# Relative Entropy Measures of Asymmetry with Applications

**Bhaskar Bhattacharya**

Department of Mathematics

Southern Illinois University

Carbondale, IL 62901-4408, USA

**Abstract**

We propose a measure of asymmetry of a probability density function (pdf) by considering the relative entropy between itself and its (appropriately defined) mirror image. The measure is shown to be useful for detecting asymmetry in distributions of categorical or continuous random variables. Asymmetries of a pdf near its center and away from the center are investigated. This measure leads to generalizations of asymmetric categorical models. Comparison (using examples) with the asymmetry measures of MacGillivray (1986) shows the proposed measures are useful for non-monotonic asymmetry. For square contingency tables with same row and column classifications, the sampling distributions of the measures are studied asymptotically. Applications are discussed for two-way tables and in linear regression models. Monte Carlo simulations show that the proposed measures/tests have good size and power properties when compared with competitors, even for smaller samples. Two illustrative examples are analyzed.

*Key words and phrases:* categorical, central asymmetry, coefficients, isotonic regression, mirror image, skewness, tail asymmetry, unimodal ordering.

## 1 Introduction

The concept of symmetry has played an important role in statistics. In nonparametric statistics, validity of many procedures depend on this assumption. In parametric inference, there are often situations when the assumption of normality can be replaced by the assumption of any symmetric distribution (Chaffin and Rhiel, 1993). For a continuous random variable $X$ with distribution function (df) $F(x)$ and probability density function (pdf) $f(x)$, let the mean, median and mode be denoted by $\mu$, $m$ and $M$, respectively. The historically known measures for detecting departures from symmetry include $(\mu-M)/\sigma$, $E(X-\mu)^3/\sigma^3$, $(Q_3+Q_1-2m)/(Q_3-Q_1)$ and $(\mu-m)/\sigma$, where $\sigma$ is the standard deviation of $X$, and $Q_1, Q_3$ are the first and third quartiles, respectively. See MacGillivray (1986) for the related references who also discussed $\gamma_p(F) = \frac{F^{-1}(1-p)+F^{-1}(p)-2m}{F^{-1}(1-p)-F^{-1}(p)}, p \in (0, .5)$ and $(\mu - m)/E|X - m|$. An influence function approach to describing the skewness of a distribution using some of these measures is given by Groeneveld (1991). Note that there exists asymmetric distributions with the third central moment equal to zero (Ord, 1968; Johnson and Kotz, 1970).

van Zwet (1964) defined a df $G$ having 'greater skewness to the right' than a df $F$ if $G^{-1}(F(x))$ is convex on $A_F = \{x : 0 < F(x) < 1\}$. Oja (1981) studied the function $G^{-1}(F(x))-x$ on $A_F$ using various partial orderings of distributions. Doksum (1975) proposed the symmetry function $\theta_F(x) = [x - \overline{F}^{-1}(F(x))]/2$ ($\overline{F}(x) = 1 - F(x)$), and defined $F$ as 'strongly skewed to the right' if and only if $\theta_F(x)$ is nonincreasing for $x < m_F$ and nondecreasing for $x \geq m_F$ and 'skewed to the right' if and only if $\theta_F(x) \geq m_F$. MacGillivray (1986) studied the interconnections between many of the above partial orderings and discussed measures that are derived from them. Two such measures are

$$\psi_1 = \frac{|\mu - m|}{\sigma} \text{ and } \psi_2 = \left| \sup_{\alpha \leq u \leq .5} \gamma_u(F) - \inf_{\alpha \leq u \leq .5} \gamma_u(F) \right| \tag{1.1}$$

for $\alpha > 0$. Both of these are in $(0, 1)$, which is a desirable property of an asymmetry measure.

In this paper, we propose a measure of departure from symmetry that compares the pdf $f(x)$ with its 'mirror image' (or reflection). The definition of mirror image depends on the context. In the continuous univariate case, the mirror image of a skewed to the right pdf becomes skewed to the left (possibly on a different domain) depending on the point of reflection. When needed, we make a linear transformation to the domain of the mirror image pdf so that this domain matches with that of the original pdf. Then the relative entropy (or the Kullback-Leibler deviation) between these two pdf's is considered. If the original pdf is symmetric to begin with, then it is identical with its mirror image, and the above deviation is zero; otherwise the deviation is nonnegative and provides us a measure of its asymmetry.

MacGillivray (1986, p.997) proposed some desirable criteria of skewness measures: they should (a) have same structure for self and different distributions comparisons, (b) compare with existing measures, (c) be able to identify the ordering the measure indicates, (d) measure skewness of any asymmetric distribution, (e) not be more complicated than necessary. Using $f(x)$ and its mirror image only (and no other information) to measure asymmetry, we satisfy (e). The criterion (b) is considered in Section 6. The criteria (a,c,d) are shown to hold in Sections 2, 3. As we will see later, the comparison between $f(x)$ and its mirror image, leads to a symmetric measure, so it does not matter which one we start with. The proposed quantity measures deviations from symmetry only and it does not detect the direction of asymmetry, right or left. Such behavior of asymmetry measures are acceptable (MacGillivray, 1986, page 1005).

In Section 2, we consider asymmetry in two-way contingency tables. Discrete and continuous univariate random variables are considered in Section 3. Section 4 addresses applications of the proposed measure in the areas of contingency tables and linear regression, respectively. Comparison with existing measures and related discussions are in Section 5. Proofs of results are deferred to the appendix at the end.

## 2 Two-way contingency tables

Let $\boldsymbol{p} = ((p_{ij}), i = 1, \ldots, r,\ j = 1, \ldots, r,\ p_{ij} \geq 0, \sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} = 1)$ be an $r \times r$ contingency table of probabilities. For another $r \times r$ contingency table of probabilities $\boldsymbol{q} = ((q_{ij}), i = 1, \ldots, r,\ j = 1, \ldots, r,)$, the *relative entropy* between $\boldsymbol{p}$ and $\boldsymbol{q}$ is

$$I(\boldsymbol{p}|\boldsymbol{q}) = \sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} \ln \frac{p_{ij}}{q_{ij}}. \tag{2.1}$$

Here we adopt the conventions that $0 \ln(0/0) = 0,\ 0 \ln 0 = 0,\ a \ln(a/0) = \infty, \forall a > 0$. It is easily shown that $I(\boldsymbol{p}|\boldsymbol{q}) \geq 0$, and $I(\boldsymbol{p}|\boldsymbol{q}) = 0$ if and only if $p_{ij} = q_{ij}, \forall i, j$.

The symmetry model is defined as

$$p_{ij} = p_{ji}, \forall i, j. \tag{2.2}$$

Due to its strict restrictions, the symmetry model may not be practical in most circumstances. To study the deviations from symmetry, one would like to know how different is $p_{ij}$ from $p_{ji}$, $\forall i > j$. We assume $\sum_{i=1}^{r}\sum_{j\neq i} p_{ij} \neq 0$. To determine if $\boldsymbol{p}$ is symmetric around the main diagonal, we propose a measure of deviation from symmetry as follows

$$\delta = \frac{1}{\sum_{i=1}^{r}\sum_{j\neq i} p_{ij}} \sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} \ln \frac{p_{ij}}{p_{ji}},$$

which can be simplified as

$$\delta = \frac{\sum\sum_{i<j} p_{ij} \ln \frac{p_{ij}}{p_{ji}} + \sum\sum_{i<j} p_{ji} \ln \frac{p_{ji}}{p_{ij}}}{\sum\sum_{i<j} p_{ij} + \sum\sum_{i<j} p_{ji}} = \frac{\sum\sum_{i<j} (p_{ij} - p_{ji}) \ln \frac{p_{ij}}{p_{ji}}}{\sum\sum_{i<j} (p_{ij} + p_{ji})}. \tag{2.3}$$

The measure $\delta$ ranges in the interval $[0, \infty)$. While 0 refers to symmetry, $\infty$ corresponds to 'complete asymmetry', which is attained when either $p_{ij} = 0$ or $p_{ji} = 0$, but not both, for some $(i, j)$.

Next we present some preliminary properties of the measure $\delta$. The following remarks and results help one understand its nature.

**Remark 2.1** To see that $\delta$ is truly a divergence between two probability distributions, define $T = \sum_{i=1}^{r} \sum_{j \neq i} p_{ij}$, then it is easy to see that $\delta$ denotes divergence between: $\{p_{ij}/T : i \neq j\}$ and $\{p_{ji}/T : i \neq j\}$. These two probability distributions use the same $p_{ij}$'s in different order. $\square$

In general, for two probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$, it is well-known that $I(\boldsymbol{p}|\boldsymbol{q}) \neq I(\boldsymbol{q}|\boldsymbol{p})$. However, it is easy to see that the following symmetry holds.

**Proposition 2.1** Let $\boldsymbol{p}^{\dagger}$ be the transpose of $\boldsymbol{p}$. Then $I(\boldsymbol{p}|\boldsymbol{p}^{\dagger}) = I(\boldsymbol{p}^{\dagger}|\boldsymbol{p}), \forall \boldsymbol{p}$. $\square$

Thus $\delta = I(\boldsymbol{p}|\boldsymbol{p}^{\dagger})/T$ is invariant to the order of $\boldsymbol{p}$ and $\boldsymbol{p}^{\dagger}$. Although $\delta$ compares $\{p_{ij}/T : i \neq j\}$ with its transpose, the following lemma shows that $\delta$ can also be expressed as a divergence between two separate probability distributions.

**Proposition 2.2** Let $\boldsymbol{p}^{(1)}$ be the probability distribution with entries as $p_{ij}^{(1)} = p_{ij}/\sum\sum_{i>j} p_{ij}$ for $i > j$, and $\boldsymbol{p}^{(2)}$ be the probability distribution with entries as $p_{ij}^{(2)} = p_{ij}/\sum\sum_{i<j} p_{ij}$ for $i < j$. If $\alpha = \sum_{i>j} p_{ij}/\sum\sum_{i \neq j} p_{ij}$, then

$$\delta = \alpha I(\boldsymbol{p}^{(1)}|\boldsymbol{p}^{(2)}) + (1-\alpha) I(\boldsymbol{p}^{(2)}|\boldsymbol{p}^{(1)}) + (2\alpha - 1) \ln \frac{\alpha}{1-\alpha}. \quad \square$$

**Remark 2.2** Proposition 2.2 shows that, in general, $\delta$ is related to a weighted sum of $I(\boldsymbol{p}^{(1)}|\boldsymbol{p}^{(2)})$ and $I(\boldsymbol{p}^{(2)}|\boldsymbol{p}^{(1)})$. In particular, when $\alpha = .5$ (or, $\sum\sum_{i>j} p_{ij} = \sum\sum_{i<j} p_{ij}$), $\delta$ is (1/2) times Jeffrey's divergence between $\boldsymbol{p}^{(1)}$ and $\boldsymbol{p}^{(2)}$. $\square$

To obtain a measure between 0 and 1, which makes it easier to understand its behavior (Bishop *et al.*, 1975), we make the following transformation

$$\psi = 1 - e^{-\delta}. \tag{2.4}$$

This transformation is appropriate for ordinal categorical random variables (assumed to derive from an underlying continuous random variable) (Joe, 1989; Soofi et al., 1995). Thus $\psi = 0$ corresponds to 'symmetry' and $\psi = 1$ corresponds to 'complete asymmetry' as $p_{ji} = 0,\ p_{ij} > 0$ for some pair $(j, i)$.

It is insightful to evaluate the measure $\delta$ for well-known asymmetry models as follows, which show how $\delta$ is related to the source of asymmetry.

i) *Conditional symmetry* holds if $p_{ij} = p_{ji}k$ for $i < j$ and a constant $k > 0$. Since $\sum_{j \neq i} p_{ij} = (k+1)\sum_{i<j} p_{ji}$, we get from (2.3),

$$\delta = \frac{(k-1)\ln k \sum_{i<j} p_{ji}}{(k+1)\sum_{i<j} p_{ji}} = \frac{(k-1)}{(k+1)} \ln k,$$

which is an increasing (decreasing) function of $k$ for $k > 1$ ($k < 1$). For large $k$, $(k-1)/(k+1) \approx 1$, we have $\delta \approx \ln k$. Note asymmetries for $k$ and $1/k$ are same.

(ii) *Diagonal parameter symmetry* holds if $p_{ij} = p_{ji}k_t$, $t = j - i$ for $i < j$, $k_t > 1$. Here we get from (2.3),

$$\delta = \frac{\sum\sum_{i<j} p_{ji}(k_t - 1)\ln k_t}{\sum\sum_{i<j} p_{ji}(k_t + 1)} = \frac{\sum_{t=1}^{r-1}\sum_{i=1}^{r} p_{i+t,i}(k_t - 1)\ln k_t}{\sum_{t=1}^{r-1}\sum_{i=1}^{r} p_{i+t,i}(k_t + 1)},$$

which is an increasing function of $k_t$'s, $\forall t$. Then $\delta = \sum_{t=1}^{r-1} w_t \ln k_t$ where $w_t = \sum_{i=1}^{r} p_{i+t,i}(k_t - 1)/\sum_{t=1}^{r-1}\sum_{i=1}^{r} p_{i+t,i}(k_t + 1)$. For large $k_t$, $(k_t - 1) \approx (k_t + 1)$, $\forall t$, we have $\sum_{t=1}^{r-1} w_t = 1$.

(iii) *Linear diagonal parameter symmetry* holds if $p_{ij} = \phi^{j-i} p_{ji}$, $\phi > 1$ for $i < j$. Define $p_{i\cdot} = \sum_{j=1}^{r} p_{ij}, \forall i$. Here we get from (2.3), $\delta = \frac{\ln\phi \sum_{t=1}^{r-1} t(\phi^t - 1)p_{t\cdot}}{\sum_{t=1}^{r-1}(\phi^t + 1)p_{t\cdot}}$, which is an increasing function of $\phi$. Then $\delta = (\ln\phi)[\sum_{t=1}^{r-1} t w_t]$ where $w_t = (\phi^t - 1)p_{t\cdot}/\sum_{t=1}^{r-1}(\phi^t + 1)p_{t\cdot}$. For large $\phi$, $(\phi^t - 1) \approx (\phi^t + 1), \forall t$, we have $\sum_{t=1}^{r-1} w_t = 1$.

(iv) *Quasi-symmetry* holds if $p_{ij} = \alpha_i \beta_j \gamma_{ij}$ where $\gamma_{ij} = \gamma_{ji} > 0$ for $i < j$. Here we get from (2.3), $\delta = \dfrac{\sum\sum_{i<j}(\alpha_i\beta_j - \alpha_j\beta_i)\gamma_{ij} \ln \frac{\alpha_i\beta_j}{\alpha_j\beta_i}}{\sum\sum_{i<j}(\alpha_i\beta_j + \alpha_j\beta_i)\gamma_{ij}}$. This is equal to 0 if $\alpha_i = \beta_i$, $\forall i$, which is *marginal homogeneity*. Since symmetry model is equivalent to marginal homogeneity and quasi-symmetry together, so $\delta > 0$ is possible when quasi-symmetry holds but not marginal homogeneity. Thus here $\delta$ is a measure of the marginal *in*homogeneity of the $p_{ij}$'s when quasi-symmetry holds. For example, if $\beta_i = c_i\alpha_i$, $c_i \geq 1\ \forall i$, then $\delta = \sum\sum_{i<j} w_{ij}(c_j - c_i)\ln(c_j/c_i)$, $w_{ij} = \alpha_i\alpha_j\gamma_{ij}/\sum\sum_{i<j}(c_j + c_i)\alpha_i\alpha_j\gamma_{ij}$. □

For nominal categorical variables, we suggest the transformation

$$\tau = -\frac{\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} \ln \frac{p_{ij}}{p_{ji}}}{\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} \ln p_{ji}}. \tag{2.5}$$

Since $0 \leq \sum_{i=1}^r \sum_{j=1}^r p_{ij} \ln p_{ij} - \sum_{i=1}^r \sum_{j=1}^r p_{ij} \ln p_{ji} \leq -\sum_{i=1}^r \sum_{j=1}^r p_{ij} \ln p_{ji}$, it follows that $0 \leq \tau \leq 1$, where $\tau = 0$ corresponds to 'symmetry' and $\tau = 1$ corresponds to 'complete asymmetry' as $p_{ij} = 1$ for some pair $(i, j)$.

Next we consider the sampling distribution of an estimate of $\delta$. Let $n_{ij}$ denote the observed frequency in the $i$th row and $j$th column of the square table resulting from a full multinomial sampling. Let $\hat{\delta}$ be the sample version of $\delta$ obtained by replacing $p_{ij}$ with $\hat{p}_{ij} = n_{ij}/n$ where $n = \sum_{i=1}^r \sum_{j=1}^r n_{ij}$ in (2.3). Let $\hat{\psi} = 1 - e^{-\hat{\delta}}$. As $n \to \infty$, $\hat{\psi}$ converges almost surely to $\psi$. Let $\boldsymbol{p}$ ($\hat{\boldsymbol{p}}$) denote the vector obtained by stacking the elements of $p_{ij}$ ($\hat{p}_{ij}$) in a column. Since $\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p}) \to \boldsymbol{N}(\boldsymbol{0}, Diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^T)$, using the multivariate delta method (Serfling, 1980), it can be shown that if $\delta > 0$, $\sqrt{n}(\hat{\delta} - \delta)$ converges in distribution to a normal distribution with mean 0 and variance $\sigma_\delta^2$ where

$$\sigma_\delta^2 = \frac{\sum\sum_{i \neq j} \left[ \frac{(p_{ij} - p_{ji})^2}{p_{ij}} + \left( \ln \frac{p_{ij}}{p_{ji}} \right)^2 p_{ij} + 2(p_{ij} - p_{ji}) \ln \frac{p_{ij}}{p_{ji}} \right]}{(\sum\sum_{i \neq j} p_{ij})^2} - \delta^2.$$

The rate of convergence to asymptotic normality is slower as $\delta$ gets closer to 0. If $\delta = 0$, then using the second-order delta method it follows that $n(\hat{\delta} - \delta)$ converges in distribution to a chi-squared distribution with 1 degree of freedom. Provided $\hat{\delta}$ and $n$ are sufficiently large, an estimated standard error for $\hat{\delta}$ is $\hat{\sigma}_\delta/\sqrt{n}$ where $\hat{\sigma}_\delta^2$ denote the estimated variance obtained by replacing $p_{ij}$ with $\hat{p}_{ij}$ in the expression of $\sigma_\delta^2$. Then $\hat{\delta} \pm z_{\alpha/2}\hat{\sigma}_\delta/\sqrt{n}$ is an approximate $100(1-\alpha)\%$ confidence interval for $\delta$, where $z_{\alpha/2}$ is the $\alpha/2$th upper quantile of the standard normal distribution. From this, an interval estimate for $\psi$ can be obtained by using the transformation (2.4). Alternatively, an interval estimate can be obtained through the standard error $\hat{\sigma}_\psi^2 = e^{-2\hat{\delta}}\hat{\sigma}_\delta^2/\sqrt{n}$.

$\hat{\delta}$ is also normal under other sampling schemes if $\delta > 0$. Consider independent row multinomial sampling in which a random sample of size $n_i$ is taken from the subpopulation of each row. Let $n = \sum_i n_i$, and, $p_{i\cdot} = n_i/n$. Writing $p_{ij} = p_{j|i}p_{i\cdot}$, where $p_{j|i}$ is the conditional probabilities in the $i$th row. Denoting the $\delta$ in this case by $\delta_r$, it can be written as $\delta_r = \dfrac{\sum\sum_{i<j}(p_{j|i}p_{i\cdot} - p_{i|j}p_{j\cdot}) \ln \frac{p_{j|i}p_{i\cdot}}{p_{i|j}p_{j\cdot}}}{\sum\sum_{i<j}(p_{j|i}p_{i\cdot} + p_{i|j}p_{j\cdot})}$. Let $\boldsymbol{p}_{\cdot|i} = (p_{1|i}, p_{2|i}, \ldots, p_{r|i})^T$

be the vector of conditional probabilities of the $i$th row with transpose $\boldsymbol{p}_{\cdot|i}^T$. Let $\hat{\delta}_r$ be obtained from $\delta_r$ by replacing $p_{j|i}$ by $\hat{p}_{j|i} = n_{ij}/n_i$. Since $\sqrt{n_i}(\hat{\boldsymbol{p}}_{\cdot|i} - \boldsymbol{p}_{\cdot|i}) \to \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ where $\boldsymbol{\Sigma}_i = Diag(p_{1|i}, p_{2|i}, \ldots, p_{r|i}) - \boldsymbol{p}_{\cdot|i}\boldsymbol{p}_{\cdot|i}^T$, it can be shown that as $\min_i n_i \to \infty$, $\sqrt{n}(\hat{\delta}_r - \delta_r)$ converges in distribution to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance given by $\sigma^2_{\delta_r} = \sum_{i=1}^r p_{i\cdot} \left[\sum_{j=1}^r p_{j|i} a_{ij}^2 - \left(\sum_{j=1}^r p_{j|i} a_{ij}\right)^2\right]$, where $a_{ij} = (p_{j|i}p_{i\cdot} - p_{i|j}p_{j\cdot})/p_{j|i}p_{i\cdot} + \ln(p_{j|i}p_{i\cdot}/p_{i|j}p_{j\cdot})$. An estimated standard error can be obtained from the above expression of $\sigma^2_{\delta_r}$ by replacing $p_{j|i}$ by $\hat{p}_{j|i}$. Asymptotic variances for other sampling schemes can be obtained in a similar way.

# 3 Univariate Random Variables

For univariate discrete random variables taking $2k+1$ $[2k]$ values $(-k, \ldots, 0, \ldots, k)$ $[(-k, \ldots, -1, 1, \ldots, k)]$ with probabilities $\boldsymbol{p} = (p_{-k}, \ldots, p_k)$, *symmetry* (around 0) corresponds to $p_i = p_{-i}$, for $1 \le i \le r$. The asymmetry measure is

$$\delta = \frac{1}{1-p_0} \sum_{i=-k}^{k} p_i \ln \frac{p_i}{p_{-i}} \qquad \left[\delta = \sum_{i=-k}^{k} p_i \ln \frac{p_i}{p_{-i}}, \quad \text{respectively}\right]. \tag{3.1}$$

In any case, let $\psi = 1 - e^{-\delta}$ so that $0 \le \psi \le 1$.

**Example 3.1** For a binomial random variable (Bin$(n,p)$), the probability function is $p_i = b_{ni}p^i(1-p)^{n-i}$, $-k \le i \le k$, where $p = .5 + c$, $-.5 \le c \le .5$, and $b_{ni}$ is the binomial coefficient. Its mirror image is also a binomial random variable (Bin$(n, 1-p)$) with probability function as $p_{-i} = b_{ni}(1-p)^i p^{n-i}$. Then, as $1 - p = .5 - c$, it follows that $\delta(c) = 2nc \ln\left(\frac{.5+c}{.5-c}\right)$, $\psi(c) = 1 - e^{-\delta(c)} = 1 - \left(\frac{.5-c}{.5+c}\right)^{2nc}$. For $n = 11$, Table 1 sh ows values of $\delta(c)$, $\psi(c)$ increase fast with $c$. The values of $\delta(c)$, $\psi(c)$ remain unchanged

Table 1: The $\delta(c)$ and $\psi(c)$ for the binomial distribution with $n = 11, p = .5 + c$

| c | .00 | .03 | .06 | .09 | .12 | .15 | .18 | .21 | .24 | .50 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\delta(c)$ | .0 | .079 | .318 | .721 | 1.292 | 2.043 | 2.985 | 4.137 | 5.523 | $\infty$ |
| $\psi(c)$ | .0 | .076 | .273 | .514 | .725 | .870 | .949 | .984 | .996 | 1.0 |

with sign of $c$. Such tables are useful for calibrating (McCullagh, 1989) asymmetry, as discussed in Section 5. □

Now we consider the case of continuous random variables. For a continuous random variable $X$ of pdf $f(x)$ with a mirror image $f^\dagger(x)$ around $c$, we define

$$\delta(c) = \int f(x) \ln\left(\frac{f(x)}{f^\dagger(x)}\right) dx, \tag{3.2}$$

and, $\psi(c) = 1 - e^{-\delta(c)}$, to get a number between 0 and 1. The mirror image $f^\dagger(x)$ and $\delta(c)$ in (3.2) depend on $c$ chosen. If two mirror images of $f$ are considered around the points of reflection $c_1$ and $c_2$, then one can be obtained from the other by moving left or right a distance of $|c_1 - c_2|$. So, for measuring asymmetry of $f$, we prescribe convenient choices of $c$ depending on the domain: midpoint for a bounded interval, zero for the real line, and then compute $\delta(c)$.

Since $f^\dagger(x) = f(y)$ for some $x, y$, and $(f^\dagger)^\dagger(x) = f(x)$, we get $\int f(x) \ln(f(x)/f^\dagger(x))$ $dx = \int f^\dagger(x) \ln(f^\dagger(x)/f(x)) dx$, as stated in the next result.

**Proposition 3.1** The measure $\delta(c)$ in (3.2) is invariant to the choice of $f, f^\dagger$. □

Using the df $F$, it is easy to see that $\delta(c)$ can be expressed as $\delta(c) = \int_0^1 \ln\left(\frac{f(F^{-1}(u))}{f^\dagger(F^{-1}(u))}\right)$ $du$. The next result, easy to verify, gives a sufficient condition for $f$ (with df $F$) to be more asymmetric than another pdf $g$ (with df $G$).

**Proposition 3.2** If the dfs $F, G$ are strictly increasing and $\frac{f(F^{-1}(u))}{f^\dagger(F^{-1}(u))} > \frac{g(G^{-1}(u))}{g^\dagger(G^{-1}(u))}, \forall 0 \leq u \leq 1$, then $\delta_f(c) \geq \delta_g(c)$. □

We consider three different cases based on the domain of $f(x)$.

**Case 1.** If the pdf $f(x)$ has support on a bounded interval, such as $(a, b)$, then we consider $c = (a + b)/2$, and compare $f(x)$ with $f^\dagger(x) = f(a + b - x)$ (mirror image around the midpoint) as follows:

$$\delta = \delta((a + b)/2) = \int_a^b f(x) \ln\left(\frac{f(x)}{f(a + b - x)}\right) dx. \tag{3.3}$$

Then consider $\psi = 1 - e^{-\delta}$.

**Example 3.2** Let $X \sim \text{beta}(\alpha, \beta)$. Here $a = 0, b = 1$. It is well known that $1 - X \sim \text{beta}(\beta, \alpha)$. If the two corresponding pdf's are denoted by $f(x)$ and $f^\dagger(x)$, respectively, then it can be shown from (3.3) that $\delta = (\alpha - \beta)(\phi(\alpha) - \phi(\beta))$ where $\phi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha) = (d/d\alpha)(\ln \Gamma(\alpha))$ is the digamma function. The graph in Figure 1 shows the values of $\psi = 1 - e^{-\delta}$ versus $\alpha$ when $\beta = 3$. When $\alpha = 3$, its value is 0, which leads to symmetry. The asymmetry behavior can be seen as one moves away from $\alpha = 3$ in either direction. This graph is not symmetric around $\alpha = 3$. So, for example, the asymmetry (skewed right) of $\text{beta}(\alpha = 1, \beta = 3)$ pdf is different from that of (skewed left) $\text{beta}(\alpha = 5, \beta = 3)$ pdf; in fact, they are $\psi = .95$ and $\psi = .69$, respectively. Thus the former is more asymmetric than the latter. Figure 1 also includes graphs of $\psi_1, \psi_2$ from (1.1). Values of $\psi$ are much larger than those of $\psi_1, \psi_2$ for most of the range although near symmetry $\psi$ is smaller than both. □

**Case 2.** Next we consider the pdf's which have support on the entire real line. For such a pdf $f$ with median $m$, let $f_1(x) = f(x - m)$. Then consider asymmetry of $f_1$ around zero as follows:

$$\delta = \int_{-\infty}^{\infty} f_1(x) \ln\left(\frac{f_1(x)}{f_1(-x)}\right) dx = \int_{-\infty}^{\infty} f(x - m) \ln\left(\frac{f(x - m)}{f(m - x)}\right) dx. \tag{3.4}$$

Then consider $\psi = 1 - e^{-\delta}$ to get a number between 0 and 1.

**Example 3.3** Let $Z \sim \text{SN}(\lambda), \lambda \in \mathcal{R}$ (Azzalini, 1985), known as the *skew-normal* distribution, which reduces to the standard normal when $\lambda = 0$. Here $f(z) = 2\phi(z)\Phi(\lambda z),\ z \in \mathcal{R}$, where $\phi(z), \Phi(z)$ are the standard normal pdf and df, respectively. As $\lambda \to \infty$, the distribution approaches to a half-normal distribution on $(0, \infty)$. Also $-Z \sim \text{SN}(-\lambda)$. Then from (3.4), with $m_\lambda =$ median of $Z$, we get

$$\delta(\lambda) = \int_{-\infty}^{\infty} 2\phi(z - m_\lambda)\Phi(\lambda(z - m_\lambda)) \ln\left(\frac{\Phi(\lambda(z - m_\lambda))\phi(z - m_\lambda)}{\Phi(\lambda(m_\lambda - z))\phi(m_\lambda - z)}\right) dz.$$

The graph in Figure 2 shows values of $\psi(\lambda) = 1 - e^{-\delta(\lambda)}$ versus $\lambda$. As expected, $\lambda = 0$ touches the $x$-axis which corresponds to symmetry (or, the standard normal distribution). Unlike the beta distribution (in Example 3.2), here asymmetries on the left and right sides of zero are identical. As $\lambda$ deviates from 0, the pdf becomes increasingly more asymmetric. $\square$

**Case 3.** Next we consider the pdf's which have support on half-infinite intervals, e.g. $(0, \infty)$. Such pdf's are always skewed to the right. Still, for some such pdf's, there exists a $c_0$ such that pdf's are approximately symmetric on $(0, c_0)$, and for $c > c_0$, the pdf becomes increasingly more asymmetric on $(0, c)$. This is truly the case for many gamma pdf's, although not the exponential (which is always strongly skewed to the right). For such cases we propose to compare the properly normalized versions of $f(x)$ and $f^\dagger(x) = f(c - x)$ (mirror image around $c/2$) on the interval $(0, c)$, for $c > 0$. Thus we define

$$\delta(c) = \frac{1}{T} \int_0^c f(x) \ln\left(\frac{f(x)}{f(c-x)}\right) dx \tag{3.5}$$

where $T = \int_0^c f(x)dx$. Then consider $\psi(c) = 1 - e^{-\delta(c)}$ to get a number between 0 and 1. As $c$ moves from 0 to $\infty$, the asymmetry pattern on $(0, \infty)$ becomes evident.

**Example 3.4** Let $X \sim \text{gamma}(\alpha, \beta)$. Then it can be shown from (3.5) that

$$\delta(c) = \frac{\displaystyle\int_0^c x^{\alpha-1} e^{-x/\beta} \left[(\alpha - 1) \ln\left(\frac{x}{c-x}\right) + \frac{c-2x}{\beta}\right] dx}{\displaystyle\int_0^c x^{\alpha-1} e^{-x/\beta} dx}.$$

The graph in Figure 3 shows the values of $\psi(c) = 1 - e^{-\delta(c)}$ versus $c$ for various gamma pdf's. It is seen that all pdf's are close to symmetry for some $c$. For larger values of $\alpha, \beta$ (e.g., 15, 17, respectively), the pdf's are close to symmetry for a larger interval of $c$. For smaller values of $\alpha, \beta$, the pdf's are close to symmetry for a smaller interval of $c$, and then become asymmetric very quickly. With close inspection, it can be seen that although these graphs look like inverted normal pdf (except exponential), they are *not* symmetric. For the case of exponential distribution (included in Figure 3), the above expression simplifies to $\delta(c) = (c - \beta + (c + \beta)e^{-c/\beta})/(\beta(1 - e^{-c/\beta}))$. $\square$

Oja (1981) suggested that skewness measures should keep the same absolute value for a linear transformation $h_1X + h_2$ of a random variable $X$. MacGillivray (1986)

.

**Fig. 1: Asymmetry measures for beta distribution with varying alpha when beta = 3**

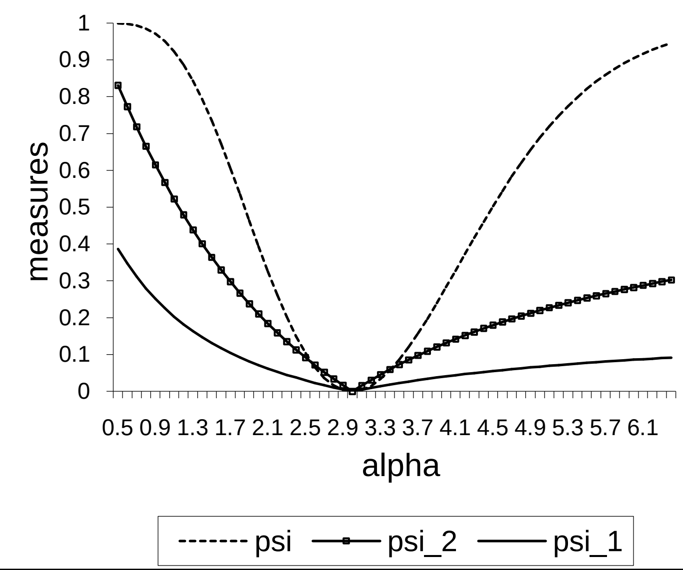



**Fig. 2: Skew-normal (lambda) pdf**

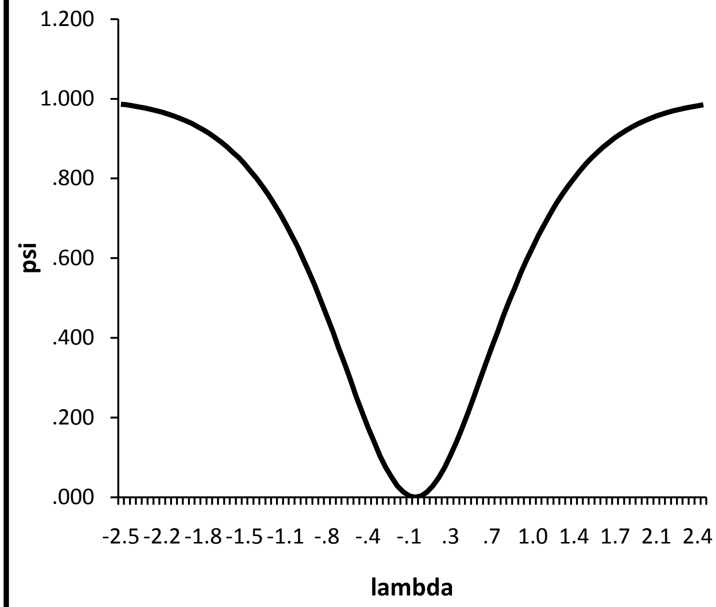
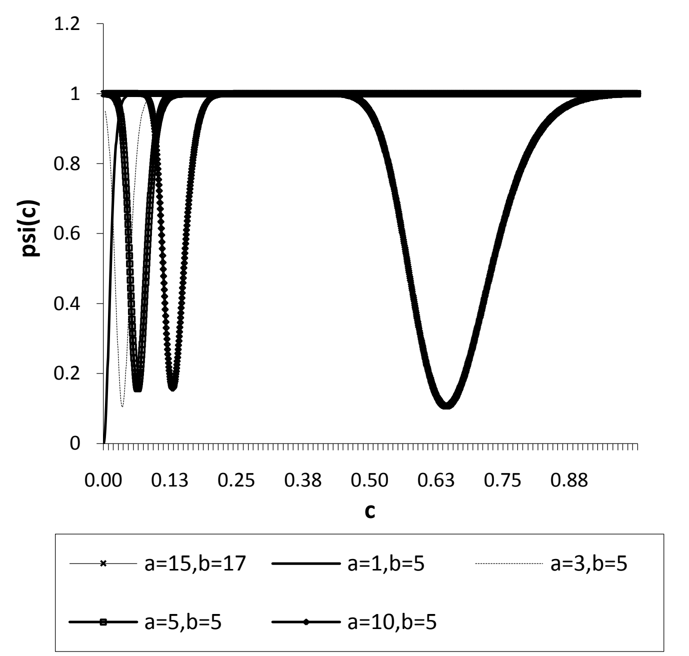



**Fig. 3: Gamma pdf, psi(c) versus c**



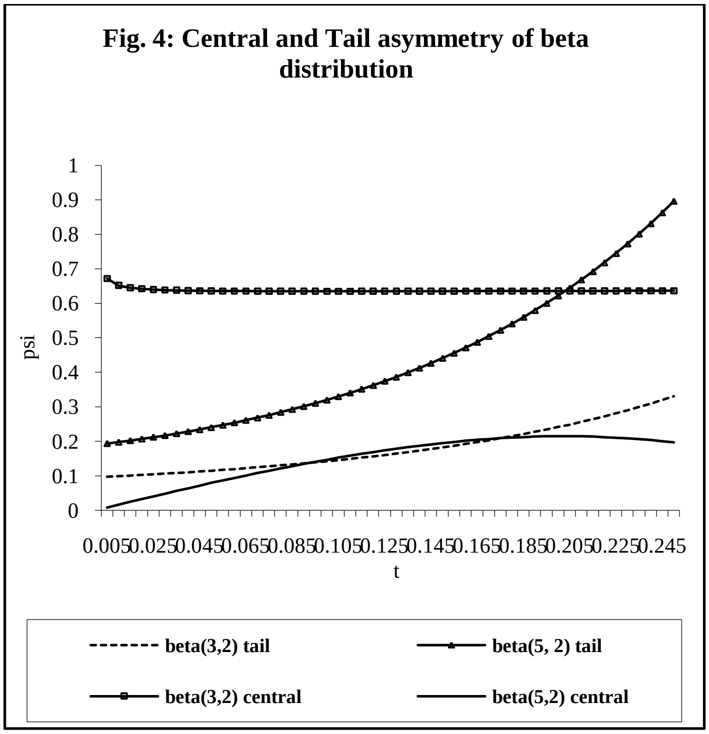


Fig. 4: Central and Tail asymmetry of beta distribution
psi
t
0.005 0.025 0.045 0.065 0.085 0.105 0.125 0.145 0.165 0.185 0.205 0.225 0.245
beta(3,2) tail
beta(5, 2) tail
beta(3,2) central
beta(5,2) central

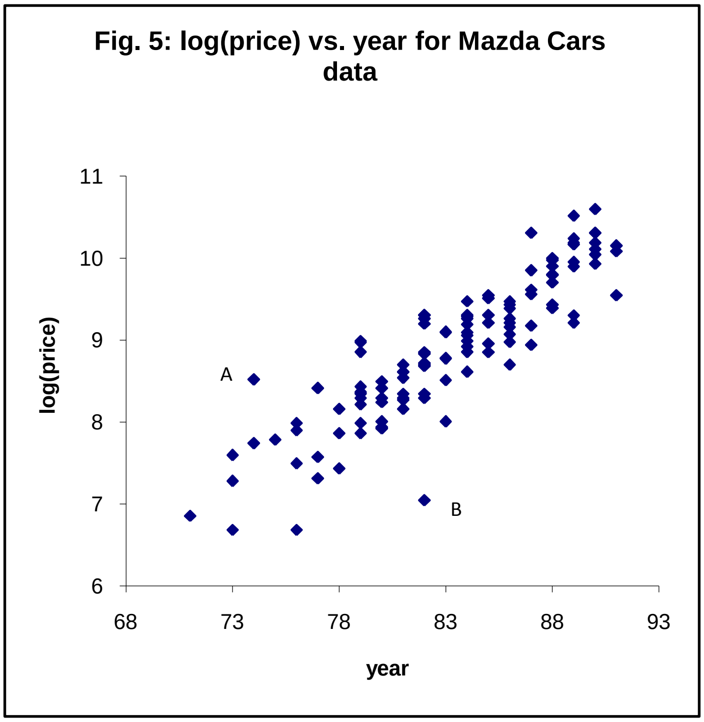


Fig. 5: log(price) vs. year for Mazda Cars data
log(price)
year
A
B

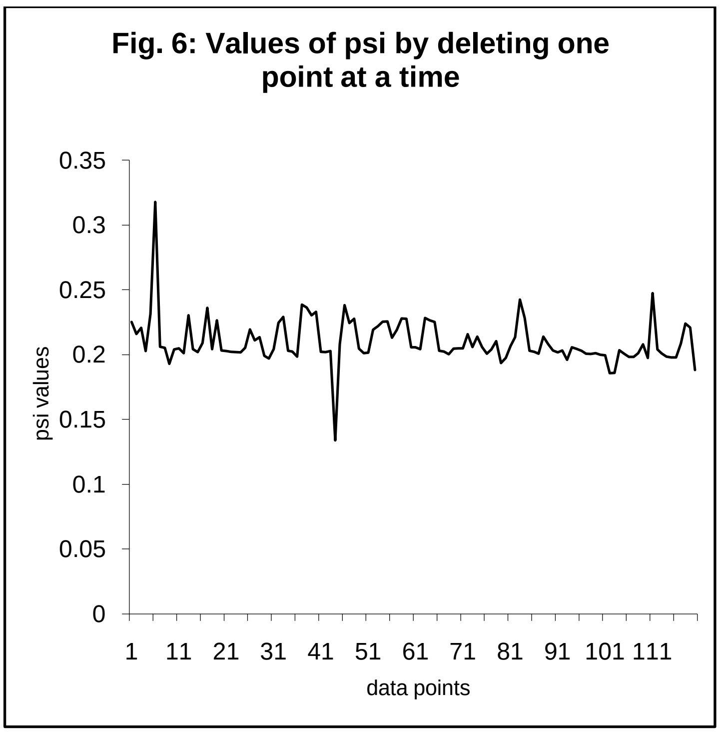


Fig. 6: Values of psi by deleting one point at a time
psi values
data points

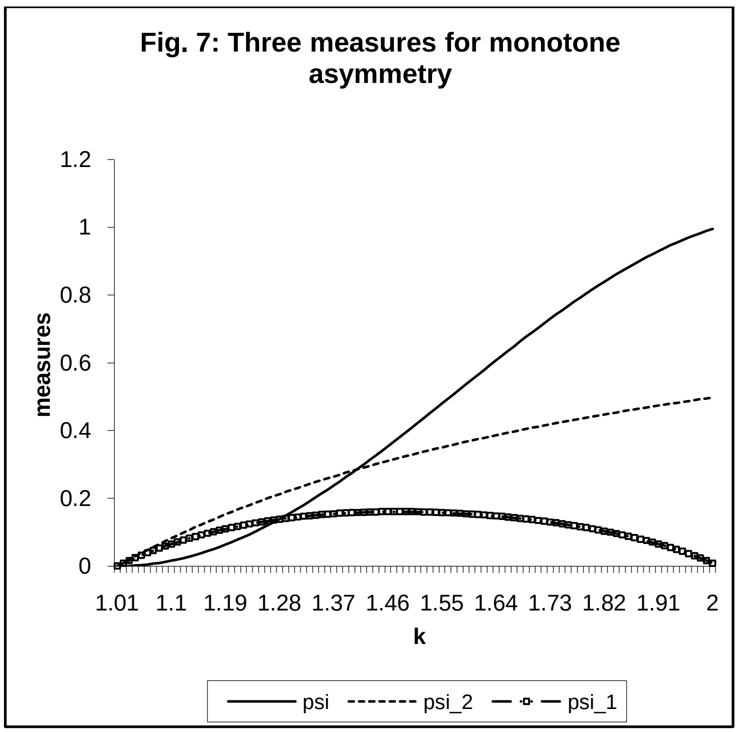


Fig. 7: Three measures for monotone asymmetry
measures
k
psi
psi_2
psi_1

showed this to be true for the measures she suggested. The same property holds for the measure $\delta$, for specific choices of $c$ when the domain of $f$ is of the form of $(a, b)$ or $\mathcal{R}$. When $f$ has domain $(0, \infty)$, we consider bounded intervals $(0, c)$, as $c \to \infty$.

**Proposition 3.3** For given constants $h_1, h_2$, $\delta_X(c) = \delta_{h_1X+h_2}(c)$, where $c =$ midpoint of $(a, b)$ when domain of $f$ is $(a, b)$, and, $c = 0$ when domain of $f$ is $\mathcal{R}$. □

The above $\delta$ in (3.2) considers the asymmetry of the entire pdf $f$. Sometimes the asymmetry of $f$ in more specific parts of it is of interest. Below we consider the asymmetries of $f$ near and away from the center.

When considering the 'central asymmetry' of $f$ with its median as center, we like to compare the asymmetries of $f$ to the left and right sides of its median. Thus we consider the pdf's,

$$\begin{aligned} f_1(x) &= f(x)/(F(m+t) - .5),\ m \le x \le m+t, \\ f_2(x) &= f(x)/(.5 - F(m-t)),\ m-t \le x \le m. \end{aligned} \tag{3.6}$$

These pdf's have different domains. We make a transformation to $f_2$ so that both pdf's are defined on the same domain, $(m, m+t)$ for $t > 0$. Thus we define

$$f_2^\dagger(x) = f(2m - x)/(.5 - F(m-t)),\ m \le x \le m+t. \tag{3.7}$$

The relative entropy between $f_1(x)$ and $f_2^\dagger(x)$ is

$$\begin{aligned} \delta^c(t) &= \int_m^{m+t} \frac{f(x)}{F(m+t) - .5)} \ln\left(\frac{f(x)/(F(m+t) - .5)}{f(2m-x)/(.5 - F(m-t))}\right) dx \\ &= \frac{.5 - F(m-t)}{F(m+t) - .5} + \int_m^{m+t} \frac{f(x)}{F(m+t) - .5)} \ln\left(\frac{f(x)}{f(2m-x)}\right) dx \end{aligned}$$

Define $\psi^c(t) = 1 - e^{-\delta^c(t)},\ 0 < t < t_0^c$ , where $t_0^c$ is the largest value of $t$ for which $\delta^c(t)$ is defined. Plotting $\psi^c(t)$ versus $t$ $(0 < t < t_0^c)$, the central asymmetry of $f$ around the median $m$ is revealed.

When considering 'tail asymmetry', we like to compare the behavior of left and right tails of $f$ away from the median. Thus we like to compare between the distributions of $X|X > m+t$, and, $X|X < m-t$ for $t > 0$. The pdf's of these random variables are given by $f_3(x) = f(x)/(1-F(m+t)),\ x > m+t$, and, $f_4(x) = f(x)/F(m-t),\ x < m-t$, respectively. Considering the mirror image of $f_4(x)$ so that it has the same domain as $f_3(x)$, we define $f_4^\dagger(x) = f(2m-x)/F(m-t),\ x > m+t$. Using the relative entropy between $f_3(x)$ and $f_4^\dagger(x)$, we define

$$\begin{aligned}\delta^u(t) &= \int_{m+t}^{\infty} \frac{f(x)}{1-F(m+t)} \ln\left(\frac{f(x)/(1-F(m+t))}{f(2m-x)/F(m-t)}\right) dx \\ &= \ln\left(\tfrac{F(m-t)}{1-F(m+t)}\right) + \tfrac{1}{1-F(m+t)} \int_{m+t}^{\infty} f(x) \ln\left(\frac{f(x)}{f(2m-x)}\right) dx.\end{aligned}$$

Defining $\psi^u(t) = 1 - e^{-\delta^u(t)},\ 0 < t < t_0^u$ , where $t_0^u$ is the largest value of $t$ for which $\delta^u(t)$ is defined. Plotting $\psi^u(t)$ versus $t$, $(0 < t < t_0^u)$ the tail asymmetry of $f$ away from the median $m$ is revealed.

**Example 3.5** To compare the central asymmetry of beta(3,2) and beta(5,2) with respective medians $m = .614$ and $m = .736$, first form $f^\dagger(x) = f(2m-x)$ in each case. Then $\psi^c(t)$ and $\psi^u(t)$ are calculated and graphed in Figure 4 for $0 < t < .25$.

Note that for both pdf's the tail asymmetry increases at a faster rate as one moves away from the respective medians, more for beta(5,2) than beta(3,2). The asymmetry between the two pdf's is substantially more when $t = .25$ than when $t$ is closer to 0. The central asymmetry of beta(3,2) starts at around .68, initially it decreases slightly as $t$ increases, then stays constant at around .6; whereas for beta(5,2) the central asymmetry starts from being close to 0 increases up to around .2 at $t = .17$, but then decreases slightly. □

It is of interest to find when the central (or tail) asymmetry is changing with $t$, if at all. The next result gives a necessary and sufficient condition for $\delta^c(t)$ and $\delta^u(t)$ to be nonincreasing (nondecreasing) in $t > 0$.

**Proposition 3.4** Let

$$\begin{aligned}h_{t,y}^c(f) &= \left(\tfrac{f(F^{-1}(F(m+t)-(1-y)(F(m+t)-.5))}{f(2m-F^{-1}(F(m+t)-(1-y)(F(m+t)-.5))}\right)\left(\tfrac{.5-F(m-t)}{F(m+t)-.5}\right), \\ h_{t,y}^u(f) &= \left(\tfrac{f(F^{-1}(1-(1-y)\overline{F}(m+t)))}{f(2m-F^{-1}(1-(1-y)\overline{F}(m+t)))}\right)\left(\tfrac{F(m-t)}{1-F(m+t)}\right).\end{aligned}$$

Then $\delta^c(t)$ $[\delta^u(t)]$ is nonincreasing (nondecreasing) according as $h^c_{t,y}(f)$ $[h^u_{t,y}(f)]$ is nonincreasing (nondecreasing) in $t > 0$ for every $y \in (0, 1)$. $\square$

**Example 3.6** For the exponential(1) distribution, $F(x) = 1-e^{-x}, x > 0$, or $F^{-1}(y) = -\ln(1-y)$. Thus it follows that $f(F^{-1}(1-(1-y)\overline{F}(m+t))) = (1-y)e^{-m-t}$ and $f(2m - F^{-1}(1-(1-y)\overline{F}(m+t))) = (1-y)^{-1}e^{-m+t}$. Here $\ln(h^u_{t,y}(f)) = 2\ln(1-y) + (m-t) + \ln(1-e^{-m+t})$, which is nonincreasing in $t$ for every $y \in (0, 1)$. Since $\int_0^1 \ln(1-y)^2 dy = -2$, it follows that $\delta^u(t) = 2e^{m-t}(e^{-m+t} - 2)$. Thus the tail asymmetry of the exponential(1) distribution is nonincreasing in $t$. The expressions for $h^c_{t,y}(f)$, $\delta^c(t)$ can be calculated. $\square$

It would be of interest to find when $\delta^c(t)$ (or, $\delta^u(t)$) is free of $t$. The next result states this fact, the proof of which follows along similar lines as in Ebrahimi and Kirmani (1996).

**Proposition 3.5** $\delta^c(t)$ $[\delta^u(t)]$ is free of $t$ if and only if $F(m-t) = (1 - F(m+t))^{\beta}$, $[.5 - F(m-t) = (F(m+t) - .5)^{\beta}]$, $\beta > 0, \forall t > 0$. $\square$

# 4 Applications

**4.1 Two-way tables.** Tomizawa (1994) proposed the indices $\phi_1 = \frac{1}{b\ln 2}\sum\sum_{i\neq j} p_{ij} \ln\left(\frac{2p_{ij}}{p_{ij}+p_{ji}}\right)$, and, $\phi_2 = \frac{1}{b}\sum\sum_{i<j}\frac{(p_{ij}-p_{ji})^2}{p_{ij}+p_{ji}}$, where $b = \sum\sum_{i\neq j} p_{ij}$ for measuring asymmetry. The sample versions satisfy $\hat{\phi}_1 = \frac{G^2}{(2\ln 2)\sum\sum_{i\neq j} n_{ij}}$, $\hat{\phi}_2 = \frac{X^2}{\sum\sum_{i\neq j} n_{ij}}$, where $G^2 = 2\sum\sum_{i\neq j} n_{ij} \ln\left(\frac{2n_{ij}}{n_{ij}+n_{ji}}\right)$, $X^2 = 2\sum\sum_{i<j}\frac{(n_{ij}-n_{ji})^2}{n_{ij}+n_{ji}}$, are the likelihood ratio and the Pearson chi-square statistics, respectively. Both $\phi_1$, $\phi_2$ are scale-invariant and between 0 and 1. As the performance of these measures are similar, we compare the performance of $\psi$ in (2.4) with that of $\phi = \phi_1$ only using a simulation study.

**Monte Carlo Experiments 4.1.** We consider two models: Model 1: $p_{ij} = \theta\pi_{ij},\ i > j,\ p_{ii} = \pi_{ii},\ p_{ij} = (2-\theta)\pi_{ij},\ i < j$, for $0 < \theta < 2$, where $\pi_{ij} = \pi_{ji} = b_{ki}b_{kj}q^{i+j}(1-q)^{2k-i-j},\ 0 \le i, j \le k,\ 0 < q < 1$ ($b_{ki}$ = binomial coefficient). Symmetry corresponds to $\theta = 1$. Asymmetry results when $\theta \neq 1$. Model 2: $p_{ij} = c$, $\forall i, j$, $p_{21} = c + \theta$, $\theta > 0$, then $p_{ij}$ are normalized to sum to 1.

We consider 10,000 replications using a sample of size $n = 500$ on a $4 \times 4$ table ($k = 3$) using $q = .5$. In Table 2, we find the 95% confidence interval for each index and count the number of times 0 is in the confidence interval for different values of $\theta$. For each index, when $\theta = 1$ (or close) for model 1 and when $\theta = 0$ (or close) for model 2, approximately 95% intervals contain 0, and this percentage gets much lower as $\theta$ deviates from the symmetry value, as expected. This is more pronounced for $\psi$ than for $\phi$. Thus from Table 2, $\psi$ performs well compared to $\phi$. □

Table 2: Percent of times the confidence interval contains zero for different indices

Model 1

| $\theta$ | 1 | 1.02 | 1.04 | 1.06 | 1.08 | 1.10 | 1.14 | 1.18 | 1.22 | 1.26 | 1.30 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\phi$ | .9742 | .9719 | .9606 | .9404 | .9076 | .8478 | .6713 | .4105 | .1868 | .0557 | .0107 |
| $\psi$ | .9705 | .9706 | .9566 | .9341 | .8963 | .8323 | .6443 | .3806 | .1655 | .0503 | .0120 |

Model 2

| $\theta$ | 0 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 | .10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\phi$ | .9773 | .9708 | .9404 | .8813 | .7952 | .6641 | .5183 | .3706 | .2524 | .1465 | .0881 |
| $\psi$ | .9726 | .9648 | .9301 | .8661 | .7680 | .6377 | .4879 | .3439 | .2297 | .1322 | .0773 |

**Example 4.1. Information analysis of eye vision data.** We consider the famous data set of unaided distance vision of 7477 women (age 30 to 39) employed in Royal Ordinance factories from 1943 to 1946 (Stuart, 1953). The variables are the right eye grade and the left eye grade. From (2.4), we find $\psi = .01763$. A 95% confidence interval of $\psi$ is given by $(.00199, .03327)$. Since zero is not in this interval, we conclude the table is asymmetric (at 95% confidence level). Our interpretation is that perhaps there are some constraints which are operational in the data-generating process causing asymmetry. One possible constraint is that 'right-eye vision is better than left-eye vision'.

Two ways to quantify such constraints are (1) $p_{ij} \geq p_{ji}, \forall i < j$, and, (2) $\sum_{i=1}^{m} \sum_{j=1}^{4}$

$p_{ij} \geq \sum_{i=1}^{m} \sum_{j=1}^{4} p_{ji}, \forall 1 \leq m \leq 3$. Clearly, (1) is a stronger ordering than (2). The (I-projection) estimates under (1) are calculated as follows: if $\hat{p}_{ij}, \forall (i,j)$ satisfy (1), then they are the estimates under (1), otherwise the estimate in $(i,j)$th cell is $(p_{ij}p_{ji})^{1/2}$, and then normed to sum to 1. We find under model (1), $\psi^{(1)} = .01754$ (slightly smaller than $\psi$ as expected). Comparing $\psi$ with $\psi^{(1)}$, the relative increment=.0051, or .51% of asymmetry is explained by constraints (1) over unrestricted. Since the unrestricted estimates already satisfy (2), those are the estimates under (2) also. Here $I(\hat{\boldsymbol{p}}_2|\hat{\boldsymbol{p}}_1) = .005168$, which gives a measure of difference between constraints (1) and (2) with respect to this data set. It is somewhat weak as the estimates are same under (1) and (2) except at only two cells. Bhattacharya and Dykstra (1997) gives an algorithm for finding restricted estimates under convex constraints such as (1) and (2) above. Bhattacharya (2006) generalized this algorithm to work for the continuous case. □

**4.2 Linear Regression Models.** Consider the linear regression model $y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n$, where $\boldsymbol{x}_i$ is a $d \times 1$ vector of regressors (fixed) that may include a constant term, $\boldsymbol{\beta}$ is a $d \times 1$ vector of parameters, $\epsilon_i \sim f(\epsilon_i)$ are iid with mean zero. If $f$ is known to be normal, then the maximum likelihood estimator of $\boldsymbol{\beta}$ is the ordinary least square estimator. However, if $f$ is only known to be symmetric around zero, then the exact maximum likelihood estimator does not exist. In this case, Bickel (1982) has shown that $\boldsymbol{\beta}$ can be estimated adaptively and Newey (1988) has constructed adaptive estimators of $\boldsymbol{\beta}$ using a generalized method of moments approach. Thus it is useful to construct consistent tests that can verify the assumption of symmetry of the error distribution in regression.

To test $H_0 : f(\epsilon) = f(-\epsilon)$ a.e. against the alternative $H_1 : f(\epsilon) \neq f(-\epsilon)$ a.e., the appropriate discrepancy measure from (3.4) is

$$\delta = \int_{-\infty}^{\infty} f(\epsilon) \ln \left( \frac{f(\epsilon)}{f(-\epsilon)} \right) d\epsilon. \tag{4.1}$$

Using a sample of size $n$, we propose to estimate $\delta$ in (4.1) by

$$\frac{1}{n}\left[\sum_{i=1}^{n}\left(\ln \hat{f}(\hat{\epsilon}_i) - \ln \hat{f}(-\hat{\epsilon}_i)\right)\right] \tag{4.2}$$

where $\hat{\epsilon}_i = y_i - \boldsymbol{x}_i^T\boldsymbol{b}$, $\boldsymbol{b}$ is the ordinary least square estimator of $\boldsymbol{\beta}$ and $\hat{f}$ is the non-parametric kernel density estimator of $f$. Robinson (1991) has proposed an estimator, similar to (4.2), when testing symmetry of univariate observations, and also has derived its asymptotic distribution. His test statistic, when modified to the regression setting, turns out to be

$$R = \frac{n^{-1/2}\left(\sum_{t:\hat{\epsilon}_t \le m} \ln \hat{f}(\hat{\epsilon}_t) - \sum_{t:\hat{\epsilon}_t > m} \ln \hat{f}(\hat{\epsilon}_t)\right)}{\sqrt{\frac{1}{n_\nu}\sum_{t=1}^{n} c_t(\ln \hat{f}(\hat{\epsilon}_t))^2 - \left(\frac{1}{n_\nu}\sum_{t=1}^{n} c_t \ln \hat{f}(\hat{\epsilon}_t)\right)^2}}$$

where $c_t = 1$ if $\hat{\epsilon}_t \le m$, and $= 0$, otherwise, $m$ is the sample median, and $n_\nu = n$, if $n$ is odd, otherwise, $n_\nu = n + \nu$, $\nu \ge 0$. Robinson (1991) has shown that $R$ has an asymptotic standard normal distribution. For testing $H_0 : \delta = 0$ versus $H_1 : \delta > 0$, we reject $H_0$ when $R > 1.645$ for $\alpha = .05$. We denote this as 'R-test'.

Often the error distribution is assumed to be unimodal in nature. This information can be incorporated when estimating $f(x)$ in (4.1) using isotonic regression as follows (Robertson, Wright and Dykstra, 1988). Let $y_1, \ldots, y_n$ be the ordered sample from a pdf $f$. Assume the mode is $M$ (known), and there exists $a$ such that $y_a < M < y_{a+1}$. The density estimate is assumed to be zero on $(-\infty, y_1) \cup (y_n, \infty)$, and constant on each of the intervals $[y_{j-1}, y_j), 2 \le j \le a, [y_a, M), (M, y_{a+1}], (y_j, y_{j+1}], a+1 \le j \le n-1$. Define a partial order $\preceq$ on $\{1, \ldots, n\}$ by $1 \preceq 2 \preceq \ldots \preceq a,\ a+1 \succeq \ldots \succeq n$. The estimate (a step function) $\hat{f}$ is the isotonic regression with respect to $\preceq$ of $\boldsymbol{g} = (g_1, \ldots, g_n)$ with weights $\boldsymbol{w} = (w_1, \ldots, w_n)$, where $w_i = y_{i+1} - y_i, i = 1, \ldots, a-1, w_a = M - y_a, w_{a+1} = y_{a+1} - M, w_i = y_i - y_{i-1}, i = a+2, \ldots, n$ and $g_i = 1/nw_i, \forall i$. We use a test statistics as

$$D = \int_{-\infty}^{\infty} \hat{f}(\epsilon) \ln\left(\frac{\hat{f}(\epsilon)}{\hat{f}(-\epsilon)}\right) d\epsilon. \tag{4.3}$$

As the distribution of $D$ is unknown, we have simulated its percentiles in Table 3. We denote this as 'D-test'.

Next we describe two already available tests for testing the symmetry of the residuals. Fan and Gencay (1995) proposed the test statistic $T_\gamma^* = \frac{\sqrt{n}(\lambda_\gamma^* - 1)}{\tilde{\sigma}_{\gamma,0}^*}$, where $\lambda_\gamma^* = \tilde{\delta}_\gamma^* / \tilde{\Delta}(f)$, $\tilde{\delta}_\gamma^* = n_\gamma^{-1} \sum_{i=1}^n C_i(\gamma) \hat{f}(-e_i)$, $C_i(\gamma) = 1 + \gamma$ for $i$ odd, $= 1 - \gamma$ for $i$ even, $n_\gamma = n$ for $n$ even, $= n + \gamma$ for $n$ odd, $0 < \gamma \leq 1$, $\hat{\Delta}(f) = \int \hat{f}^2(\epsilon) d\epsilon$, and, $\hat{\sigma}_{\gamma,0}^{*2} = \frac{\gamma^2 \int \hat{f}(\epsilon)[\hat{f}(-\epsilon) - \tilde{\delta}_\gamma^*]^2 d\epsilon}{\hat{\Delta}^2(f)}$ with $\gamma = .65$, $\hat{f}$ the nonparametric kernel density estimator of $f$, $e_i$ is the $i$th residual, $\tilde{\Delta}(f), \tilde{\sigma}_{\gamma,0}^*$ are obtained from $\hat{\Delta}(f), \hat{\sigma}_{\gamma,0}^*$ respectively, by replacing $\epsilon$ with $e_i$. We denote this as 'FG-test'.

Bai and Ng (2005) has presented a test of symmetry of the residuals based on two odd moments, say $r_1$th and $r_2$th. For a sample $(X_1, \ldots, X_n)$, let $Y_n' = (\sum_{t=1}^n (X_t - \overline{X})^{r_1} / \sqrt{n}, \sum_{t=1}^n (X_t - \overline{X})^{r_2} / \sqrt{n})$. Defining $\alpha = [1 \;\; 0 \;\; -r_1 \mu_{r_1 - 1} \backslash 0 \;\; 1 \;\; -r_2 \mu_{r_2 - 1}]'$, $Z_t = ((X_t - \mu)^{r_1}, (X_t - \mu)^{r_2}, (X_t - \mu))'$ $\Gamma = \lim_{n \to \infty} n E(\overline{Z}\overline{Z}')$, $\overline{Z}$ is the sample average of the $Z_t$'s, the test statistic is $Y_n' (\hat{\alpha} \hat{\Gamma} \hat{\alpha}')^{-1} Y_n$, which has a chi-square distribution with 2 degrees of freedom, where $(\hat{\alpha} \hat{\Gamma} \hat{\alpha}')$ is a consistent estimate of $(\alpha \Gamma \alpha')$. They used $r_1 = 3, r_2 = 5$. We denote this as 'BN-test'.

**Monte Carlo Experiments 4.2.** We considered $n = 50$ and 100 with 2000 replications in each case. The symmetric distributions used are standard normal, $t$ with 5 degrees of freedom, and, double exponential. Listed below are asymmetric distributions:

A1: chi-squared with two degrees of freedom, A2: lognormal,

A3: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}] / \lambda_2, (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, 1, 1.4, .25)$

A4: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}] / \lambda_2, (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -1, -.0075, -.03)$,

A5: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}] / \lambda_2, (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -1, -.1, -.18)$,

A6: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}] / \lambda_2, (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -1, -.001, -.13)$,

A7: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1 - u)^{\lambda_4}] / \lambda_2, (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, -1, -.0001, -.17)$,

of which A3-A7 contain the inverse distribution functions of the generalized lambda family with lambda parameters taken from table 1 of Randles et al. (1980). The

$n \times d$ regressor matrix $\boldsymbol{X}$ is constructed the same way as Fan and Gencay (1995).

Table 3: Size of $R$ test and percentiles of the $D$ test under symmetry

| | Size of $R$ test | | | | Percentiles of the $D$ test | | | |
|---|---|---|---|---|---|---|---|---|
| | $n = 50$ | | $n = 100$ | | $n = 50$ | | $n = 100$ | |
| Distributions | 5% | 10% | 5% | 10% | 90% | 95% | 90% | 95% |
| $N(0, 1)$ | .047 | .095 | .047 | .083 | .450 | .517 | .268 | .313 |
| $t_5$ | .056 | .110 | .050 | .102 | .448 | .492 | .300 | .338 |
| DE | .068 | .112 | .052 | .101 | .477 | .542 | .299 | .343 |

With two regressors (including an intercept term), Table 3 presents the simulated size of the $R$ test. The overrejection or underrejection is limited by 2% of the nominal size in each case. The percentiles of the D test in Table 3 under different symmetric distributions are relatively insensitive, more so for higher $n$. Critical values of the $D$ test are constructed from this table by averaging these values. Similar observations are revealed with four and six regressors (available from author).

Table 4: Power comparison of the BN, FG and R tests at 5% level

| | $n = 50$ | | | | $n = 100$ | | | |
|---|---|---|---|---|---|---|---|---|
| Distributions | BN | FG | R | D | BN | FG | R | D |
| | Two regressors | | | | | | | |
| $A1$ | .685 | .934 | .996 | .991 | .957 | .998 | 1.0 | 1.0 |
| $A2$ | .399 | .994 | .999 | .999 | .627 | 1.0 | 1.0 | 1.0 |
| $A3$ | .336 | .465 | .793 | .761 | .733 | .744 | .968 | .988 |
| $A4$ | .349 | .545 | .851 | .703 | .816 | .806 | .989 | .963 |
| $A5$ | .090 | .287 | .480 | .330 | .245 | .420 | .737 | .569 |
| $A6$ | .548 | .946 | 1.0 | .995 | .847 | .998 | 1.0 | 1.0 |
| $A7$ | .511 | .951 | 1.0 | .994 | .799 | .997 | 1.0 | .1.0 |

For the asymmetric alternatives, the power values of the four tests are given in Table 4. Power values for all tests improve substantially with increase in $n$. All four tests have relatively lower powers under the alternative A5, which is close to symmetry. Powers are lower with increase in the number of regressors for all tests (not shown). This decrease is less when $n = 100$ than $n = 50$. Overall both proposed $R$ and $D$ tests performed very well in this simulation compared with BN and FG tests. The $R$ test uses sample median, and the $D$ test uses the unimodal order, whereas the BN and FG tests do not use this order information. □

**Example 4.2. Information analysis of Mazda cars data.** We apply the methods developed on a data set (available at statsci.org) of prices of 124 Mazda cars versus the year of purchase, see Figure 5. Rousseeuw and Struyf (2004) analyzed this data with a logarithmic transformation of the response variable (price) and established presence of linearity. Thus the model becomes $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where $Y$ = logarithm of the purchase price, and $X$ is the year of purchase, $i = 1, \ldots, 124$.

The calculated symmetry test from (4.2), using the residuals of the regression model,is -.7396, thus we do not reject the null hypothesis of symmetry at $\alpha = .05$ level. As the residuals have a unimodal pattern, one would like to use this information. Using (4.3), we find the observed asymmetry as $D = .2105$ using 120 nontied observations. Although this is not very large, to find out which pair(s) of values is/are most responsible for the observed asymmetry, we have considered 120 separate regressions of 119 pairs each, keeping $i$th ($1 \leq i \leq 120$) pair out. The 120 values of $D$ from these regressions are graphed in Figure 6. When we keep the point B out, the value of $D$ is closest to symmetry (spike down), and when we keep the point A out, $D$ is farthest from symmetry (spike up). Keeping any other observation out does not make as much difference as these two. It is immediate that points A and B are most influential in assessing asymmetry of this data. By inspecting data, it can be seen that the point B corresponds to a unusually low price for the year 1982, and the point A corresponds to a unusually high price for the year 1974. Usual regression analysis in this example shows that the errors are homoscedastic, and not normal. Thus the OLS estimator is not the most efficient one, and a generalized method of moments estimator such as Newey (1988) may be adopted. □

# 5 Comparison and Discussion

Although our proposed measure depends on the choice of relative entropy, our method of using $I(\boldsymbol{p}|\boldsymbol{p}^{\dagger})$ instead of $I(\boldsymbol{p}|\boldsymbol{p}^{s})$ or $I(\boldsymbol{p}^{s}|\boldsymbol{p})$, $\boldsymbol{p}^{s}$ being the symmetrized version of $\boldsymbol{p}$, (similarly for the continuous case) has a clear advantage. The estimators $\hat{\boldsymbol{p}}^{s}$ take different forms based on the criterion used (see Pardo, 2006, for related references): for two-way contingency tables, $p^{s}_{ij} = (\hat{p}_{ij} + \hat{p}_{ji})/2$ (maximum likelihood estimators); $p^{s}_{ij} = c(\hat{p}_{ij}\hat{p}_{ji})^{1/2}$, $c =$ normalizing constant (minimum modified likelihood estimators); $p^{s}_{ij} = c(\hat{p}^{2}_{ij} + \hat{p}^{2}_{ji})^{1/2}$, $c =$ normalizing constant (minimum chi-square estimators); etc. Thus different levels of asymmetry would be obtained when one calculates $I(\hat{\boldsymbol{p}}|\hat{\boldsymbol{p}}^{s})$ or $I(\hat{\boldsymbol{p}}^{s}|\hat{\boldsymbol{p}})$ for same $\hat{\boldsymbol{p}}$. Using $I(\hat{\boldsymbol{p}}|\hat{\boldsymbol{p}}^{\dagger})$ instead, we avoid this ambiguity.

For the two-way contingency tables, the well-known test statistics $G^2$, $X^2$ (both with df $= r(r-1)/2$ for an $r \times r$ table) are not scale-invariant. The symmetry model imposes no restriction on the diagonal cell probabilities. Hence it seems natural that the measures of degree of departure from symmetry do not depend on the diagonal probabilities. The ranges of $G^2/n$ and $X^2/n$ depend on diagonal proportions (e.g. $0 \le G^2/n \le 2\ln 2(1 - \sum n_{ii}/n)$ and $0 \le X^2/n \le (1 - \sum n_{ii}/n)$). On the contrary, the estimate $\hat{\psi}$ does not depend on diagonal proportions and range in $(0,1)$.

There is no relationship between our ordering and the orderings given by $\psi_1, \psi_2$ in (1.1). For example, from Figure 1 for the beta distribution $\psi, \psi_1, \psi_2$ are all larger for $\alpha = 2$ than for $\alpha = 3$, but $\psi(\alpha = 2.4) < \psi(\alpha = 3.7)$, however $\psi_1(\alpha = 2.4) > \psi_1(\alpha = 3.7)$, $\psi_2(\alpha = 2.4) > \psi_2(\alpha = 3.7)$.

For monotone asymmetry such as $f(x) = k,\ 0 \le x \le .5, = ck,\ .5 \le x \le 1,\ 1 \le k \le 2$, $c$ is a norming constant, the values of $\psi, \psi_1, \psi_2$ are similar for a longer range (see Figure 7) than for nonmonotone asymmetry such as beta pdf's (Figure 1). From

Figure 7, when asymmetry is high, $\psi_1$ goes down, and rate of increment in values of $\psi_2$ slows down, but $\psi$ increases fast. Thus the behavior of $\psi$ is different from $\psi_1, \psi_2$ for such monotone asymmetries.

One could construct a general class of distance or divergence measures that includes the relative entropy by defining $\tau = \int f\phi\left(\frac{g}{f}\right) d\mu$, where $\phi$ is a strictly convex function on $\mathcal{R}$ satisfying $\phi(1) = 0$. Taking $\phi(u) = u \ln u$ leads to relative entropy, and $\phi(u) = u^2 - 1$ leads to $\tau_P = \int f^2/g d\mu - 1 = \int (f-g)^2/g d\mu$, which is Pearson's chi-square discrepancy. Thus for the density $f$ with its mirror image $f^\dagger = g$, $\tau_P = \int f^2/f^\dagger d\mu - 1$. However, when $f$ is beta$(\alpha, \beta)$ so that $f^\dagger$ is beta$(\beta, \alpha)$, it can be shown that $\tau_P = \frac{\Gamma(2\alpha-\beta)\Gamma(2\beta-\alpha)}{\Gamma(\alpha)\Gamma(\beta)}$. Thus $\tau_P$ is defined only for those $\alpha, \beta$ which satisfy $2\alpha - \beta > 0$ and $2\beta - \alpha > 0$. Since, $\tau_P$ is not defined for all $\alpha > 0, \beta > 0$ (because the gamma function is undefined when $2\alpha - \beta < 0$ or $2\beta - \alpha < 0$), so $\tau_P$ is not a very useful measure of asymmetry. Similarly, other choices of the convex function $\phi$, such as, $\phi(u) = u^t - 1, 1 < t < 2$, in the divergence $\int f\phi\left(\frac{g}{f}\right) d\mu$ will not lead to suitable measures of asymmetry.

Two pdf's $f_1, f_2$ with $\psi_1, \psi_2$ can be compared: if $\psi_1 \geq \psi_2$ then $f_1$ is more asymmetric than $f_2$, and the relative increment in asymmetry of $f_1$ over $f_2$ is $r(f_1|f_2) = 1 - \psi_2/\psi_1$. When $\psi_1 \approx \psi_2$, the asymmetry levels of $f_1, f_2$ are about the same. As $\delta$ cannot be expressed as a difference of two entropies, we have refrained from naming $r$ as a distinguishability index (Soofi et al, 1995) although it serves similar purpose.

For calibration purposes, one can seee from Table 1, binomial distributions with $n = 10$ and $p \geq .85$ or $p \leq .15$ have similar asymmetries (all with $\psi = 1$, approximately). Also, the distributions binomial($n = 10$, $p = .65$ or $p = .35$), beta($\alpha = 1.3, \beta = 3$), beta($\alpha = 5.6, \beta = 3$), skew-normal($\lambda = 1.45$) all have $\psi = .84$ (approximately), thus they have similar asymmetries. From Table 1, $r(\text{bin}(n = 10, p = .65)|\text{bin}(n = 10, p = .6)) = (.8439 - .5556)/.8439 = .3416$. Thus bin($n = 10, p = .65$) is 34.16% more asymmetric than bin($n = 10, p = .6$). □

## Appendix

*Proof of Proposition 2.2* Let $A = \sum\sum_{i>j} p_{ij},\ B = \sum\sum_{i<j} p_{ij}$. Then we get

$$\sum\sum_{i>j} p_{ij} \ln \frac{p_{ij}}{p_{ji}} + \sum\sum_{i>j} p_{ji} \ln \frac{p_{ji}}{p_{ij}} = AI(\boldsymbol{p}^{(1)}|\boldsymbol{p}^{(2)} + BI(\boldsymbol{p}^{(2)}|\boldsymbol{p}^{(1)}) + (A-B)\ln \frac{A}{B}.$$

Dividing both sides by $A+B$ and setting $\alpha = A/(A+B)$, one gets the result. □

*Proof of Proposition 3.3* Without loss of generality, assume $0 \le x \le 1$, and $y = ax+b$. For $a > 0$, $b \le y \le a+b$, and for $a < 0$, $a+b \le y \le b$. The result follows from the fact that the mirror image satisfies $g^\dagger(y) = g(a+2b-y)$. □

*Proof of Proposition 3.4* For $t > s > 0$, let $F_{m+t}(s) = P(m < X \le m+s | m < X \le m+t) = \frac{F(m+s)-.5}{F(m+t)-.5}$. Setting $\frac{F(m+s)-.5}{F(m+t)-.5} = y$, one gets $F(m+s) = F(m+t) - (1-y)(F(m+t) - .5)$, or

$$m + s = F^{-1}(F(m+t) - (1-y)(F(m+t) - .5)). \tag{5.1}$$

With $x = m+s$ in (3.6), (3.7), one gets $f_{m+t}(s) = f(m+s)/(F(m+t) - .5)$, and, $f^\dagger_{m+t}(s) = f(m-s)/(.5 - F(m-t))$. Then

$$\begin{aligned} \delta^c(t) &= \int_m^{m+t} \frac{f(x)}{F(m+t)-.5)} \ln\left(\frac{f(x)/(F(m+t)-.5)}{f(2m-x)/(.5-F(m-t))}\right) dx \\ &= \int_0^1 \ln\left(\frac{f_{m+t}(s)}{f^\dagger_{m+t}(s)}\right) dF_{m+t}(s). \end{aligned}$$

Using $y = F_{m+t}(s)$, forms of $f_{m+t}, f^\dagger_{m+t}$, and (5.1), the result follows. □


### Acknowledgments

The author thanks two referees and the editor for helpful comments, which have improved this manuscript. The author is grateful to Professors Y. Fan and R. Gencay, who made the Fortran codes of their procedure available. The author also thanks the National Science Foundation for partial support (DMS 0706041).

# References


Azzalini, A. (1985), A class of distributions which include the normal ones, *Scandinavian Journal of Statistics*, **12**, 171-178.

Bai, J. and Ng, S. (2005), Tests for skewness, kurtosis, and normality for time series data, *Journal of Business, Economics and Statistics*, **23**, 49–60.

Bhattacharya, B. (2006), An iterative procedure for general probability measures to obtain I-projection onto intersection of convex sets, *Annals of Statistics*, **34**, 878-902.

Bhattacharya, B. and Dykstra, R. (1997), A Fenchel duality aspect of iterative I-projection procedures, *Annals of the Institute of Statistical Mathematics.* **49**, 435-446.

Bickel, P. (1982), On adaptive estimation, *Annals of Statistics*, **10**, 647-671.

Chaffin, W. W. and Rhiel, G. S. (1993). The effect of skewness and kurtosis on the one-sample $t$ test and the impact of the knowledge of the population standard deviation. *J. Stat. Comp. Sim.* **46**, 79-90.

Doksum, K. A. (1975), Measures of location and asymmetry, *Scandinavian Journal of Statistics*, **2**, 11-22.

Ebrahimi, N. and Kirmani, S. N. U. A. (1996), A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions, *Biometrika*, **83**, 233-235.

Fan, Y. and Gencay, R. (1995), A consistent nonparametric test of symmetry in linear regression models, *Journal of the American Statistical Association*, **90**, 551-557.

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

Joe, H.(1989), Relative entropy measures of multivariate dependence, *Journal of the American Statistical Association*, **84**, 157-164.

Johnson, N. L. and Kotz, S. (1970), *Continuous Univariate Distributions*, Houghton-Mifflin, Boston.

MacGillivray, H. L. (1986), Skewness and asymmetry: measures and orderings, *Annals of Statistics*, **14**, 3, 994-1011.

McCulloch, R. E. (1989), Local model influence, *Journal of the American Statistical Association*, **84**, 473-478.

Newey, W. K. (1988), Adaptive estimation of regression models via moment restrictions, *Journal of Econometrics*, **38**, 301-339.

Oja, H. (1981), On location, scale, skewness and kurtosis of univariate distributions, *Scandinavian Journal of Statistics*, **8**, 154-168.

Ord, J. K. (1968), The discrete student's $t$ distribution, *Annals of Mathematical Statistics*, **39**, 1513-1516.

Pardo, L. (2006), *Statistical Inference Based on Divergence Measures*, Chapman & Hall/CRC, Boca Raton, FL.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* John Wiley & Sons, Inc., New York.

Robinson, P. M. (1991), Consistent nonparametric entropy based testing, *The Review of Economic Studies*, **58**, 437-453.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Soofi, E. S., Ebrahimi, N. and Habibullah, M. (1995), Information distinguishability with application to analysis of failure data, *Journal of the American Statistical Association*, **90**, 657-668.

Tomizawa, S. (1994), Two kinds of measures of departure from symmetry in square contingency tables having nominal categories, *Statistica Sinica*, **4**, 325-334.

van, Zwet, W. R. (1964), Convex transformations of random variables, *Math. Centrum.*, Amsterdam.

Wand, M.P. and Jones, M.C. (1996), *Kernel Smoothing*, Chapman and Hall.