness functional and the one-to-one relationship ensures that p can also be considered as a skewness parameter.

Next to these distributions, the Zipfian distribution also meets the requirements:

$$f(x; s, N) = \frac{1/x^s}{\sum_{y=1}^{N} 1/y^s}, \quad x = 1, 2, \ldots, k, \ N \in \mathbb{N}, \ s > 0$$

with $y = 1, 2, ...k$. By varying the parameter s, the skewness of the distribution can be manipulated.

Another way of generating skewed distributions is by considering the Lehmann alternative $F(x)^\lambda$ for $\lambda \neq 1$. The Wilcoxon signed rank test is locally optimal for $\gamma F + (1-\gamma)F^2$, with F being the cumulative distribution function, $\gamma \in [0, 1]$. Lehmann (1953) presented this as being the 'simplest choice' of generating a specific alternative to the null hypothesis of equal distributions. Let $p = (p_1, \ldots, p_k)$ be symmetric and $P_i' = P_i^\lambda$, $\lambda > 1$, then

$$Q_i' \leq 0.5, \ i = 1, 2, \ldots, k-1.$$

Thus, $p' = (p_1', \ldots, p_k')$ with $p_i' = P_i^\lambda - P_{i-1}^\lambda$, $i = 1, 2, \ldots, k$, and $P_0 = 0$ is skewed to the left (Lehmann (1953)).

Using these distributions, we compared the power of tests based on the new class of skewness statistics with generating functions b_1 and b_2 to the power of the alternative tests of skewness presented in section 5. As the S^K test is the uniformly most powerful for testing on symmetry for the exponential distribution family on simple hypotheses, power comparisons for the one sample case were not only conducted using binomially distributed samples, but also by generating skewness using the method of Lehmann (1953). Furthermore, for the two sample cases, the tests' power was evaluated using two truncated negative binomially distributed samples as well as two Zipfian distributions. In addition, the skewness parameters were stochastically determined to not only show the new skewness tests' performance on simple, but also on composite hypotheses.

The number of categories for all simulations was set at $k = 7$. Moreover, sample sizes were set to be $m = n = 20$, $m = n = 100$, and $m = n = 1000$ to cover small, moderate and large sample sizes. Simulation repetitions were performed 10 000 times.

6.1. *Power Comparisons for the One Sample Case*

For the one sample case, power comparisons were performed on the binomial distribution and using the Lehmann alternative. Skewness was manipulated for the binomial distribution by varying the parameter p_1 from 0.50 (symmetry) to 0.9 (skewed to the left) (see table 3). For the Lehmann alternative, a random probability vektor p was drawn from a Dirichlet distribution to generate p with hyperparameter $\alpha = (1, 1, ..., 1)$ of length k. Let $p = (p_1, \ldots, p_k) \in \Delta_k$, then $p^s = (p + (p_k, \ldots, p_1))/2$ provides a symmetric probability vector with cumulated probabilties $P_i^s = \sum_{j=1}^{i} p_j^s$. Therefore, $P_i = (P_i^s)^\lambda$, $i = 1, 2, \ldots, k$ is an alternative of 'skewed to the left' to the null hypothesis of symmetry. λ was varied from 1 (symmetry) to 5 (skewed to the left) (see table 4).

Results of the power comparisons for the one sample case demonstrate that the linear skewness tests introduced in section 4 provide good power for the binomially distributed sample (see table 3) as well as for the Lehmann alternativ (see table 4). Table 3 displays that at binomial distributions, the S^K test, by being the uniformly most powerful test, outperforms all alternative tests even at low sample sizes. However, regarding the Lehmann alternative, table 4 displays that the Wilcoxon test's power is slightly higher than that of tests based on the new class of linear skewness functionals. Throughout all conditions for the one sample case, the S^K test and the S^G test show to have similar performances, while the S^K test provides a slightly higher power. As expected, the $\sqrt{b_1}$ test provides the lowest power, as it is influenced by the categories' assigned values.

Table 3: Power comparisons for the one sample case on binomial distribution with $p_0 = 0.5$; k=7

	S^K		S^G		$\sqrt{b_1}$		Wilcoxon		GLR	
p_1	exact	asy.	exact	asy.	exact	asy.	exact	asy.	exact	asy.
					$n = 20$					
0.5	0.059	0.043	0.055	0.028	0.049	0.05	0.051	0.022	0.051	0.032
0.52	0.128	0.099	0.124	0.070	0.056	0.058	0.116	0.054	0.087	0.059
0.55	0.311	0.261	0.306	0.203	0.068	0.069	0.281	0.158	0.199	0.145
0.6	0.729	0.680	0.723	0.604	0.089	0.091	0.672	0.494	0.534	0.448
0.7	0.999	0.998	0.999	0.996	0.155	0.157	0.997	0.983	0.987	0.975
0.8	1.000	1.000	1.000	1.000	0.277	0.278	1.000	1.000	1.000	1.000
0.9	1.000	1.000	1.000	1.000	0.603	0.604	1.000	1.000	1.000	1.000
					$n = 100$					
0.5	0.051	0.048	0.052	0.040	0.051	0.053	0.052	0.037	0.055	0.038
0.52	0.263	0.249	0.261	0.226	0.067	0.068	0.260	0.202	0.198	0.155
0.55	0.793	0.781	0.791	0.756	0.115	0.118	0.786	0.725	0.657	0.596
0.6	1.000	1.000	1.000	0.999	0.218	0.220	1.000	0.999	0.996	0.995
0.7	1.000	1.000	1.000	1.000	0.546	0.550	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	0.910	0.911	1.000	1.000	1.000	1.000
0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					$n = 1000$					
0.5	0.056	0.055	0.055	0.051	0.045	0.047	0.054	0.044	0.054	0.034
0.52	0.927	0.925	0.925	0.919	0.128	0.132	0.920	0.907	0.846	0.793
0.55	1.000	1.000	1.000	1.000	0.388	0.396	1.000	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000	0.873	0.876	1.000	1.000	1.000	1.000
0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Power comparisons for the one sample case with Lehmann Alternative

	S^K		S^G		$\sqrt{b_1}$		Wilcoxon		GLR	
δ	exact	asy.	exact	asy.	exact	asy.	exact	asy.	exact	asy.
					$n = 20$					
1.0	0.056	0.098	0.051	0.060	0.051	0.099	0.052	0.014	0.050	0.053
1.05	0.069	0.119	0.068	0.079	0.067	0.122	0.075	0.021	0.061	0.064
1.1	0.090	0.154	0.089	0.105	0.086	0.153	0.102	0.033	0.074	0.078
1.2	0.158	0.235	0.147	0.172	0.130	0.213	0.170	0.062	0.107	0.112
1.5	0.370	0.501	0.356	0.392	0.294	0.400	0.447	0.236	0.275	0.285
2.0	0.728	0.825	0.697	0.734	0.511	0.614	0.809	0.620	0.629	0.638
3.0	0.948	0.964	0.933	0.942	0.710	0.772	0.980	0.945	0.942	0.945
5.0	0.993	0.997	0.980	0.983	0.812	0.852	0.998	0.996	0.995	0.995
					$n = 100$					
1.0	0.048	0.053	0.046	0.036	0.047	0.077	0.045	0.039	0.048	0.041
1.05	0.092	0.112	0.093	0.078	0.089	0.132	0.102	0.093	0.082	0.070
1.1	0.175	0.196	0.173	0.143	0.144	0.200	0.192	0.177	0.136	0.120
1.2	0.401	0.432	0.399	0.342	0.294	0.370	0.449	0.431	0.317	0.292
1.5	0.927	0.939	0.921	0.903	0.681	0.730	0.951	0.946	0.892	0.878
2.0	1.000	1.000	1.000	1.000	0.835	0.854	1.000	1.000	0.999	0.999
3.0	1.000	1.000	1.000	1.000	0.880	0.890	1.000	1.000	1.000	1.000
5.0	1.000	1.000	1.000	1.000	0.922	0.929	1.000	1.000	1.000	1.000
					$n = 1000$					
1.0	0.059	0.091	0.059	0.034	0.048	0.063	0.050	0.058	0.050	0.043
1.05	0.357	0.453	0.356	0.267	0.232	0.277	0.363	0.391	0.249	0.223
1.1	0.781	0.842	0.782	0.703	0.540	0.585	0.806	0.825	0.665	0.634
1.2	0.993	0.999	0.993	0.990	0.850	0.866	0.999	0.999	0.995	0.994
1.5	1.000	1.000	1.000	1.000	0.919	0.922	1.000	1.000	1.000	1.000
2.0	1.000	1.000	1.000	1.000	0.924	0.926	1.000	1.000	1.000	1.000
3.0	1.000	1.000	1.000	1.000	0.934	0.935	1.000	1.000	1.000	1.000
5.0	1.000	1.000	1.000	1.000	0.961	0.962	1.000	1.000	1.000	1.000

6.2. Power Comparisons for the Two Sample Case

For the two sample case power comparisons were performed using two truncated negativ binomially distributed samples as well as two Zipfian distributed samples. The truncated negative binomial distributions are applied with stochastically determining the parameter p_1 by drawing a random number from the uniform distribution, while parameter $p_2 = (p_1 + \delta) \leq 1$, with δ varied between 0 (identical skewness) to 0.4 (sample two is more skewed to the right). In addition, both distributions were designed to be slightly skewed to the left by setting $r = 5$ (see table 5). For the Zipfian distributions, the parameter p_1 was as well randomly drawn from a uniform distribution, while parameter $p_2 = (p_1 + \delta) \leq 1$, with δ varied between 0 (identical skewness) to 0.4 (sample two is more skewed to the right) (see table 6).

Results of the two sample power comparisons demonstrate that the tests based on the new linear skewness functionals provide good power in comparison to the alternative tests even at composite hypotheses. At both two sample cases, the S^K test together with the Wilcoxon-Mann-Whitney test, outperform the alternative tests. For the two truncated negative binomially distributed samples, the S^K shows slightly higher power than the Wilcoxon-Mann-Whitney test, especially at low sample sizes. For the two Zipfian distributions the situation is reversed. Moreover, as the test based on the third moment ($\sqrt{b_1}$ test) is sensitive to the assigned values of the categories, to ensure a positive value of the test's estimated variance, the categories' values were set to equal $1.5, ..., 7.5$ in this special case. Nonetheless, this test shows through all power comparison cases the worst performance.

Table 5: Power comparisons for the two sample case on two truncated negativ binomial distributions with k=7; r=5

δ	S^K exact	S^K asy.	S^G exact	S^G asy.	$\sqrt{b_1}$ exact	$\sqrt{b_1}$ asy.	Wilcoxon exact	Wilcoxon asy.	GLR exact	GLR asy.
					$n = 20$					
0.0	0.054	0.054	0.054	0.079	0.054	0.089	0.052	0.050	0.051	0.039
0.02	0.079	0.077	0.078	0.106	0.058	0.107	0.071	0.069	0.056	0.042
0.05	0.119	0.120	0.120	0.171	0.074	0.108	0.105	0.104	0.078	0.054
0.1	0.241	0.243	0.244	0.314	0.101	0.180	0.217	0.233	0.133	0.108
0.2	0.583	0.584	0.553	0.616	0.205	0.325	0.560	0.545	0.375	0.326
0.3	0.829	0.830	0.787	0.833	0.399	0.521	0.814	0.814	0.645	0.602
0.4	0.960	0.961	0.938	0.957	0.618	0.716	0.954	0.952	0.856	0.836
					$n = 100$					
0.0	0.052	0.047	0.053	0.077	0.055	0.014	0.049	0.048	0.052	0.029
0.02	0.101	0.097	0.102	0.146	0.073	0.038	0.097	0.096	0.070	0.042
0.05	0.253	0.244	0.243	0.306	0.122	0.047	0.236	0.249	0.146	0.095
0.1	0.569	0.557	0.504	0.566	0.288	0.243	0.562	0.548	0.370	0.298
0.2	0.912	0.914	0.819	0.882	0.684	0.688	0.914	0.913	0.762	0.710
0.3	0.996	0.995	0.981	0.992	0.911	0.861	0.996	0.996	0.959	0.935
0.4	0.999	0.999	1.000	1.000	0.988	0.978	1.000	1.000	0.999	0.999
					$n = 1000$					
0.0	0.046	0.049	0.048	0.078	0.046	0.020	0.047	0.048	0.046	0.025
0.02	0.304	0.306	0.280	0.348	0.132	0.097	0.292	0.289	0.182	0.123
0.05	0.745	0.740	0.604	0.676	0.484	0.462	0.758	0.749	0.535	0.470
0.1	0.980	0.981	0.906	0.948	0.868	0.901	0.984	0.983	0.880	0.836
0.2	1.000	1.000	1.000	1.000	0.998	0.999	1.000	1.000	1.000	1.000
0.3	0.999	0.999	1.000	1.000	0.999	0.999	1.000	1.000	1.000	1.000
0.4	0.999	0.999	1.000	1.000	0.999	0.999	1.000	1.000	1.000	1.000

Table 6: Power comparisons for the two sample case on two Zipfian distributions with k=7

δ	S^K exact	S^K asy.	S^G exact	S^G asy.	$\sqrt{b_1}$ exact	$\sqrt{b_1}$ asy.	Wilcoxon exact	Wilcoxon asy.	GLR exact	GLR asy.
					$n = 20$					
0.0	0.054	0.046	0.054	0.033	0.049	0.026	0.054	0.049	0.049	0.045
0.02	0.060	0.051	0.061	0.046	0.054	0.030	0.060	0.055	0.052	0.045
0.05	0.068	0.058	0.067	0.047	0.056	0.031	0.067	0.062	0.056	0.048
0.1	0.082	0.070	0.078	0.054	0.063	0.036	0.078	0.075	0.062	0.054
0.2	0.114	0.099	0.099	0.064	0.091	0.050	0.111	0.110	0.078	0.074
0.3	0.159	0.139	0.138	0.106	0.120	0.069	0.165	0.154	0.102	0.089
0.4	0.202	0.188	0.161	0.164	0.144	0.084	0.196	0.206	0.128	0.120
					$m = 100, n = 100$					
0.0	0.047	0.045	0.047	0.059	0.050	0.026	0.047	0.048	0.051	0.039
0.02	0.058	0.055	0.057	0.054	0.058	0.030	0.057	0.058	0.056	0.043
0.05	0.075	0.073	0.072	0.060	0.067	0.036	0.074	0.075	0.065	0.050
0.1	0.116	0.112	0.107	0.106	0.093	0.052	0.118	0.119	0.085	0.067
0.2	0.232	0.229	0.204	0.186	0.163	0.098	0.240	0.244	0.146	0.123
0.3	0.380	0.387	0.317	0.342	0.277	0.166	0.388	0.410	0.253	0.213
0.4	0.580	0.571	0.495	0.493	0.390	0.263	0.599	0.603	0.382	0.336
					$m = 1000, n = 1000$					
0.0	0.049	0.049	0.048	0.041	0.053	0.023	0.050	0.049	0.051	0.036
0.02	0.084	0.088	0.084	0.090	0.082	0.037	0.088	0.088	0.069	0.052
0.05	0.165	0.179	0.160	0.190	0.128	0.067	0.168	0.183	0.114	0.096
0.1	0.430	0.424	0.412	0.368	0.272	0.169	0.445	0.439	0.267	0.231
0.2	0.899	0.897	0.879	0.885	0.678	0.551	0.908	0.909	0.770	0.722
0.3	0.996	0.996	0.993	0.994	0.933	0.876	0.997	0.997	0.980	0.977
0.4	0.999	0.999	0.999	0.999	0.994	0.988	1.000	1.000	1.000	1.000

7. Applications

In this section the new class of skewness measures and tests were applied to two different datasets. The first dataset which is on reactions towards an anti-smoking advertisement was used by Bhattacharya and Nandram (1996) as one of examples that should illustrate possible gains in precision can be achieved by applying a stochastic order restriction when using bayesian methods at multinomial populations. The second dataset considers political left-right self reporting data gathered by the the German research institute for social sciences 'GESIS' in 2016.

Application 1: Anti-Smoking Advertisement

The dataset for the first application was used by Bhattacharya and Nandram (1996) and provided by Gelb and Pickett (1983). Reactions to an anti-smoking advertisement were gathered via questionnaire with one of its items asking the subjects to indicate on the question 'I dislike the advertisement' on a 5-point rating scale ranging from 'strongly agree' to 'strongly disagree'. This question was answered by 97 smokers and 281 non-smokers. Denote f^S and f^N the vectors of relative frequencies for the five ordered categories of smokers respectively non-smokers. F^S and F^N are the corresponding vectors of cumulated frequencies. The vectors Q^S and Q^N are given by $Q_i^S = (F_i^S + F_{k-i}^S)/2$ and $Q_i^N = (F_i^N + F_{k-i}^N)/2$, $i = 1, 2, \ldots, k-1 = 4$.

Table 7: Relative frequencies of smokers and non-smokers (Gelb and Pickett (1983))

	strongly agree	agree	neutral	disagree	strongly disagree
			Smokers		
f_i^S	0.082	0.144	0.361	0.216	0.196
F_i^S	0.082	0.227	0.588	0.804	1.000
Q_i^S	0.443	0.407	0.407	0.443	
			Non-Smokers		
f_i^N	0.110	0.149	0.278	0.217	0.246
F_i^N	0.110	0.260	0.537	0.754	1.000
Q_i^N	0.432	0.399	0.399	0.432	

Table 7 shows that the median is on the middle category 'neutral' for the frequency distributions of smokers as well as non-smokers. However, as can be seen by $Q_{i'i} < 1/2$, $i' = 1, 2$, the frequency distributions of smokers as well as non-smokers are skewed to the left. This assumption was tested and thus table 8 provides results of all considered tests for the one sample case on the alternative that the smokers' distribution, respectively the non-smokers' distribution is skewed to the left.

Table 8: One sample tests on skewness to the left for smokers and non-smokers

	sample value	critical value (exact)	p value (exact)	critical value (asy)	p value (asy)
		Smokers			
S^K	-0.1495	-0.0825	0.0012	-0.0835	0.0016
S^G	-0.2753	-0.1577	0.001	-0.167	0.0033
$\sqrt{b_1}$	-0.1776	-0.2713	0.1398	-0.2696	0.1393
Wilcoxon	0.3353	0.3075	0.0077	0.3146	0.0136
GLR	6.024	4.8804	0.0308	5.1324	0.0317
		Non-Smokers			
S^K	-0.169	-0.0498	0	-0.0491	0
S^G	-0.3084	-0.0956	0	-0.0981	0
$\sqrt{b_1}$	-0.2714	-0.1547	0.0025	-0.1627	0.003
Wilcoxon	0.3324	0.2837	0	0.2843	0
GLR	18.3343	4.6746	0	5.0373	0

Almost all tests on skewness do foster the descriptive statistics' indication that both samples of smokers and non-smokers are skewed to the left. This can be interpreted as both groups rather liked than disliked the anti-smoking advertisment.

As can be seen in table 7, $Q_i^S > Q_i^N$ for $i = 1, 2, \ldots, k-1 = 4$, implies that the skewness measures from class S^b preserve the skewness ordering and the smokers' frequency distribution is more skewed to the right than the non-smokers' one. This is in line with Bhattacharya and Nandram (1996) that expected non-smokers to have a more positive reaction towards the anti-smoking advertisment than smokers. However, Bhattacharya and Nandram

(1996) found that the stochastic ordering restriction is not appropriate for this dataset and thus no gains in precision for bayesian inference could be achieved by applying this ordering restriction. Nonetheless, table 8's descriptive results foster the expectation of f^N being more skewed to the left than f^S. Moreover, due to the skewness properties provided in section 3 the non-smokers' distribution is more skewed to the left than the smokers' distribution as $S^K_{Non-Smokers} < S^K_{Smokers}$ and $S^G_{Non-Smokers} < S^G_{Smokers}$.

Table 9 shows the two sample tests' results of all considered tests on skewness for the alternative that the non-smokers' distribution is more skewed to the right than the smokers' ones. In line with Bhattacharya and Nandram (1996), no significant differences in skewness

Table 9: Two sample tests on non-smokers more skewed to the left than smokers

	sample value	critical value (exact)	p value (exact)	critical value (asy)	p value (asy)
S^K	0.0196	0.1392	0.4118	0.1638	0.4222
S^G	0.033	0.2316	0.4095	0.2775	0.4224
$\sqrt{b_1}$	0.0938	0.2983	0.298	0.2468	0.2659
Wilcoxon	0.5133	0.435	0.6304	0.4455	0.6556
GLR	1.5533	9.2241	0.5434	5.1378	0.3363

could be found for the smokers' compared to the non-smokers' distribution. Despite not being significant, results from table 8 provides the indication that non-smokers did not dislike the anti-smoking advertisment as much as smokers.

Application 2: Political Left-Right Self Reporting

The German research institute for social sciences 'GESIS' periodically asks the German population on general and recent subjects using a questionnaire called ALLBUS (see GESIS (2017)). One of its items concerns the political 'left-right self reporting'. This self reporting is measured by a 10-point rating scale, while '1' means an extremely left and '10' an extremely right political position. In the 2016 ALLBUS, this question was answered by 2221 persons from West Germany and 1114 persons from East Germany (former GDR). Denote f^W and f^E the vectors of relative frequencies for the ten ordered categories in West and East Germany.

F^W and F^E are the corresponding vectors of cumulated frequencies. The vectors Q^W and Q^E are given by $Q_i^W = (F_i^W + F_{k-i}^W)/2$ and $Q_i^E = (F_i^E + F_{k-i}^E)/2$, $i = 1, 2, \ldots, k-1 = 9$.

Table 10: Relative frequencies for the political left-right self reporting for West and East Germany

i	1	2	3	4	5	6	7	8	9	10
	West Germany									
f_i^W fi	0.022	0.036	0.113	0.124	0.247	0.276	0.116	0.050	0.009	0.007
F_i^W	0.022	0.058	0.171	0.294	0.542	0.818	0.934	0.984	0.993	1.000
Q_i^W	0.507	0.521	0.552	0.556	0.542	0.556	0.552	0.521	0.507	
	East Germany									
f_i^E	0.034	0.066	0.153	0.110	0.239	0.260	0.083	0.037	0.008	0.010
F_i^E	0.034	0.101	0.253	0.364	0.602	0.863	0.945	0.982	0.990	1.000
Q_i^E	0.512	0.541	0.599	0.613	0.602	0.613	0.599	0.541	0.512	

Table 10 shows that the median is the category '5' for both frequency distributions of West and East Germany. However, as can be seen by $Q_{i'i} > 1/2$, $i' = 1, 2$, the frequency distributions of West and East Germany are skewed to the right. This can be interpreted as both groups stated to be more politically left than right.

This assumption was tested and thus table 11 provides results of all considered tests for the one sample case on the alternative that f^W, respectively f^E is skewed to the right.

Table 11: One sample tests on skewness to the right for the political left-right self reporting of West and East Germany

	sample value	critical value (exact)	p value (exact)	critical value (asy)	p value (asy)
		West Germany			
S^K	0.07	0.0116	0	0.0128	0
S^G	0.1335	0.0229	0	0.0233	0
$\sqrt{b_1}$	-0.1963	0.0709	0.9999	0.0728	1
Wilcoxon	0.2047	0.24	0	0.2402	0
GLR	-115.4078	-8.5946	0	0	0
		East Germany			
S^K	0.1408	0.0166	0	0.0194	0
S^G	0.2552	0.0324	0	0.0317	0
$\sqrt{b_1}$	-0.0677	0.1013	0.8674	0.1012	0.8645
Wilcoxon	0.1602	0.236	0	0.2362	0
GLR	-164.0569	-9.0464	0	0	0

Almost all tests on skewness do foster the descriptive statistics' indication that both samples of West and East Germany are skewed to the right. The only exception is the $\sqrt{b_1}$ test, whose results are not in line with the alternative tests' results throughout all tests made to this point. This behavior can be explained by its poor power performance shown in tables 3 to 6.

Moreover, table 10 shows that West Germany's distribution is more skewed to the left than East Germany's one, as $Q_i^W < Q_i^E$, $i = 1, 2, ..., k-1 = 9$. In addition, this result implies that skewness measures from the class S^b preserve the skewness ordering. Table 11 fosters this indication as $S^K_{West} < S^K_{East}$ and $S^G_{West} < S^G_{East}$. Table 12 shows the two sample tests' results of all considered tests on skewness for the alternative that West Germany's distribution is more skewed to the left than East Germany's distribution.

The tests show that West Germany's distribution is significantly more skewed to the left than East Germany's distribution. This result can be interpreted by subjects from East

Table 12: Two sample tests on the political left-right self reporting of West Germany is more skewed to the left than East Germany's one

	sample value	critical value (exact)	p value (exact)	critical value (asy)	p value (asy)
S^K	-0.0708	-0.0177	0	-0.0675	0.0421
S^G	-0.1217	-0.0325	0	-0.1184	0.0454
$\sqrt{b_1}$	-0.1286	-0.1194	0.0366	-0.1403	0.0658
Wilcoxon	0.4484	0.4857	0	0.4829	0
GLR	-30.0674	0.3319	0	0.4732	0

Germany reported to be politically more right than subjects from West Germany.

8. Conclusion

In summary, this paper first discussed skewness properties suitable for ordered categorical data and introduced a new skewness ordering which is restricted to a fixed number of categories. Based on these proposals, skewness functionals and skewness statistics were presented that are defined as weighted sums of cumulated frequencies. Therefore, a test on asymmetry for ordered categorical variables is provided. Power comparisons to alternative tests of symmetry on discrete data with restricted support demonstrated the good power performances of these new tests on asymmetry. Finally, two examples were provided to show possible applications and interpretations of the new class of linear skewness measures and tests.

Bibliography

Arellano-Valle, R. B., Gómez H. W., and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128:427–443.

Arnold, B. C. and Groeneveld, R. (1995). Measuring skewness with respect to the mode. *The American Statistician*, 49(1):34–38.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Bhapkar, V. P. (1961). Some tests for categorical data. *Annals of Mathematical Statistics*, 32(1):72–83.

Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, 61:228–235.

Bhattacharya, B. (1997). On tests of symmetry against one-sided alternatives. *Annals of the Institute of Statistical Mathematics*, 49(2):237–254.

Bhattacharya, B. and Nandram, B. (1996). Bayesian inference for multinomial populations under stochastic ordering. *Journal of Statistical Computation and Simulation*, 54:145–163.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43(244):572–574.

Bowley, A. L. (1920). *Elements of Statistics*. Scribner, New York, 4 edition.

Büning, H. and Trenkler, G. (1994). *Nichtparametrische statistische Methoden*. de Gruyter, Berlin.

Cassart, D., Hallin, M., and Paindaveine, D. (2008). Optimal detection of Fechner-asymmetry. *Journal of Statistical Planning and Inference*, 138:2499–2525.

Chanda, K. C. (1963). On the efficiency of two-sample mann-whitney test for discrete populations. *Annals of Mathematical Statistics*, 34(2):612–617.

Charlier, C. V. L. (1905). Über das Fehlergesetz. *Archiv für Mathematik, Astronomi och Fysik*, 2:8.

Conover, W. J. (1973). Rank tests for one sample, two sample, and k samples without the assumption of a continuous distribution function. *Annals of Statistics*, 1(6):1105–1125.

De Leeuw, J., Hornik, K., and Mair, P. (2010). Isotone optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32(5):1–24.

Doane, D. P. and Seward, L. E. (2011). Measuring skewness: a forgotten statistic? *Journal of Statistics Education*, 19(2):1–18.

Dykstra, R., Kochar, S., and Robertson, T. (1995). Likelihood ratio tests for symmetry against one-sided alternatives. *Annals of the Institute of Statistical Mathematics*, 74(4):719–730.

Dykstra, R., Kochar, S., and Robertson, T. (1995b). Inference for likelihood ratio ordering in the two-sample problem. *Journal of the American Statistical Association*, 90(431):1034–1040.

Edgeworth, F. Y. (1904). The law of error. *Transactions of the Cambridge Philosophical Society*, 20:26–65.

Fechner, G. T. (1897). *Kollektivmaßlehre*. Engelmann, Leipzig.

Fernández, C. and Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Ferreira, T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association*, 101:823–829.

Freidlin, B. and Gastwirth, J. L. (2000). Should the median test be retired from general use? *The American Statistician*, 54(3):161–164.

Gelb, B. D. and Pickett, C. M. (1983). Attitude-toward-the-ad: links to humor and to advertising effectiveness. *Journal of Advertising*, 12:34–42.

GESIS. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2016.

Gini, C. (1955). Variabilità e mutabilità. In Pizetti, E. and Salvemini, T., editors, *Memorie die metodologica statistica*. Libreria Eredi Virgilio Veschi, Rome.

Groeneveld, R. (1991). An influence function approach to describing the skewness of a distribution. *The American Statistician*, 45(2):97–102.

Grottke, M. (2002). *Die t-Verteilung und ihre Verallgemeinerungen als Model für Finanzmarktdaten.* Josef Eul Verlag, Lothmar.

Hájek, J. and Sidák, V. (1967). *Theory of rank tests.* Academic Press, New York.

Hotelling, H. and Solomons, L. M. (1932). The limits of a measure of skewness. *The Annals of Mathematical Statistics*, 3:141–142.

Kiesl, H. (2003). *Ordinale Streuungsmaße.* Eul-Verlag, Köln.

Klein, I. (1994). *Mögliche Skalentypen, invariante Relationen und wissenschaftliche Gesetze.* Vandenhoeck & Ruprecht, Göttingen.

Klein, I. (1999). Systematik der Schiefemessung für ordinalskalierte Merkmale. *Diskussion Paper der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Friedrich-Alexander Universität Erlangen - Nürnberg*, 26.

Klein, I. (2001). Schiefemessung ordinalskalierter Merkmale mittels Rangordnungsstatistiken. *Allgemeines Statistisches Archiv*, 85:67–78.

Klein, I. and Fischer, M. (2006). Skewness by splitting the scale parameter. *Communications in Statistics - Theory and Methods*, 35(7):1159–1171.

Lehmann, E. L. (1953). The power of rank tests. *The Annals of Mathematical Statistics*, 24(1):23–43.

Ley, C. and Paindaveine, D. (2009). Le cam optimal tests for symmetry against Ferreira and Steel's general skewed distributions. *Journal of nonparametric statistics*, 21(8):943–967.

MacGillivray, H. L. (1986). Skewness and asymmetry: measures and orderings. *Annals of Statistics*, 14:994–1011.

Majindar, K. L. (1962). Improved bounds on a measure of skewness. *The Annals of Mathematical Statistics*, 33(3):1192–1194.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, 18:154–168.

Pareto, V. (1897). *Cours d'économie politique*. F. Rouge, Lausanne.

Pearson, K. (1895). Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, 186:343–414.

Pratt, J. W. (1959). Remarks on zeros and ties in the wilcoxon signed rank procedures. *Journal of the American Statistical Association*, 54:169–187.

Premaratne, G. and Bera, A. (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics*, 3(2):169–187.

Ramsey, F. L. (1971). Small sample power functions for nonparametric tests of location in the double exponential family. *Journal of the American Statistical Association*, 66(333):149–151.

Rayner, J. C. W., Best, D. J., and Mathews, K. L. (1995). Interpreting the skewness coefficient. *Communications in Statistics - Theory and Methods*, 24(3):593–600.

Robertson, T. and Wright, F. T. (1981). Likelihood ratio tests for and against stochastic ordering between multinomial populations. *Annals of Statistics*, 9(6).

Rohatgi, V. K. and Székely, G. J. (1989). Sharp inequalities between skewness and kurtosis. *Statistics & Probability Letters*, 8(4):297–299.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons, New York.

Tabor, J. (2010). Investigating the investigative task: Testing for skewness: An investigation of different test statistics and their power to detect skewness. *Journal of Statistics Education*, 18(2):1–13.

Thas, O., Rayner, J. C. W., and Best, D. J. (2005). Tests for symmetry based on the one-sample wilcoxon signed rank statistic. *Communications in Statistics - Simulation and Computation*, 34(4):957–973.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, 44(12):1650–1661.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In Olkin, I., editor, *Contributions to Probability and Statistics*. Stanford University Press.

van Zwet, W. R. (1964). *Convex transformations of random variables*. Mathematical Centre Tract 7, Amsterdam.

von Hippel, P. T. (2005). Mean, median, and skew: correcting a textbook rule. *Journal of Statistics Education*, 13(2):1–13.

Vorlicková, D. (1972). Asymptotic properties of rank tests of symmetry under discrete distributions. *Annals of Mathematical Statistics*, 43:2013–2018.

Yanagimoto, T. and Sibuya, M. (1972). Test of symmetry of a one-dimensional distribution against positive biasedness. *Annals of the Institute of Statistical Mathematics*, 24:423–434.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, 75(6):579–652.

Zhao, Y. D., Rahardja, D., and Qu, Y. (2008). Sample size calculation for the wilcoxon-mann-whitney test adjusting for ties. *Statistics in Medicine*, 27:462–468.

Appendix

Proof of theorem 4.1:

Due to the limit theorem of deMoivre & Laplace $(F_1, F_2, \ldots, F_{k-1})$ is asymtotically normally distributed with mean $(P_1, P_2, \ldots, P_{k-1})$ and covariance matrix $\frac{1}{n}\Sigma$ with elements

$$\sigma_{ij} = Cov(F_i, F_j) = \begin{cases} P_i(1-P_i)/n & \text{für } i = j \\ P_i(1-P_j)/n & \text{für } i < j \\ P_j(1-P_i)/n & \text{für } i > j \end{cases}$$

for $i, j = 1, 2, \ldots, k-1$ (see e.g. Kiesl (2003) p. 98).

$S^b(N) = \frac{1}{k-1}\sum_{i=1}^{k-1} b(H_i)$ is a differentiable function of $(F_1, F_2, \ldots, F_{k-1})$. If the differential of $S^b(F)$ does not vanish at $\mu = (P_1, \ldots, P_{k-1})$, then $S^b(F)$ is asymptotically normally distributed with mean $\sum_{i=1}^{k-1} b(P_i)$ and variance

$$\frac{1}{n}\left(\frac{\partial S^b(F)}{\partial F_1}, \ldots, \frac{\partial S^b(F)}{\partial F_{k-1}}\right)\Sigma\left(\frac{\partial S^b(F)}{\partial F_1}, \ldots, \frac{\partial S^b(F)}{\partial F_{k-1}}\right)'$$

(see Serfling (1980), p. 124).

$$\frac{\partial S^b(F)}{\partial F_i} = \frac{1}{k-1} b'(H_i)\frac{\partial H_i}{\partial F_i},$$

while $H_i = (F_i + F_{k-i})/2$ for $i = 1, 2, \ldots, k-1$, leads to the variance of the asymptotic distribution. □

Proof of theorem 5.1:

Let $(n_1, \ldots, n_k)$ be multinomially distributed with parameters $n = \sum_{i=1}^{k} n_i$ and $p = (p_1, \ldots, p_k)$, $\sum_{i=1}^{k} p_i = 1$. Let $\nu = E(f) = p$ and Σ denotes the covariance matrix of $f = (f_1 = n_1/n, \ldots, n_k/n)$ with $\sum_{i=1}^{k} f_i = 1$, then the elements of Σ are

$$\sigma_{ij} = Cov(f_i, f_j) = \begin{cases} p_i(1-p_i)/n & \text{for } i = j \\ -p_i p_j/n & \text{for } i \neq j \end{cases}$$

for $i, j = 1, 2, \ldots, k$. If $u_1, \ldots, u_k$ are fixed values such that p_i is the probability that u_i occurs and f_i is the relative frequency of u_i in the sample of size n, then the standardized third moment

$$SK(u, f) = \frac{M_3(u, f)}{(S^2(u, f))^{3/2}}$$

with

$$M_3(u, f) = \sum_{i=1}^{k}(u_i - M_1(u, f))^3 f_i, \quad S^2(u, f) = \sum_{i=1}^{k}(u_i - M_1(u, f))^2 f_i, \quad M_1(u, f) = \sum_{i=1}^{k} u_i f_i$$

is a function of f. The gradient w.r.t f exists. The elements of this gradient are

$$\frac{\partial SK(u, f)}{\partial f_i} = -3\frac{u_i}{\sqrt{S^2(u, f)}} + \frac{(u_i - M_1(u, f))^3}{S^2(u, f)^{3/2}} - \frac{3}{2} SK(u, f)\frac{u_i^2 - 2M_1(u, f)u_i}{S^2(u, f)}$$

due to

$$\frac{\partial M_1(u,f)}{\partial f_i} = u_i,$$

$$\frac{\partial S^2(u,f)}{\partial f_i} = u_i^2 - 2M_1(u,f)u_i$$

and

$$\frac{\partial M_3(u,f)}{\partial f_i} = -3S^2(u,f)u_i + (u_i - M_1(u,f))^3$$

for $i = 1, 2, \ldots, k$. Denote $\nabla(f)$ the gradient of SK w.r.t to f. Due to Serfling (1980), p. 124

$$SK(u,f) \overset{asy}{\sim} N\left(SK(u,p), \frac{1}{n}\Sigma\right)$$

if $\nabla(f) \neq 0$ for $f = p$. $\square$