HAL
open science

Some measures for asymmetry of distributions

Georgi N. Boshnakov

To cite this version:

Georgi N. Boshnakov. Some measures for asymmetry of distributions. Statistics and Probability Letters, 2009, 77 (11), pp.1111. 10.1016/j.spl.2007.02.002 . hal-00565447

HAL Id: hal-00565447

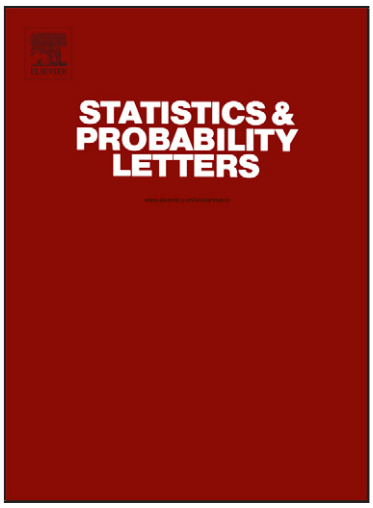
https://hal.science/hal-00565447v1

Submitted on 13 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript



www.elsevier.com/locate/stapro

Some measures for asymmetry of distributions

Georgi N. Boshnakov

PII: S0167-7152(07)00062-4
DOI: doi:10.1016/j.spl.2007.02.002
Reference: STAPRO 4583

To appear in: *Statistics & Probability Letters*

Received date: 10 July 2006
Revised date: 10 December 2006
Accepted date: 21 February 2007

Cite this article as: Georgi N. Boshnakov, Some measures for asymmetry of distributions, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.02.002

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Some measures for asymmetry of distributions

Georgi N. Boshnakov

School of Mathematics
The University of Manchester

Short running title: Measures for asymmetry

Georgi N. Boshnakov

School of Mathematics
The University of Manchester
Sackville str.
Manchester M60 1QD
UK

E-mail: `georgi.boshnakov@manchester.ac.uk`

Accepted manuscript

ACCEPTED MANUSCRIPT

Abstract

We propose several measures, functional and scalar, for asymmetry of distributions by comparing the behavior of probability densities to the right and left of the mode(s) and show how to generate classes of equivalent distributions from a given distribution, allowing for varying asymmetry but retaining some information theoretic properties of the original distribution, such as the entropy.

Key Words: asymmetry; asymmetry curve; asymmetry index; confidence characteristic

1 Introduction

The numerical characteristic normally employed to characterize the lack of symmetry of a distribution is the coefficient of skewness, a standardized version of the third central moment. A variant of this takes expectation with respect to the median rather than the mean. Various measures of asymmetry are discussed by MacGillivray (1986). The monograph Dharmadhikari and Joag-dev (1988) does not treat asymmetry directly but is still a very useful reference for its comprehensive study of closely related topics.

We propose some alternatives which measure asymmetry with respect to modes rather than means or medians. The proposed measures always exist and seem quite intuitive. Our approach provides a systematic way to create, from a reference distribution, an entire family of distributions having different asymmetries but sharing some fundamental properties of the original distribution, such as the differential entropy. This may be of some interest for kernel estimation as a source of asymmetric kernels.

2 Notation

We consider absolutely continuous distributions and measure asymmetry by comparing how long does it take for the density to fall to a given value on the two sides of the modes. A very useful tool for this is the confidence transformation (Boshnakov, 2003) which produces, from a given source distribution, a new distribution (called the confidence characteristic) whose density is, effectively, a rearrangement of the values of the source density in decreasing order. This transformation preserves some important properties of the source distribution and so it makes sense to say that distributions having the same confidence density belong to a family, see Boshnakov (2003) for details and related notions. The distribution function and the probability density of the

Accepted manuscript

ACCEPTED MANUSCRIPT

confidence characteristic are called the confidence distribution function and confidence density, respectively.

Given a distribution, we denote its distribution function, probability density, confidence distribution function, and confidence density by F, f, G, and g, respectively.

3 Asymmetry curves and coefficients

Suppose that f is unimodal with mode m. Let, for $l > 0$, $x = x(l)$ and $y = y(l)$ be such that $f(m+x) = f(m-y)$ and $l = x+y$. Then $x = x(l)$ and $y = y(l)$ are monotonically increasing as functions of $l > 0$. If f is strictly monotonically decreasing on each side of the mode then l represents the length of the region where f is greater than $f(m+x)$, while x and y represent the length of the part of this region where f is monotonically descreasing and increasing, respectively. To accomodate multi-modal distributions we use this property to define $x(l)$ and $y(l)$.

For any $c > 0$ let $S_c = \{z : f(z) \geq c\}$. Let also $S_{I,c}$ (respectively, $S_{D,c}$) be the subregion of S_c where the density f is strictly increasing (decresing). Denote the lengths of these regions by l, l_{incr}, and l_{decr}, respectively. We will assume that $l = l_{\text{incr}} + l_{\text{decr}}$.

If f is symmetric then $l_{\text{decr}} = l_{\text{incr}}$ for all l. So, for a symmetric distribution the parametric plot of $l_{\text{decr}} = l_{\text{decr}}(l)$ and $l_{\text{incr}} = l_{\text{incr}}(l)$ as a function of l is a straight line with slope one. For asymmetric distributions such a plot provides comprehensive information about asymmetry. So, we introduce the following definition.

Definition 1. *The curve* $(l_{decr}(l), l_{incr}(l))$, $l \geq 0$, *is said to be the* asymmetry curve *of* f *and* $(l_{decr}(l), l_{incr}(l) - \frac{l}{2})$ *its* detrended asymmetry curve.

The detrended asymmetry curves of symmetric distributions coincide with the positive x-axis. Asymmetry may be defined in terms of functions of $(l_{\text{decr}}(l), l_{\text{incr}}(l))$ as follows.

Definition 2. *The functions* $l_{decr}(l)/l$, $l_{incr}(l)/l$, *and* $l_{decr}(l)/l_{incr}(l)$, *where* $l > 0$, *are called right-asymmetry, left-asymmetry, and odds-asymmetry, respectively.*

Symmetric distributions may be thought of as having a constant asymmetry equal to zero. More generally:

Definition 3. *A distribution is said to have constant asymmetry if its asymmetry curve is a straight line.*

Accepted manuscript

ACCEPTED MANUSCRIPT

For many common distributions the asymmetry curve for small l is close to a straight line with slope one, indicating "approximate" symmetry in a neghbourhood of the mode. Figure 1 shows the asymmetry curves of several Gamma distributions. It makes sense to say that the distribution whose curve is "closest" to the line with slope 1 is the most symmetric one, Gamma(1,120) in this case. This is expected here since increasing the second parameter of the Gamma distribution results in a distribution closer to normal, see also the example in Section 4.3.

Various summary characteristics of the above functions may be considered candidates for the title "coefficient (or index) of asymmetry". Let

$$r_{\text{pos}} = \mathrm{E}_g\left(\frac{l_{\text{decr}}}{l}\right), \qquad r_{\text{neg}} = \mathrm{E}_g\left(\frac{l_{\text{incr}}}{l}\right), \qquad r_{\text{assym}} = r_{\text{pos}} - r_{\text{neg}}, \tag{1}$$

where E_g denotes expectation with respect to the confidence density $g(l)$. We will call r_{assym}, r_{pos}, and r_{neg} the mean asymmetry, the mean positive asymmetry and the mean negative asymmetry, respectively. The expectations above exist since $\frac{l_{\text{decr}}}{l}$ and $\frac{l_{\text{incr}}}{l}$ are positive and less than one. Moreover,

Theorem 1. *The coefficients r_{pos}, r_{neg}, and r_{assym} are always finite, r_{pos} and r_{neg} are in the interval $[0,1]$, r_{assym} is in $[-1,1]$.*

The mean odds-asymmetry is defined by

$$r_{\text{odds}} = \mathrm{E}_g\left(\frac{l_{\text{decr}}(l)}{l_{\text{incr}}(l)}\right), \tag{2}$$

and infinity is a possible value for it.

From the definitions above it is easy to see that symmetric distributions have constant asymmetry.

Theorem 2. *Let f be symmetric with median m, i.e., $f(m+x) = f(m-x)$. Then f has constant asymmetry; the right-asymmetry, left-asymmetry, r_{pos}, and r_{neg} are equal to $\frac{l}{2}$; $r_{odds} = 1$, $r_{assym} = 0$.*

Non-symmetric distributions may also have constant asymmetry. A typical example may happen when the density has several modes and decreases symmetrically around each one of them.

It is clear that the introduced measures of asymmetry are invariant with respect to a shift. Some are invariant with respect to change of scale as well, others are not. Indeed, let $Y = cX$, where $c > 0$ is a positive constant, X is a random variable. We will use the above notation with an additional index

Accepted manuscript

ACCEPTED MANUSCRIPT

x or y for the probability characteristics of X and Y. The two densities are related by

$$f_y(y) = \tfrac{1}{c} f_x(\tfrac{y}{c}).$$

Hence $f_y(a) = f_y(a+l)$ if and only if $f_x(a/c) = f_x((a+l)/c)$. In the particular case when f_x is unimodal this shows that l_x is transformed to cl_y. Similarly, it can be seen that $l_{y,\mathrm{incr}}(l) = cl_{x,\mathrm{incr}}(l/c)$ and $l_{y,\mathrm{decr}}(l) = cl_{x,\mathrm{decr}}(l/c)$. Thus a change of scale leads, in general, to a scale change in the discussed asymmetry measures. It is easy to see however that the change of the asymmetry curve, for example, corresponds to changing the units of its plot. The odds-asymmetry is invariant under a scale transformation. The asumptions of unimodality can be removed with the help of Corollary 1 from Boshnakov (2003).

4 Asymmetry of some distributions

4.1 Symmetric unimodal distributions

Let F be a symmetric unimodal distribution function with mode M. Then its confidence density is $g(l) = f(M + \frac{l}{2}) = f(M - \frac{l}{2})$, i.e., $l_{\mathrm{decr}}(l) = l_{\mathrm{incr}}(l) = \frac{l}{2}$ in this case. So, the asymmetry curve is $(\frac{l}{2}, \frac{l}{2})$, $r_{\mathrm{pos}} = r_{\mathrm{neg}} = \frac{1}{2}$, $r_{\mathrm{assym}} = 0$, $r_{\mathrm{odds}} = 1$.

4.2 Triangular distribution

Let

$$f(x) = \begin{cases} 2x/H & 0 \le x \le H, \\ 2(1-x)/(1-H) & H \le x \le 1, \end{cases}$$

where $H \in (0,1)$. The system $f(H - l_{\mathrm{incr}}) = f(H + l_{\mathrm{decr}})$, $l = l_{\mathrm{decr}} + l_{\mathrm{incr}}$, gives $l_{\mathrm{decr}}(l) = l(1-H)$ and $l_{\mathrm{incr}}(l) = lH$. So, $l_{\mathrm{decr}}(l) = \frac{1-H}{H} l_{\mathrm{incr}}(l)$, $l_{\mathrm{decr}}(l)/l = 1 - H$, $l_{\mathrm{incr}}(l)/l = H$, $l_{\mathrm{decr}}(l)/l - l_{\mathrm{incr}}(l)/l = 1 - 2H$, where $l \in [0,1]$. Hence, the asymmetry of the triangular distribution is constant. The distribution is symmetric if $H = \frac{1}{2}$, skewed to the right if $H < \frac{1}{2}$ and skewed to the left otherwise. The odds-asymmetry is equal to $\frac{1-H}{H}$.

4.3 Γ-distribution

Let f be a Γ-density (Johnson et al., 1994),

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \qquad x \ge 0. \tag{3}$$

Accepted manuscript

ACCEPTED MANUSCRIPT

We assume here that $\alpha > 1$. In this case the distribution is unimodal with mode $M = (\alpha - 1)/\lambda$ such that

$$f(M) > f(x), \quad \text{for every } x \neq M. \tag{4}$$

The equation $f(M + l_{\text{decr}}) = f(M - l_{\text{incr}})$, where $l_{\text{decr}} \geq 0$ and $l_{\text{incr}} \geq 0$, will be satisfied if

$$(M + l_{\text{decr}})^{\alpha-1} e^{-\lambda(M+l_{\text{decr}})} = (M - l_{\text{incr}})^{\alpha-1} e^{-\lambda(M-l_{\text{incr}})},$$

which can be written as

$$\left(\frac{M + l_{\text{decr}}}{M - l_{\text{incr}}}\right)^{\alpha-1} = e^{\lambda(M+l_{\text{decr}}-M+l_{\text{incr}})} = e^{\lambda l}. \tag{5}$$

Also,

$$\frac{M + l_{\text{decr}}}{M - l_{\text{incr}}} = \frac{(M - l_{\text{incr}}) + (l_{\text{incr}} + l_{\text{decr}})}{M - l_{\text{incr}}} = 1 + \frac{l}{M - l_{\text{incr}}}. \tag{6}$$

From equations (5) and (6) we get

$$1 + \frac{l}{M - l_{\text{incr}}} = e^{\lambda l/(\alpha-1)}.$$

Hence,

$$\frac{l}{M - l_{\text{incr}}} = e^{\lambda l/(\alpha-1)} - 1.$$

So,

$$M - l_{\text{incr}} = \frac{l}{e^{\lambda l/(\alpha-1)} - 1}.$$

Finally, using the identity $l_{\text{decr}} + l_{\text{incr}} = l$, we get

$$l_{\text{incr}} = M - \frac{l}{e^{\lambda l/(\alpha-1)} - 1}$$
$$l_{\text{decr}} = l - M + \frac{l}{e^{\lambda l/(\alpha-1)} - 1}$$

Hence, when $\alpha > 1$ the confidence density of the Γ-distribution is

$$\begin{aligned} g(l) &= f(M + l_{\text{decr}}), \qquad l \geq 0 \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(l + \frac{l}{e^{\lambda l/(\alpha-1)} - 1}\right)^{\alpha-1} e^{-\lambda\left(l + \frac{l}{e^{\lambda l/(\alpha-1)} - 1}\right)} \end{aligned}$$

Accepted manuscript

ACCEPTED MANUSCRIPT

From the way we defined l_{decr} and l_{incr}, they should have the following limiting behavior

$$l_{\text{decr}} \xrightarrow[l\to\infty]{} \infty \qquad\qquad l_{\text{decr}} \xrightarrow[l\to 0]{} 0$$
$$l_{\text{incr}} \xrightarrow[l\to\infty]{} M \qquad\qquad l_{\text{incr}} \xrightarrow[l\to 0]{} 0$$

This is indeed so, since it is easy to verify that for any $c > 0$

$$\frac{l}{e^{cl}-1} \xrightarrow[l\to\infty]{} 0 \qquad \text{and} \qquad \frac{l}{e^{cl}-1} \xrightarrow[l\to 0]{} \frac{1}{c}.$$

We also have $g(0) = f(M)$, as expected. Thus, the odds-asymmetry tends to infinity as $l \to \infty$, the right-asymmetry and the asymmetry tend to 1.

We see that the asymmetry curve of the Γ-distribution depends only on $M = (\alpha - 1)/\lambda$. In other words, for Γ-distributions having the same mode the asymmetry curves and all measures of asymmetry derived from it are the same. For comparison, the usual coefficient of skewness is equal to $2/\sqrt{\alpha}$ and so depends on α but not on λ.

5 Distributions with given asymmetry

The confidence transformation has a number of desirable properties. In particular, it preserves the entropy and other information theoretic properties of the original distribution, see Boshnakov (2003) for details. It is therefore justifiable to classify distributions by their confidence characteristics. By reverting the above process distributions with specified asymmetry properties may be generated.

5.1 Constant asymmetry

Suppose that g is the confidence density of some unimodal distribution and we wish to create a distribution with the same confidence characteristic but with odds-asymmetry $c > 0$. Using the established notation, the following relations should be satisfied:

$$\frac{l_{\text{decr}}}{l_{\text{incr}}} = c, \quad l_{\text{decr}} + l_{\text{incr}} = l, f(M + l_{\text{decr}}) = f(M - l_{\text{incr}}) = g(l). \tag{7}$$

Let $x \geq 0$, $y \geq 0$. The required density is

$$f(M + x) = g(\tfrac{1+c}{c}x)$$
$$f(M - y) = g((1 + c)y).$$

Accepted manuscript

ACCEPTED MANUSCRIPT

5.2 An asymmetric normal family

The confidence distribution function and confidence density of the standard normal distribution are $G(l) = 2\Phi(l/2) - 1$ and $g(l/2) = \varphi(l/2) = \exp(-l^2/8)/\sqrt{2\pi}$. The above formulae then give

$$f(M+x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1+c}{c}\right)^2 x^2/8},$$
$$f(M-y) = \frac{1}{\sqrt{2\pi}} e^{-(1+c)^2 y^2/8}.$$

The distributions obtained by varying c and M form a family of distributions having constant asymmetry and the same confidence characteristic as the standard normal distribution.

5.3 Another family

It may be more convenient in some circumstances to express l and l_{incr} in terms of l_{decr}. So, let

$$l_{\text{incr}} = u(l_{\text{decr}}), \quad l = l_{\text{decr}} + u(l_{\text{decr}}), \tag{8}$$

where $u(.)$ is an appropriate function. Then we may define a new density f by

$$f(M + l_{\text{decr}}) = f(M - l_{\text{incr}}) = g(l) = g(l_{\text{decr}} + u(l_{\text{decr}})). \tag{9}$$

For example if we take $g(z) = \lambda e^{-\lambda z}$ to be the the exponential density and $u(x) = x^2$, then we get

$$f(M + l_{\text{decr}}) = f(M - l_{\text{decr}}^2) = g(l_{\text{decr}} + u(l_{\text{decr}})) = \lambda e^{-\lambda(l_{\text{decr}} + l_{\text{decr}}^2)}.$$

This can be written as

$$f(M + l_{\text{decr}}) = \begin{cases} g(l_{\text{decr}} + l_{\text{decr}}^2) = \lambda e^{-\lambda(l_{\text{decr}} + l_{\text{decr}}^2)}, & \text{if } l_{\text{decr}} \geq 0 \\ g(|l_{\text{decr}}| + \sqrt{|l_{\text{decr}}|}) = \lambda e^{-\lambda(|l_{\text{decr}}| + \sqrt{|l_{\text{decr}}|})}, & \text{if } l_{\text{decr}} < 0. \end{cases}$$

Now the asymmetry is not constant. For example, the odds-asymmetry is $l_{\text{decr}}/l_{\text{decr}}^2 = 1/l_{\text{decr}}$.

6 Conclusion

We defined some measures of asymmetry which provide useful information about this type of property. The asymmetry curve and its variants provide comprehensive information, while their averaged counterparts summarize it to single numbers. These measures provide a systematic way to generate asymmetric distributions.

Accepted manuscript

ACCEPTED MANUSCRIPT

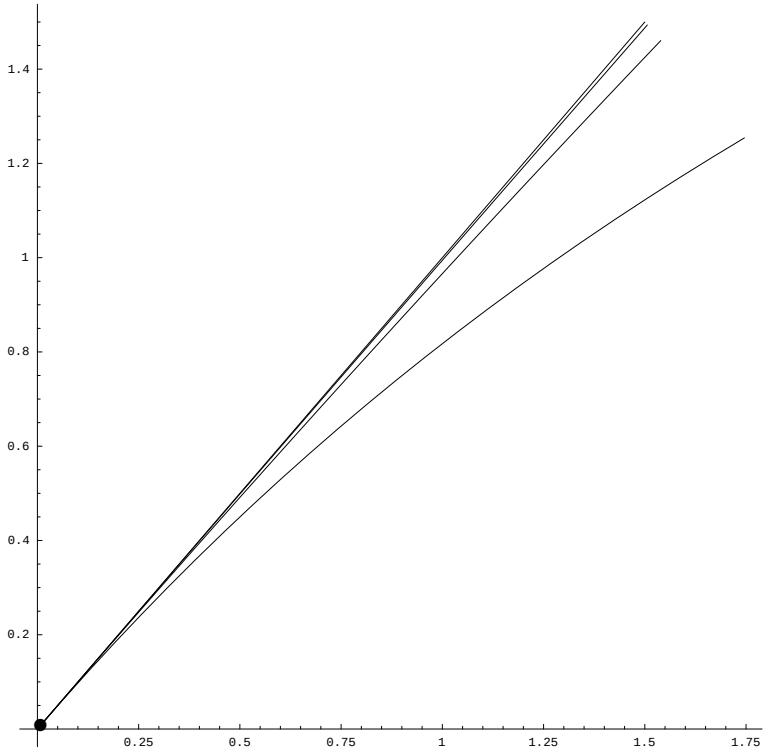


Figure 1: Asymmetry curves, $(l_{\text{decr}}(l), l_{\text{incr}}(l))$, $l > 0$, of Gamma(1,4), Gamma(1,20) and Gamma(1,120). When the second parameter is increased, the curves are closer to the straight line with slope 1 (almost the same in the case of Gamma(1,120) over the plotted range), the interpretation being that the corresponding distributions become more symmetric. All pictured curves however have horizontal asymptotes because the left limit of the support of the Gamma distribution is finite.

References

Boshnakov, G. N. (2003) *Confidence characteristics of distributions*, Statistics & Probability Letters **63/4**, 353–360.

Dharmadhikari, S. and K. Joag-dev (1988) *Unimodality, convexity and applications*, Probability and Mathematical Statistics. London: Academic Press.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994) *Continuous univariate distributions. Vol. 1.*, 2nd ed., Chichester: Wiley.

MacGillivray, H. L. (1986) *Skewness and asymmetry: Measures and orderings.*, Ann. Stat. **14**, 994–1011.

Accepted manuscript