

AUTHOR QUERY FORM

	<p>Journal: AMC</p> <p>Article Number: 19552</p>	<p>Please e-mail or fax your responses and any corrections to:</p> <p>E-mail: corrections.esch@elsevier.sps.co.in</p> <p>Fax: +31 2048 52799</p>
---	--	--

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult <http://www.elsevier.com/artworkinstructions>.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Please confirm that given name(s) and surname(s) have been identified correctly.
Q2	The number of keywords provided exceeds the maximum allowed by this journal. Please delete one keyword.
Q3	Please note that the reference style has been changed from Name–Date style to Numbered style as per the journal specifications.
Q4	Please check the word “simplectic” has been changed to “symplectic”. Please check and correct if necessary.
Q5	Please provide an update for Ref. [7].

Please check this box if you have no corrections to make to the PDF file

Thank you for your assistance.



Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc



Kullback–Leibler life time testing

M. Stehlík^{a,b,*}, P. Economou^c, J. Kiseřák^d, W.-D. Richter^e

^a Department of Applied Statistics, Johannes Kepler University in Linz, Austria

^b Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile

^c Department of Civil Engineering, University of Patras, Greece

^d Institute of Mathematics, Faculty of Science, P.J. Šafárik University in Košice, Slovakia

^e Institute of Mathematics, University of Rostock, Rostock, Germany

ARTICLE INFO

Keywords:

Information divergence
Deconvolution
Statistical inference
Homogeneity and scale testing
Exact distribution
Lambert W
Intersection percentage function

ABSTRACT

The paper deals with testing the hypotheses for homogeneity and point null value of the scale parameter in the gamma family. Tests suggested here are based upon the Kullback–Leibler divergence from an observed vector to the canonical parameter (see Pázman, 1993 [14]), and upon its decomposition. The latter is used to derive the exact distribution of the test statistic by convolution. A geometric integration method is used alternatively to derive the distribution directly. Because it is observed by simulation, that the test's performance is poor when the shape parameter is estimated from the data, an interval method is developed and its use is demonstrated in an analysis of real data.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

It is natural to hope that by adopting optimal statistical procedures one is using somehow natural forms of information merging or decomposition. In this regard, we recall recent development in the deconvolution of Kullback–Leibler information divergence (I – divergence for short) and its relation to optimal statistical testing. For a book review on statistical inference based on standard divergence measures the reader is referred to [13].

In this paper, we concentrate on the I – divergence statistic $I_N(y, \gamma)$ from an observed vector y to the canonical parameter γ (see [14]) of the gamma family. A test based on $I_N(y, \gamma)$ can be well utilized for joint by testing the homogeneity and point null value of the scale parameter in the gamma family. We illustrate the applicability of such a test for several real data sets, among them one on light indicators for aeroplanes.

Information divergence can be decomposed in special cases (as in the gamma case), to construct a natural measure of heterogeneity or a test separately for homogeneity and a point null value hypothesis with respect to (w.r.t.) the scale parameter. Such a measure of heterogeneity is useful not only for testing, but also for clustering, since there the application of automatic methods hoping that data will disclose the true structure is deceptive (see e.g. discussion [25]). More specifically, the exact distribution of the I – divergence statistic can be derived by a convolution of the distributions of two independent random variables R_N, S_N corresponding to the likelihood ratio statistics for testing the hypotheses of homogeneity and point null value of the scale. We show the complexities of the exact distribution of the I – divergence statistic, and derive

* Corresponding author. Address: Institut für Angewandte Statistik, Altenberger Straße 69, A-4040 Linz, Austria.

E-mail addresses: Milan.Stehlik@jku.at (M. Stehlík), peconom@upatras.gr (P. Economou), jozef.kiselak@gmail.com (J. Kiseřák), wolf-dieter.richter@uni-rostock.de (W.-D. Richter).

<http://dx.doi.org/10.1016/j.amc.2014.04.027>

0096-3003/© 2014 Elsevier Inc. All rights reserved.

approximations of distributions for higher sample sizes. Such approximations are working well, and jointly with simulations they help to determine critical test values.

The paper is organized as follows. In Section 1 we recall the I – divergence deconvolution result in [22]. Hereafter, we study in Section 2 deconvolution of the exact distributions of the likelihood ratio test statistics in the gamma and normal families. According to these results, we introduce a test based on $I_N(\mathbf{y}, \gamma)$ and motivate it by a real data example on airplane indicator light operating times. In Section 3, the exact distributions of the I – divergence statistic are derived for small samples by means of convolutions and geometric integration theory developed in [8]. In Section 4, the critical values of the I – divergence statistic are computed for various samples sizes and shape parameters, based on a detailed simulation study, and a certain difficulty of applying the proposed test is discussed when the shape parameter is unknown and has to be estimated. In Section 5, we apply the proposed test to real data and construct credible regions for the gamma parameters. Section 6 concludes.

2. Deconvolution of the I – divergence statistic

Let y_1, y_2, \dots, y_N , be N independent, but not necessary identically distributed observations according to the gamma probability density function

$$f(y_i|\gamma_i) = \begin{cases} \gamma_i^{v_i} \frac{y_i^{v_i-1}}{\Gamma(v_i)} \exp(-\gamma_i y_i), & y_i > 0, \\ 0, & y_i \leq 0. \end{cases} \quad (1)$$

Here $\gamma := (\gamma_1, \dots, \gamma_N)$ is the vector of the unknown scale parameters, which are the parameters of interest, and $v = (v_1, \dots, v_N)$ is the vector of the known shape parameters.

This structure is motivated, for example, by a situation when we observe time intervals between $(N + 1)$ successive random events in a Poisson process. In this case, the parameters γ_i are equal to the (usually parametrized) intensity γ , and the shape parameters are all equal to 1.

The densities in (1) build a regular exponential family (see [1]) and thus the sufficient statistic for the canonical parameter $\gamma \in \Gamma$ has the form $t(\mathbf{y}) = -\mathbf{y}$ where $\Gamma = \{(\gamma_1, \dots, \gamma_N), \gamma_i > 0; i = 1, \dots, N\}$. The “covering” property $\{t(\mathbf{y}) : \mathbf{y} \in Y\} \subseteq \{E_\gamma[t(\mathbf{y})] : \gamma \in \Gamma\}$ (see [14]) together with the relation $E_\gamma[t(\mathbf{y})] = \frac{\partial \kappa(\gamma)}{\partial \gamma}$, where $\kappa(\gamma) = N \ln(\Gamma(v)) - v \sum_{i=1}^N \ln(\gamma_i)$, makes it possible to associate with each value of $t(\mathbf{y})$ a value $\hat{\gamma}_y \in \Gamma$ which satisfies

$$\left. \frac{\partial \kappa(\gamma)}{\partial \gamma} \right|_{\gamma=\hat{\gamma}_y} = t(\mathbf{y}). \quad (2)$$

It follows from (2) that $\hat{\gamma}_y$ is the MLE of the canonical parameter γ in the family (1). According to (2), we can define the I – divergence of the observed vector \mathbf{y} from γ , in the sense of [14], as $I_N(\mathbf{y}, \gamma) := I(\hat{\gamma}_y, \gamma)$, thus obtaining

$$I_N(\mathbf{y}, \gamma) = -\sum_{i=1}^N \{v_i - v_i \ln(v_i)\} + \sum_{i=1}^N \{y_i \gamma_i - v_i \ln(y_i \gamma_i)\}, \quad (3)$$

which can be simplified to

$$I_N(\mathbf{y}, \gamma) = -N(v - v \ln(v)) + \sum_{i=1}^N \{y_i \gamma_i - v \ln(y_i \gamma_i)\} \quad (4)$$

under the assumption that all the observations share the same shape parameter v , $v_i = v \forall i$. Let the **Kullback–Leibler** divergence between the distributions P_{γ^*} and P_γ be defined by

$$I(\gamma^*, \gamma) := \begin{cases} \int \ln \frac{dP_{\gamma^*}}{dP_\gamma} dP_{\gamma^*}, & \text{if } P_{\gamma^*} \ll P_\gamma, \\ +\infty, & \text{otherwise.} \end{cases}$$

The main advantage of the I – divergence is that it can be used to test the homogeneity hypothesis

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_N \text{ versus } H_A : \text{non } H_0 \quad (5)$$

and the simple scale hypothesis

$$H_0 : \gamma = \gamma_0 \text{ versus } H_A : \gamma \neq \gamma_0. \quad (6)$$

This can be justified by the statistical decomposition of the I – divergence. Let us denote the likelihood ratio (LR) of the test for simple hypothesis $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \neq \gamma_0$ by λ_1 and the LR of the test for homogeneity $H_0 : \gamma_1 = \dots = \gamma_N$ in the family (1) by λ_2 . Then the following decomposition for every vector of canonical parameters $(\gamma_0, \dots, \gamma_0) \in \Gamma^N$ holds:

$$I_N(\mathbf{y}, (\gamma_0, \dots, \gamma_0)) = -\ln \lambda_1 + (-\ln \lambda_2 | \gamma_1 = \dots = \gamma_N). \quad (7)$$

Here, the variables $-\ln \lambda_1$ and $-\ln \lambda_2 | \gamma_1 = \dots = \gamma_N$, (i.e. $-\ln \lambda_2$ under the condition $H_0 : \gamma_1 = \dots = \gamma_N$) are independent. The deconvolution (7) of I_N is the consequence of Theorem 4 in [22]. Both tests are asymptotically optimal in the Bahadur sense ([19,20]).

The above mentioned structural relationship of likelihood ratio test statistics may be useful in more complex setups, e.g. when a nuisance parameter elimination for proportional likelihood ratio models is needed (see e.g. [3]). In the present case of I -divergence in the gamma family, we have discussed above a decomposition of the I -divergence from an observed vector to the canonical parameter. In [9], a generalized family of measures of divergence are investigated and applied successfully to statistical inference. Deconvolution ideas similar to that stressed here, will be of further interest for such families of divergences, too. For more open problems, e.g. those related to ϕ -divergences and statistical information, see [24].

The above discussed case of gamma distributions extends ANOVA to that case of distributions. ANOVA, in the case of normal observations, is an archetypical example for testing homogeneity of parameters. In Section 3.5 the I -divergence in the normal regression is presented, showing that this pseudo-distance became squared Euclidean distance.

2.1. Motivation: life testing of light indicators

In practical cases it may be observed that practitioners (e.g. reliability engineers) are testing of a particular life time (i.e. testing a scale hypothesis) while silently ignoring the uncertainty about homogeneity in the sample. As an example of such a situation consider the data presented in Table B.1 regarding the cumulative operating times of aeroplane indicator light during successive failures taken from RAC database (see [4]).

It is true that the gamma distribution is a natural generalization of the instinctive choice of the exponential distribution for modeling these data. The authors in [4] give the MLEs of the gamma parameters under the assumption of the gamma distributed individual times-to-failure (shape parameter = 0.7 and scale parameter = 0.0000484). If the homogeneity of the scale parameters is indeed statistically significant, we could use a directed test for scale parameter (with simple null, and composite alternative hypothesis, typically). For example, we can rely on the exact LR test for the scale parameter using the sum of the shape parameters $\omega = 38 \times 0.7 = 26.6$ and the total at risk $\sum_{j=1}^6 T_j = 552, 400$. Although, the homogeneity hypothesis is at least questionable, since not only the light indicators may have different characteristics, but also the successive failures may cause correlation resulting to changes of the scale parameter.

One remedy in such a situation is to test for both homogeneity and the scale hypotheses in single step based on I -divergence $I_N(y, \gamma_0)$, i.e. to statistically measure the deviation of the observed vector y from the hypothesized canonical parameter γ_0 . This is illustrated in the next sections, and especially in Section 5. Stehlík et al. [26] illustrated some preliminary simulation results in this direction.

3. Exact distribution of $I_N(y, \gamma)$

Stehlík [22] derived (in Theorems 1 and 2) the exact distribution of $I_N(y, \delta(1, 1, \dots, 1))$ as the distribution of the sum of two independent variables R_N, S_N , where y is a sample of size N from Exp (1), $\delta > 0$ is the perturbation parameter given by the ratio of the unknown true scale parameter and the hypothesized scale parameter, and $(1, 1, \dots, 1)$ denotes the N -dimensional vector with all entries equal to 1.

The c.d.f. of the random variable R_N has the form

$$F_N(\rho) = \begin{cases} \mathcal{F}_N(-\frac{N}{\delta} W_{-1}(-\exp(-1 - \frac{\rho}{N}))) - \mathcal{F}_N(-\frac{N}{\delta} W_0(-\exp(-1 - \frac{\rho}{N}))), & \rho > 0, \\ 0, & \rho \leq 0 \end{cases} \quad (8)$$

and the density of R_N has the form

$$f_N(\rho) = \begin{cases} h(N, 1, \rho, \delta^{-1}) - h(N, 0, \rho, \delta^{-1}), & \text{for } \rho > 0, \\ 0, & \text{for } \rho \leq 0. \end{cases}$$

Here \mathcal{F}_N is the c.d.f. of the $\Gamma(N, 1)$ -distribution, and for $\tau, r, s > 0; k \in \mathbf{Z}$ we define

$$h(N, k, r, s) = \frac{(-N)^{N-1} s^N}{\Gamma(N)} \frac{\{W_{-k}(-\exp(-1 - \frac{r}{N}))\}^N}{1 + W_{-k}(-\exp(-1 - \frac{r}{N}))} \times \exp\{NsW_{-k}(\exp(-1 - \frac{r}{N}))\},$$

where W_0, W_{-1} are the two real-valued branches of the Lambert-W function (see [5]).

Under the null hypothesis of homogeneity, the cdf of S_2 has the form (see [22]):

Table B.1
Aeroplane indicator light reliability data.

Number of failures	2	9	8	8	6	5
Cumulative operating time T_j	51,000	194,900	45,300	112,400	104,000	44,800

$$G_2(x) = \begin{cases} \sqrt{1 - \exp(-x)}, & \text{for } x > 0, \\ 0, & \text{for } x \leq 0. \end{cases}$$

161

162 **Q4** The density of S_3 is given in [22], and for larger N , we use nonlinear function for symplectic component S_N to obtain it
163 indirectly. The density of the statistic $I_N(y, \delta(1, 1, \dots, 1)) = R_N + S_N$ is given by convolution of the densities of the two sum-
164 mands and will be presented in the following subsections.

165 One can generally work with R_N and S_N in a separate manner, not only because one can determine their null distributions,
166 but also because R_N and S_N correspond to the likelihood ratio test statistics for homogeneity (hypothesis (5)) and simple scale
167 hypothesis (hypothesis (6)), respectively. Of course, one can also work simultaneously with both statistics, by studying
168 directly the I – divergence, as we will do in the following sections.

169 To our best knowledge, the exact likelihood ratio test for the scale and homogeneity in the complete sample from gamma
170 family has been derived in [22]. Other studies of these tests for several setups has been made in [23] and references therein.
171 These tests have been shown to be optimal in the sense of Bahadur (see [19,20,22]). The exact LR test of scale follows asymp-
172 totically a χ^2 distribution (see [27]).

173 In Sections 3.1–3.3, we derive the exact distribution of the I – divergence $I_N(y, \delta(1, 1, \dots, 1))$ by means of convolutions of
174 densities of random variables R_N and S_N for $N = 2, 3, 4$, respectively. However, for $N > 4$, the exact distribution becomes non-
175 tractable, and thus we approximate the convolution (Section 3.4) by exploiting the approximation for S_N given by Bartlett
176 and Kendall [2].

177 Validation of theoretical results is made by simulation of critical constants. Additionally, in Section 3.6, we develop a dif-
178 ferent approach based on geometrical integration. The I – divergence in the normal regression is given in Section 3.5.

179 **3.1. $N = 2$**

180 As already mentioned, the density z_2 of the sum $R_N + S_N$, ($N = 2$) is a convolution of the densities f_2 and g_2 and is given by

$$z_2(u) = \int_{-\infty}^{\infty} f_2(\rho) g_2(u - \rho) d\rho,$$

183

184 where g_2 is density of S_2 and f_2 is the density of R_2 obtained by the c.d.f in (8). But then, the c.d.f of $R_2 + S_2$ is

185

$$Z_2(s) = \int_{-\infty}^s z_2(u) du = \int_{-\infty}^s \int_{-\infty}^{\infty} f_2(\rho) g_2(u - \rho) d\rho du$$

187

188 and using the Fubini theorem, we conclude that it has the form

189

$$Z_2(u) = \int_{-\infty}^{\infty} f_2(\rho) G_2(u - \rho) d\rho.$$

191

192 In the upper half of Table B.2, the critical values of $S_2 + R_2$ are shown. These values were computed using two different
193 methods. Firstly, we relied on the convolution type expression of the distribution of the I – divergence (for both exact and
194 approximated distribution of S_2). We had to rely on numerical integration since the direct computation of certain involved
195 integrals was not possible. More specifically, we have used the function `NIntegrate()` of software Mathematica to numer-
196 ically evaluate the convolution integral, by sampling the integrand at a sequence of 10,000 points (`MaxPoints = 10,000`) of Z_n
197 and then using numerical integration on real axis. Additionally to that approach, 100,000 (`nsim = 100,000`) random samples
198 were generated and the $1 - \alpha$ empirical quantiles of I_2 were computed to obtain c_z . It is clear that the two procedures give
199 similar critical values although their difference is more severe as we move to more extreme α demonstrating in that way the
200 weakness of the numerical integration approach.

201 **3.2. $N = 3$**

202 In order to obtain the distribution of the I – divergence statistic, we first prove the c.d.f. of S_3 to have the form

203

Table B.2
Critical values C_α for $S_2 + R_2$ (upper table) and $S_3 + R_3$ (lower table) ($\sigma = 1, \nu = 1, \gamma = 1$).

$S_2 + R_2$	$C_{0.1}$	$C_{0.05}$	$C_{0.01}$	$C_{0.001}$
Convolution (<code>MaxPoints = 10⁴</code>)	2.649	3.426	5.204	7.699
Convolution (approx. S_2 , <code>MaxPoints = 10⁴</code>)	2.684	3.493	5.371	8.074
I – divergence (<code>nsim = 10⁵</code>)	2.651643	3.414478	5.239796	7.592669
$S_3 + R_3$	$C_{0.1}$	$C_{0.05}$	$C_{0.01}$	$C_{0.001}$
Convolution (<code>MaxPoints = 10³</code>)	3.59	4.47	6.42	9.09
Convolution (approx. S_3 , <code>MaxPoints = 10³</code>)	3.64	4.56	6.63	9.52
I – divergence (<code>nsim = 10⁵</code>)	3.581297	4.443605	6.453668	9.041494

$$G_3(x) = \begin{cases} 2 \int_{a(x)}^{b(x)} \frac{1}{s} \sqrt{s^2(1-s)^2 - s \frac{4}{27} \exp(-x)} ds, & \text{for } x > 0, \\ 0, & \text{for } x \leq 0 \end{cases}$$

where $0 < a(x) < b(x) < 1$ are solutions of the algebraic equation $t(1-t)^2 = \frac{4}{27} e^{-x}$. To be more specific, we obtain three roots

$$t_k(x) = \frac{2}{3} + \frac{2}{3} \cos\left(\frac{2k\pi}{3} - \frac{1}{3} \arccos(2e^{-x} - 1)\right), \quad k = 0, 1, 2,$$

where $a(x) = t_2(x)$, $b(x) = t_1(x)$ and for the third root, $c(x) = t_0(x)$, $c(x) > 1$ for $x > 0$. Notice that Cardano's method leads to a complex form of the roots, thus the trigonometric method should be used.

In order to obtain the c.d.f. of S_3 , we need to calculate the complete elliptic integrals of the first, second and third kinds given respectively by

$$K(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k^2 \sin^2 \theta}} = \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k^2 t^2)}},$$

$$E(k) = \int_0^{\pi/2} \sqrt{1-k^2 \sin^2 \theta} d\theta = \int_0^1 \frac{\sqrt{1-k^2 t^2}}{\sqrt{1-t^2}} dt$$

$$\text{and } \Pi(n, k) = \int_0^{\pi/2} \frac{d\theta}{(1-n \sin^2 \theta) \sqrt{1-k^2 \sin^2 \theta}}.$$

Notice that the total circumference of an ellipse can be given in terms of complete elliptic integrals. In the view of these integrals, the c.d.f. of S_3 can be expressed as follows,

$$G_3(x) = \begin{cases} \frac{K(m(x))k(x)+E(m(x))e(x)+\Pi\left(1-\frac{a(x)}{b(x)}, m(x)\right)\pi(x)}{2\sqrt{b(x)(c(x)-a(x))}}, & \text{for } x > 0, \\ 0, & \text{for } x \leq 0, \end{cases}$$

where $m(x) = \sqrt{\frac{c(x)(b(x)-a(x))}{b(x)(c(x)-a(x))}}$, $k(x) = a(x)b(x)^2 + a(x)^2b(x) - 5a(x)b(x)c(x)$, $e(x) = b(x)c(x)^2 + b(x)^2c(x) - a(x)b(x)^2 - a(x)^2b(x)$ and $\pi(x) = 2a(x)^2b(x) + 2a(x)^2c(x) + 2a(x)b(x)c(x) - a(x)c(x)^2 - a(x)b(x)^2 - a(x)^3$. Using same arguments as for $N = 2$, we conclude that c.d.f. of $S_3 + R_3$ has the form

$$Z_3(u) = \int_{-\infty}^{\infty} f_3(\rho) G_3(u - \rho) d\rho.$$

In the lower half of Table B.2, the critical values are presented which were obtained with the two methods described in the two dimensional case. The only difference is that we used 1000 points for the numerical integration. It is worth mentioning that the duration of this approach was 35 min. From the results it is clear that the critical values have similar or smaller differences compared with the corresponding values of the two dimensional case. For graphical comparison of c.d.fs. of $S_2 + R_2$ and $S_3 + R_3$, see Fig. B.1.

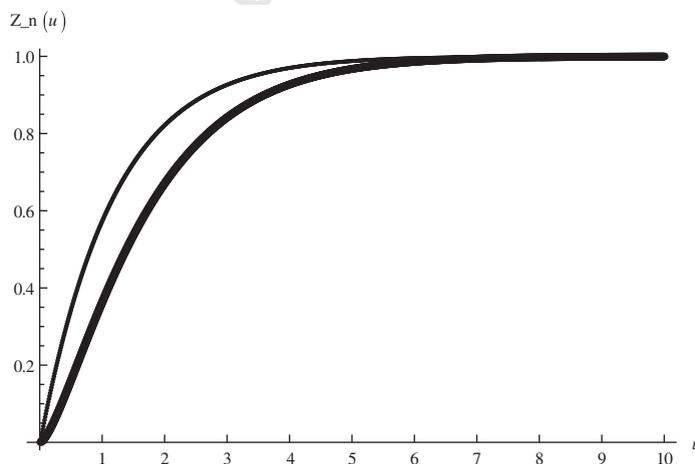


Fig. B.1. Plot of the distribution functions of $S_2 + R_2$ and $S_3 + R_3$ (thick line).

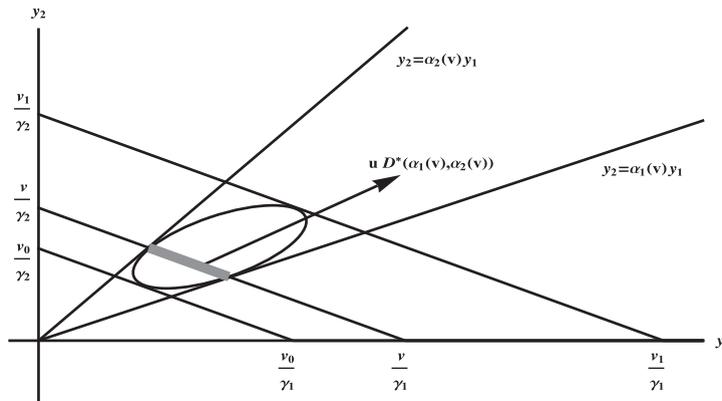


Fig. B.2. The geometric definition of $D(\mu_1, \mu_2)$ and $\text{Sector}_\gamma(D(\mu_1, \mu_2), v)$.

3.3. $N = 4$

The c.d.f. of S_4 has form

$$G_4(x) = \frac{2}{3!} \int_{a(x)}^{b(x)} \int_{s_2(x,t)}^{s_1(x,t)} \frac{1}{st} \sqrt{s^2 t^2 (s+t-1)^2 - st} \frac{e^{-x}}{64} ds dt$$

for $x > 0$, and is 0 for $x \leq 0$, where $0 < a(x) < b(x) < 1$ are solutions of the algebraic equation $-\frac{27}{128} e^{-x} = 2t(t-1)^3$ (result of the domain of arccos in roots T_k below – see Appendix A). Moreover,

$$T_k(x, t) = \frac{2}{3} (1-t) \left(1 + \cos \left(\frac{1}{3} \arccos \left(\frac{-\frac{27}{128} \frac{e^{-x}}{t} - (t-1)^3}{(t-1)^3} \right) - \frac{2k\pi}{3} \right) \right), \quad k = 0, 1, 2$$

are the solutions of the algebraic equation $64st(s+t-1)^2 = e^{-x}$ in s . They satisfy the relationship $0 < s_2(x, t) < s_1(x, t) < 1 < s_0(x, t)$ for $t \in (0, 1)$, $x \in (0, \infty)$. Using elliptic integrals, we can obtain the c.d.f. $G_4(x)$ in the form

$$\frac{1}{12} \int_{a(x)}^{b(x)} \frac{K(m_x(t)) k_x(t) + E(m_x(t)) e_x(t) + \Pi \left(1 - \frac{s_2(x,t)}{s_1(x,t)}, m_x(t) \right) \pi_x(t)}{t \sqrt{s_1(x,t)(s_0(x,t) - s_2(x,t))}} dt$$

for $x > 0$, and being 0 otherwise. Here $m_x(t) = \sqrt{\frac{s_0(x,t)(s_1(x,t) - s_2(x,t))}{s_1(x,t)(s_0(x,t) - s_2(x,t))}}$, $k_x(t) = s_2(x, t)s_1(x, t)^2 + s_2(x, t)^2 s_1(x, t) - 5s_2(x, t)s_1(x, t)s_0(x, t)$, $e_x(t) = s_1(x, t)s_0(x, t)^2 + s_1(x, t)^2 s_0(x, t) - s_2(x, t)s_1(x, t)^2 - s_2(x, t)^2 s_1(x, t)$ and $\pi_x(t) = 2s_2(x, t)^2 s_1(x, t) + 2s_2(x, t)s_0(x, t) + 2s_2(x, t)s_1(x, t)s_0(x, t) - s_2(x, t)s_0(x, t)^2 - s_2(x, t)s_1(x, t)^2 - s_2(x, t)^3$.

So, the c.d.f. of $S_4 + R_4$ is given by $Z_4(u) = \int_{-\infty}^{\infty} f_4(\rho) G_4(u - \rho) d\rho$.

3.4. $N > 4$

As we have seen in the previous sections, complexity of the exact distribution increases tremendously. Therefore, for $N > 4$, we prefer the following approximation of the distribution of S_N given by Bartlett and Kendall [2]:

$$\frac{1}{2} \left(1 + \frac{1 + \frac{1}{N}}{6} \right) \chi_{N-1}^2.$$

By letting $c := \frac{1}{2} \left(1 + \frac{1 + \frac{1}{N}}{6} \right)$, the c.d.f. of the approximated S_N is expressed as

$$\tilde{G}_N(x) = \frac{\gamma \left(\frac{N-1}{2}, \frac{x}{2c} \right)}{\Gamma \left(\frac{N-1}{2} \right)},$$

where $\gamma(a, x) := \int_0^x t^{a-1} e^{-t} dt$ denotes the lower incomplete gamma function. Then the cdf of $R_N + S_N$ can be expressed as the convolution $f_N * \tilde{G}_N$. From Tables B.2, B.3, it is clear that the quality of the approximation increases, as N increases and that the approximation oversized the value of the critical constants.

3.5. Normal distribution

There is an analogy between the radial component of the I – divergence in the normal regression and in the model (3) (for more details see [22]). For the normal regression with $y \sim \mathcal{N}_N(\vartheta, \sigma^2 I)$, the I – divergence is given by $I_N(t(y), \vartheta) = \frac{1}{2\sigma^2} \|y - \vartheta\|^2$ where $\vartheta \in \Theta$ is an unknown parameter of the interest and σ being the known variance parameter. Fig. B.2.

Thus the LR test of the hypothesis $H_0 : \vartheta = \vartheta_0$ versus $H_1 : \vartheta \neq \vartheta_0$, based on the statistic $-2 \ln \lambda = 2R_N^*(r, \vartheta_0)$, has a χ_N^2 distribution under the null hypothesis.

For a comparison with the exponential model, consider the LR test of the hypothesis $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma \neq \gamma_0$ in the simple generalized linear model $y \sim \exp(\gamma, \dots, \gamma)$ (homogeneous Poisson process). Under the null hypothesis, the LR statistics $-2 \ln \lambda = 2R_N(r, \gamma_0)$ is asymptotically χ_1^2 -distributed. More extensive discussion about this test can be found in Section 5 of [22].

3.6. Geometric integration method to obtain exact distribution

In this section, we apply alternatively a geometric integration method developed in [8] to derive the exact distribution of I_2 . Such a construction aims to illustrate the possible approach to the derivation of the exact distribution and can be generalized for arbitrary dimension N as was demonstrated in [15,17] along with several advanced applications and a generalization of this method. However, we concentrate here after some general results only on the simplest case, $N = 2$, to avoid unnecessary prolongation of the paper.

Let us consider the vector

$$Y = (Y_1, \dots, Y_n)^T \sim (E(\gamma_1), \dots, E(\gamma_n))^T = E_\gamma$$

with joint exponential density $f_\gamma(y) = \prod_{i=1}^n \gamma_i e^{-\gamma_i y_i}$, $y \in \mathbb{R}_+^n$. In order to obtain the c.d.f. of the I -divergence, we need first to define some necessary functions. More specifically, we define a modified l_1 -norm (see [8]): $|y|_\gamma = \sum_{i=1}^n \gamma_i y_i$, and a simplicial radius statistic $R_\gamma := |Y|_\gamma$, which is nonnegative, Gamma(n,1) distributed. Furthermore, the symplectic component $U_\gamma = \frac{Y}{R_\gamma}$, $|U_\gamma|_\gamma = 1$, is generalized uniformly distributed on the modified simplex $S_\gamma = \{y \in \mathbb{R}_+^n : |y|_\gamma = 1\}$.

Definition 1. The simplicial coordinate transformation $\text{Sim}_\gamma : [0, 1]^{(n-1)} \times [0, \infty) \rightarrow \mathbb{R}_+^n$, where $y = \text{Sim}_\gamma(\mu_1, \dots, \mu_{n-1}, \nu)$ is defined by

$$y_i = \begin{cases} \frac{\nu}{\gamma_i} \prod_{j=1}^{i-1} (1 - \mu_j) \mu_i, & i = 1, \dots, n-1, \\ \frac{\nu}{\gamma_n} \prod_{j=1}^{n-1} (1 - \mu_j), & i = n. \end{cases}$$

Lemma 3.1. The simplicial coordinate ν allows the representation $\nu = \sum_{i=1}^n \gamma_i y_i = |y|_\gamma$.

Proof. Notice that

$$\begin{aligned} \sum_{i=1}^n \gamma_i y_i &= \nu \sum_{i=1}^{n-1} \prod_{j=1}^{i-1} (1 - \mu_j) \mu_i + \nu \prod_{j=1}^{n-1} (1 - \mu_j) \\ &= \nu \sum_{i=1}^{n-2} \prod_{j=1}^{i-1} (1 - \mu_j) \mu_i + \nu(1 - \mu_1) \cdots (1 - \mu_{n-2}) \mu_{n-1} + \nu(1 - \mu_1) \cdots (1 - \mu_{n-2})(1 - \mu_{n-1}) \\ &= \nu \sum_{i=1}^{n-2} \prod_{j=1}^{i-1} (1 - \mu_j) \mu_i + \nu \prod_{j=1}^{n-2} (1 - \mu_j), \end{aligned}$$

where $\prod_{j=1}^0 = 1$. The assertion of the lemma follows by analogously repeating these calculations $(n-2)$ times. \square

Lemma 3.2. For the Jacobian J of the coordinate transformation Sim_γ , one gets

$$J = \frac{D(y_1, \dots, y_n)}{D(\nu, \mu_1, \dots, \mu_{n-1})} = \frac{(-\nu)^{n-1}}{\gamma_1 \cdots \gamma_n} \prod_{i=1}^{n-2} (1 - \mu_i)^{n-1-i}.$$

Proof. We have that $\tilde{J} = \frac{D(\gamma_1 y_1, \dots, \gamma_n y_n)}{D(\nu, \mu_1, \dots, \mu_{n-1})} = (-\nu)^{n-1} \prod_{i=1}^{n-2} (1 - \mu_i)^{n-1-i}$ (see [8]), and thus $J = \frac{D(y_1, \dots, y_n)}{D(\gamma_1 y_1, \dots, \gamma_n y_n)} \tilde{J}$. \square

For $N = 2$, we denote by $D(\mu_1, \mu_2) := \{\text{Sim}_\gamma(\mu, 1) | \mu_1 \leq \mu \leq \mu_2\} \in \mathfrak{B}(S_\gamma)$ a Borel measurable line segment (see Fig. B.3 (left plot)). Notice that $\rho D(\mu_1, \mu_2) = \{\text{Sim}_\gamma(\mu, \rho) | \mu_1 \leq \mu \leq \mu_2\}$ for any $\rho > 0$, since $\rho \text{Sim}_\gamma(\mu, 1) = \text{Sim}_\gamma(\mu, \rho)$.

Let us denote by $\text{Sector}_\gamma(D(\mu_1, \mu_2), \nu) := \bigcup_{\rho=0}^\nu \rho D(\mu_1, \mu_2)$ a set-union of line segments $D(\mu_1, \mu_2)$, see again Fig. B.3 (left plot), which can be formally understood to be a sector.

The next definition introduces the generalized arc-length measure L_γ .

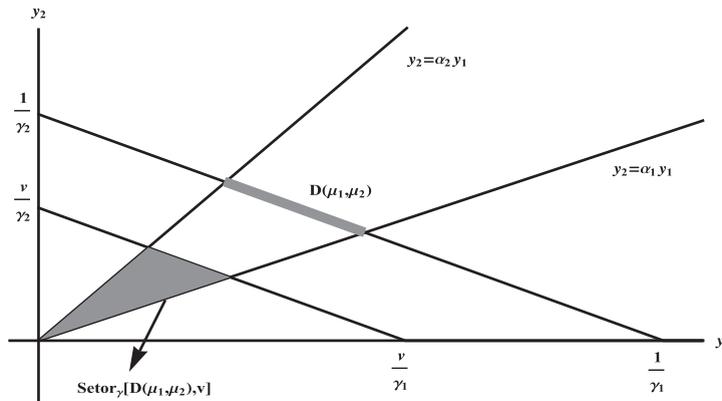


Fig. B.3. The geometric definition of set partitions based on $D^*(\alpha_1(v), \alpha_2(v))$.

310 **Definition 2.** Let $L_\gamma(D(\mu_1, \mu_2)) := \frac{d}{dv} \lambda(\text{Sector}_\gamma(D, v))|_{v=1}$.

311 Note that the generalized arc-length measure L_γ is closely connected with $O_{2,1}$ and $O_{\infty,1}$ where $O_{n,p}$ denotes the p -generalized
312 surface content in dimension n which was studied in [15,16].

313 In the following lemma, we calculate the section measure L_γ for the set $D(\mu_1, \mu_2)$, which enables us to define the geomet-
314 ric measure representation.

315 **Lemma 3.3.** The arc-length measure of the set $D(\mu_1, \mu_2)$ is $L_\gamma(D(\mu_1, \mu_2)) = \frac{\mu_2 - \mu_1}{\gamma_1 \gamma_2}$.

316 **Proof.** We have that $\lambda(\text{Sector}_\gamma(D(\mu_1, \mu_2), v)) = \int_{\rho=0}^v \int_{\mu=\mu_1}^{\mu_2} \frac{\rho}{\gamma_1 \gamma_2} d(\rho, \mu)$. Hence, according to Definition 3.2,
317 $L_\gamma(D(\mu_1, \mu_2)) = \int_{\mu_1}^{\mu_2} \frac{v}{\gamma_1 \gamma_2} d\mu|_{v=1}$. \square

318 **Lemma 3.4.** The line segment $D(\mu_1, \mu_2)$ may be represented as

321
$$D(\mu_1, \mu_2) = D^*\left(\frac{\gamma_1}{\gamma_2} \cdot \frac{1 - \mu_2}{\mu_2}, \frac{\gamma_1}{\gamma_2} \cdot \frac{1 - \mu_1}{\mu_1}\right)$$

322 for $0 \leq \mu_1 \leq \mu_2 \leq 1$, where $D^*(\alpha_1, \alpha_2) = \{(y_1, y_2) \in S_\gamma : \alpha_1 y_1 \leq y_2 \leq \alpha_2 y_1\}$ defines a line segment in a slightly different way than
323 $D(\cdot, \cdot)$ does.

324 **Proof.** Since $N = 2$, we obtain $\gamma_1 y_1 = v \mu$, $\gamma_2 y_2 = v(1 - \mu)$, and vice versa, $v = \gamma_1 y_1 + \gamma_2 y_2$ and $\mu = \frac{\gamma_1 y_1}{\gamma_1 y_1 + \gamma_2 y_2}$. Hence,

327
$$\mu_1 \leq \mu \leq \mu_2 \text{ iff; } \mu_1 \leq \frac{\gamma_1 y_1}{\gamma_1 y_1 + \gamma_2 y_2} \leq \mu_2 \text{ which is nothing else than } \frac{\gamma_1}{\gamma_2} \cdot \frac{1 - \mu_2}{\mu_2} \cdot y_1 \leq y_2 \leq \frac{\gamma_1}{\gamma_2} \cdot \frac{1 - \mu_1}{\mu_1} \cdot y_1. \quad \square$$

328 \blacktriangleleft

329 **Corollary 3.1.** The line segment D^* can be expressed in terms of the segment D as $D^*(\alpha_1, \alpha_1) = D\left(\frac{\gamma_1}{\gamma_1 + \alpha_2 \gamma_2}, \frac{\gamma_1}{\gamma_1 + \alpha_1 \gamma_2}\right)$.

330 **Proof.** Recognize that $\alpha_1 = \frac{\gamma_1}{\gamma_2} \cdot \frac{1 - \mu_2}{\mu_2}$ is the same as $\frac{\gamma_2 \alpha_1}{\gamma_1} = \frac{1}{\mu_2} - 1$ or $\frac{1}{\mu_2} = \frac{\alpha_1 \gamma_2 + \gamma_1}{\gamma_1}$ or finally $\mu_2 = \frac{\gamma_1}{\gamma_1 + \alpha_1 \gamma_2}$; Analogously,
331 $\alpha_2 = \frac{\gamma_1}{\gamma_2} \cdot \frac{1 - \mu_1}{\mu_1}$ is valid if and only if $\mu_1 = \frac{\gamma_1}{\gamma_1 + \alpha_2 \gamma_2}$. \square

332 The following definition specifies a class of sets which are appropriate for geometric integration method introduced in
333 this section.

335 **Definition 3.** A set $A \in \mathfrak{B}_+^2$ belongs to the system \mathcal{M}^* of subsets of R_+^2 , if there exist functions $\alpha_i : [0, \infty) \rightarrow [0, \infty)$, $i \in \{1, 2\}$,
336 with $0 < \alpha_1 < \alpha_2 < \infty$, and numbers $0 \leq v_0 < v_1 \leq \infty$, such that A allows the decomposition
337

339
$$A = \bigcup_{v=v_0}^{v_1} vD^*(\alpha_1(v), \alpha_2(v)),$$

340 see also Fig. B.3. Let \mathfrak{C}_γ denote the probability measure induced by $(E(\gamma_1), E(\gamma_2))$ on \mathfrak{B}_+^2 .

For $A \in \mathfrak{B}^2$ we have $\mathfrak{E}_\gamma(A) = \int_A f_\gamma(y) dy$, and from Lemma 3.2 it follows that

$$\mathfrak{E}_\gamma(A) = \int_{\text{Sim}_\gamma^{-1}(A)} \left(\prod_{i=1}^2 \gamma_i \right) e^{-v} - v \frac{d(\mu, \nu)}{\gamma_1 \gamma_2}.$$

Thus we have shown the following lemma.

Lemma 3.5. The measure satisfies the representation $\mathfrak{E}_\gamma(A) = \int_{\text{Sim}_\gamma^{-1}(A)} \nu e^{-v} d(\mu, \nu)$, $A \in \mathfrak{B}^2$.

Theorem 3.1. For $A \in \mathcal{M}^*$,

$$\mathfrak{E}_\gamma(A) = \gamma_1 \gamma_2 \int_{v_0}^{v_1} \nu e^{-v} \frac{\alpha_2(v) - \alpha_1(v)}{(\gamma_1 + \gamma_2 \alpha_1(v))(\gamma_1 + \gamma_2 \alpha_2(v))} d\nu.$$

Proof. We have that $\mathfrak{E}_\gamma(A) = \int_{\text{Sim}_\gamma^{-1}(A)} \nu e^{-v} d(\mu, \nu)$, and from Definition 3 that

$$A = \bigcup_{v=v_0}^{v_1} \nu D^*(\alpha_1(v), \alpha_2(v)).$$

So, from Corollary 3.1, we have $A = \bigcup_{v=v_0}^{v_1} \nu D\left(\frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_2(v)}, \frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_1(v)}\right)$ and from Definition 1, we obtain

$$\mathfrak{E}_\gamma(A) = \int_{v=v_0}^{v_1} \int_{\frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_2(v)}}^{\frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_1(v)}} \nu e^{-v} d(\mu, \nu).$$

Lemma 3.3 gives $\mathfrak{E}_\gamma(A) = \int_{v_0}^{v_1} \nu e^{-v} \gamma_1 \gamma_2 \mathcal{L}_D d\nu$ where

$$\mathcal{L}_D = L_\gamma\left(D\left(\frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_2(v)}, \frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_1(v)}\right)\right) = \frac{1}{\gamma_1 \gamma_2} \left(\frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_1(v)} - \frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_2(v)} \right) = \frac{\alpha_2(v) - \alpha_1(v)}{(\gamma_1 + \gamma_2 \alpha_1(v))(\gamma_1 + \gamma_2 \alpha_2(v))},$$

which gives the required result. \square

Remark 3.1. The L_γ arc-length measure of the simplex S_γ is

$$L_\gamma(S_\gamma) = \lim_{\alpha_2 \rightarrow \infty} L_\gamma\left(D\left(\frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_2}, \frac{\gamma_1}{\gamma_1 + \gamma_2 \alpha_1}\right)\right) \Big|_{\alpha_1=0} = \lim_{\alpha_2 \rightarrow \infty} \frac{\alpha_2}{\gamma_1(\gamma_1 + \gamma_2 \alpha_2)} = \frac{1}{\gamma_1 \gamma_2}.$$

Definition 4. The probability law $\omega_\gamma(D) = \frac{L_\gamma(D)}{L_\gamma(S_\gamma)}$, $D \in \mathfrak{B}(S_\gamma)$ is called the γ -generalized uniform distribution on the Borel subsets $\mathfrak{B}(S_\gamma)$ of S_γ .

Definition 5. The function $\mathcal{F}_\gamma(A, \nu) = \omega_\gamma(\frac{1}{\nu} A \cap S_\gamma)$, $\nu > 0$ is, for every $\gamma > 0$, called the $|\cdot|_\gamma$ -related intersection percentage function of the set A , $A \in \mathcal{M}^*$.

For analogous notions being used in the theory of $l_{n,p}$ -symmetric distributions, we refer to [15,17,8]. Note that the sets $\frac{1}{\nu} A \cap S_\gamma$ can be interpreted within a generalized method of indivisibles as discussed in these papers, too.

Corollary 3.2. For $A \in \mathcal{M}^*$, $\mathfrak{E}_\gamma(A) = \int_{v_0}^{v_1} \nu e^{-v} \mathcal{F}_\gamma(A, \nu) d\nu$.

Proof. From Corollary 3.1 and Remark 3.1, we have

$$\mathfrak{E}_\gamma(A) = \int_{v_0}^{v_1} \nu e^{-v} \frac{L_\gamma(D^*(\alpha_1(v), \alpha_2(v)))}{L_\gamma(S_\gamma)} d\nu,$$

where $D^*(\alpha_1(v), \alpha_2(v)) = (\frac{1}{\nu} A) \cap S_\gamma$. \square

Table B.3

Critical values C_x for $S_4 + R_4$ (upper table) and $S_5 + R_5$ (lower table) ($\sigma = 1$, $\nu = 1$, $\gamma = 1$).

$S_4 + R_4$	$C_{0.1}$	$C_{0.05}$	$C_{0.01}$	$C_{0.001}$
l -divergence (nsim = 10^5)	4.447175	5.43199	7.499567	10.59929
Convolution (approx. S_4 , MaxPoints = 10^3)	4.54	5.53	7.76	10.80
$S_5 + R_5$	$C_{0.1}$	$C_{0.05}$	$C_{0.01}$	$C_{0.001}$
l -divergence (nsim = 10^5)	5.317319	6.308037	8.584385	11.64449
Convolution (approximated S_5 , MaxPoints = 10^3)	5.39	6.46	8.81	11.99

Table B.4

Critical values $C_{0.05,N,\nu}$ for $I_N(y, \gamma)$ for different sample sizes. (Critical values were computed based on 100,000 samples.)

ν	Sample size														
	5	10	15	20	25	30	35	40	45	50	60	70	80	90	100
0.15	7.743	13.124	18.066	22.875	27.560	32.084	36.627	41.013	45.496	49.964	58.622	67.306	75.911	84.482	92.692
0.20	7.563	12.675	17.529	22.080	26.455	30.951	35.306	39.630	43.927	48.152	56.498	64.772	73.090	81.204	89.470
0.40	7.003	11.643	15.970	20.163	24.173	28.181	32.091	35.932	39.920	43.604	51.272	58.587	66.113	73.524	80.822
0.60	6.659	11.097	15.199	19.106	22.980	26.668	30.378	34.049	37.668	41.294	48.317	55.459	62.380	69.379	76.179
0.80	6.482	10.754	14.639	18.506	22.179	25.811	29.377	32.818	36.384	39.784	46.590	53.295	59.984	66.790	73.549
1.00	6.307	10.502	14.312	18.051	21.616	25.104	28.589	32.021	35.356	38.787	45.563	52.048	58.637	65.094	71.562
1.20	6.213	10.294	14.060	17.688	21.226	24.636	27.998	31.469	34.739	38.056	44.618	51.060	57.492	63.854	70.175
1.40	6.132	10.147	13.896	17.486	20.861	24.229	27.683	31.066	34.222	37.561	43.899	50.282	56.550	62.938	69.083
1.60	6.073	9.990	13.747	17.267	20.621	24.019	27.320	30.603	33.902	37.041	43.394	49.749	55.877	62.132	68.284
1.80	6.040	9.970	13.593	17.089	20.477	23.789	27.075	30.398	33.572	36.728	43.029	49.262	55.443	61.600	67.645
2.00	5.960	9.894	13.493	16.994	20.319	23.687	26.840	30.081	33.327	36.390	42.713	48.888	55.091	61.129	67.181
2.50	5.888	9.715	13.273	16.717	20.027	23.302	26.591	29.606	32.807	35.975	42.117	48.205	54.226	60.216	66.172
3.00	5.814	9.656	13.179	16.595	19.841	23.101	26.250	29.344	32.482	35.590	41.701	47.769	53.693	59.705	65.554
4.00	5.737	9.539	13.025	16.365	19.602	22.790	25.902	28.999	32.080	35.171	41.245	47.095	52.992	58.901	64.632
5.00	5.716	9.462	12.903	16.265	19.469	22.641	25.698	28.790	31.915	34.864	40.844	46.803	52.602	58.553	64.128
6.00	5.676	9.375	12.838	16.120	19.395	22.513	25.591	28.594	31.776	34.711	40.665	46.489	52.342	58.169	63.855
8.00	5.638	9.309	12.779	16.051	19.239	22.402	25.380	28.414	31.464	34.404	40.397	46.297	52.059	57.691	63.468

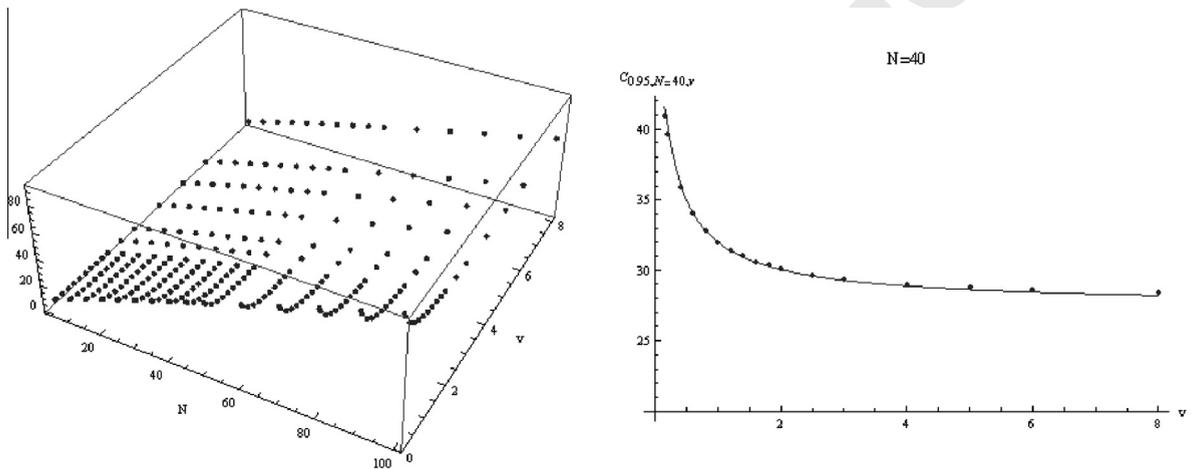


Fig. B.4. The critical values $C_{0.05,N,\nu}$ for $I_N(y, \gamma)$ with respect to the sample size N and the shape parameter ν (left plot) and the $C_{0.95,N,\nu}$ for $N = 40$ with the corresponding fraction polynomial of relationship (9) (right plot).

The expression from Corollary 3.2 well applies to the lower level sets of $I_N(y, \gamma)$, $N = 2$ i.e. to obtain the cumulative distribution function in an alternative way to convolution discussed in Sections 3.1–3.4. Moreover, it may be generalized to higher dimensions, too.

4. Simulation study

In the previous Sections, the decomposition of the I – divergence statistic was used in order to determine its null distribution. This decomposition allows to test separately the homogeneity and the scale hypotheses. In this section, we use simulations to test simultaneously both the hypotheses directly, without applying the exact distribution.

To this end, a detailed simulation study was carried out in order to compute the critical values of the I – divergence for various samples sizes and shape parameters. Furthermore, in this section, we also present the challenges to apply the proposed test when the shape parameter ν is unknown and has to be estimated.

4.1. Critical values

The simple form of the I – divergence statistic, and the fact that the null distribution of the I – divergence distance is independent from the assumed null scale parameter, make the computation of its critical values given the shape parameter ν a simple task for any sample size.

Table B.5

The simulated size of the proposed test for $\alpha = 0.05$, calculated after plugging in the estimated gamma shape parameter.

v	Sample size													
	10		20		30		40		60		80		100	
	ML	MM	ML	MM	ML	MM	ML	MM	ML	MM	ML	MM	ML	MM
0.4	0.0066	0.0990	0.0054	0.1042	0.0064	0.1074	0.0074	0.1190	0.0052	0.1316	0.0066	0.1160	0.0038	0.1020
0.8	0.0124	0.0620	0.0138	0.0734	0.0150	0.0788	0.0154	0.0936	0.0138	0.0834	0.0136	0.0936	0.0136	0.0882
1.0	0.0126	0.0590	0.0152	0.0706	0.0156	0.0758	0.0192	0.0858	0.0188	0.0788	0.0190	0.0750	0.0184	0.0806
1.2	0.0160	0.0538	0.0166	0.0612	0.0188	0.0706	0.0178	0.0688	0.0188	0.0728	0.0184	0.0752	0.0224	0.0792
1.6	0.0206	0.0512	0.0216	0.0576	0.0238	0.0656	0.0250	0.0610	0.0276	0.0736	0.0280	0.0710	0.0248	0.0622
2.0	0.0206	0.0484	0.0244	0.0540	0.0244	0.0532	0.0234	0.0550	0.0268	0.0608	0.0296	0.0624	0.0272	0.0650
4.0	0.0254	0.0418	0.0316	0.0474	0.0314	0.0508	0.0304	0.0470	0.0320	0.0506	0.0378	0.0596	0.0332	0.0526
8.0	0.0278	0.0358	0.0344	0.0430	0.0334	0.0432	0.0412	0.0490	0.0380	0.0454	0.0372	0.0442	0.0388	0.0470

Table B.6

The simulated size of the proposed test for $\alpha = 0.05$, calculated after plugging in both the estimated gamma parameters.

v	Sample size													
	10		20		30		40		60		80		100	
	ML	MM	ML	MM	ML	MM	ML	MM	ML	MM	ML	MM	ML	MM
0.4	0.0042	0.0914	0.0042	0.0958	0.0050	0.1006	0.0072	0.1136	0.0060	0.1354	0.0074	0.1278	0.0060	0.1390
0.8	0.0102	0.0568	0.0138	0.0712	0.0150	0.0786	0.0154	0.0936	0.0144	0.0868	0.0148	0.0974	0.0174	0.0968
1.0	0.0110	0.0548	0.0144	0.0688	0.0154	0.0746	0.0196	0.0862	0.0208	0.0830	0.0212	0.0786	0.0206	0.0848
1.2	0.0142	0.0502	0.0162	0.0592	0.0186	0.0702	0.0182	0.0704	0.0198	0.0744	0.0196	0.0768	0.0230	0.0798
1.6	0.0176	0.0480	0.0216	0.0576	0.0242	0.0658	0.0254	0.0614	0.0276	0.0730	0.0264	0.0688	0.0214	0.0584
2.0	0.0196	0.0464	0.0248	0.0550	0.0250	0.0548	0.0240	0.0552	0.0268	0.0608	0.0290	0.0612	0.0246	0.0616
4.0	0.0252	0.0418	0.0332	0.0512	0.0348	0.0552	0.0320	0.0486	0.0330	0.0520	0.0338	0.0542	0.0284	0.0434
8.0	0.0270	0.0348	0.0374	0.0452	0.0386	0.0490	0.0440	0.0522	0.0400	0.0470	0.0364	0.0438	0.0344	0.0430

As a consequence, in order to compute the critical values, a detailed simulation study was carried out based on the gamma distribution with scale parameter $\gamma = 1$. More specifically, 100,000 samples were generated under the null hypotheses for each combination of sample size ($N = 5, 10, \dots, 50, 60, \dots, 100$) and shape parameter ($v = 0.15, 0.2, 0.4, \dots, 2, 2.5, 3, 4, 5, 6, 8$) and the critical values for different significant levels α where computed. In Table B.4 of Appendix A, the critical values for $\alpha = 0.05$ are presented. It is clear from Table B.4 that the critical values increase as the sample size increases and decrease as the shape parameter v increases. In Fig. B.4 (left plot) the critical values $C_{0.05,N,v}$ are plotted with respect to the sample size N and the shape parameter v .

In order to determine possible non-linear functional relationships between the critical values, the sample size and the shape parameter v , we relied on fractional polynomials models [18,21]. The following relationship presents the best second degree fractional polynomials model obtained by this procedure

$$C_{0.05,N,v} = 1.28248 + 0.564928 \cdot N + 1.1303 \log N + \sqrt{\frac{N}{v}} \left(-0.496904 + 0.271317 \log \left(\frac{N}{v} \right) \right) + v^{-0.5} (0.864955 - 0.11949 \cdot v^{-3/2}). \tag{9}$$

The above relationship has an excellent fit to the critical values of Table B.4 from Appendix A. This is demonstrated both in the right plot of Fig. B.4 in which the above relationship is plotted for $N = 40$ with the corresponding critical values of Table B.4, and by the high coefficients of determination of the model ($R^2 = 0.9997$). The use of the above relationship is recommended for $0.15 \leq v \leq 8$ and $5 \leq N \leq 100$.

4.2. Plugging in an estimated shape parameter

The null distribution of the I -divergence, and as a consequence its critical values, depend not only on the sample size but also on the gamma shape parameter. Given a gamma sample with known shape parameter, the application of the I -divergence is straightforward with the help of Table B.4 (see Appendix A) or relationship (9). Unfortunately, the shape parameter is unknown in most of the real cases and has to be estimated. In such cases, it is clear that the behavior of the I -divergence depends on how good the available estimation is.

In order to investigate the performance of the I -divergence when the shape parameter is estimated and then plugged into the relationship (3), the following simulation study was carried out by generating 5000 gamma samples under the null hypotheses and for each combination of sample size ($N = 10, 20, 30, 40, 60, 80, 100$) and shape parameter ($v = 0.4, 0.8, 1, 1.2, 1.6, 2, 4$) (scale parameter was set equal to 1).

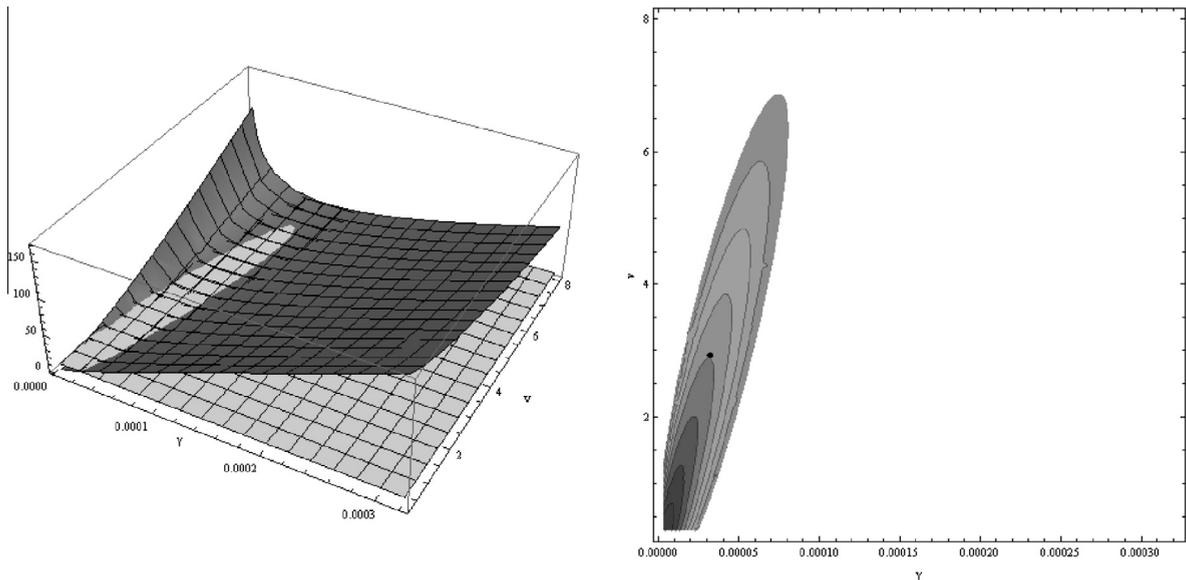


Fig. B.5. The I – divergence as a function of γ and ν for aerolane light indicators and the surface (in darker gray) determining the critical values (left plot) and the contour plot for the parameters' "credible" regions along with the point estimations obtain by different procedures (right plot).

For each sample, two of the most popular estimation procedures were used, the maximum likelihood (ML) and the method of moments (MM), and the estimation of the shape gamma parameter was obtained under the null hypotheses (homogeneity and $(H_0 : \gamma = 1)$). Then the estimated gamma shape parameter was plugged into the I – divergence and the value of the test statistic was calculated. Next, two different approaches were followed in order to calculate the critical value of the test. Firstly, the critical values from Table B.4 (Appendix A) were used (adopting the true gamma shape parameter). Secondly, the critical values were calculated from the relationship (9) with the use of the estimated shape parameter. In every case, the proportions of the samples for which the null hypotheses were rejected, or not, were recorded. The results from these approaches are presented in Tables B.5 and B.6 in Appendix A.

It follows from the results in both Tables that none of the applied procedures maintain the required size of the test ($\alpha = 0.05$) since in almost all the cases the size of the test differed more than $0.006 = 1.96\sqrt{0.05 \cdot 0.95/5000}$ from its nominal value. More specifically, the simulated size of the test depends on the underlying true value of the shape parameter, and the MLE increases while for the MM decreases as the shape parameter increases. Comparing the different approaches and estimation techniques, it seems that the use of the critical values of Table B.4 in Appendix A or of the critical values obtained from the relationship (9) with the use of the estimated shape parameter does not have a significant impact of the simulated size, and that the MLE presents generally a worse behavior than the MM. As a consequence, we can conclude that the plug-in strategy for the I – divergence does not provide an acceptable solution to the problem of unknown shape (and scale) parameter.

5. Applying the proposed test to real data

As was demonstrated in the previous section, the plug-in strategy does not seem to work well and so an alternative approach has to be developed in order to apply the proposed test to real data when no previous information is available for the gamma shape parameter. Such an approach is described just after a short discussion on the structure of the null hypotheses. Next the application of this approach is applied to real data sets in order to clarify the procedure.

5.1. The structure of the null hypotheses

The I – divergence given in relationship (3) can be used in order to test for homogeneity (5) and to check the scale (6) hypotheses simultaneously. So, if the value of the I – divergence is smaller than the critical value $C_{N,\alpha}$, both of the null hypotheses can not be rejected. On the other hand, if the value of the statistic is larger than the critical value we should conclude that at least one of the null hypothesis is rejected. Although, based on the structure of the hypotheses, we should assume that either both hypotheses are rejected or the scale hypothesis is rejected. More specifically, if homogeneity hypothesis holds then for some γ (probably not the one that we test) the $I(y, \gamma)$ should be smaller than $C_{N,\alpha}$, which is independent of the true γ . So, plugging in a value for the γ , as the MLE – which may not be a very good estimation, especially for small sample sizes, may result in a rejection of the null hypotheses due to the incorrect selection of the value of the γ

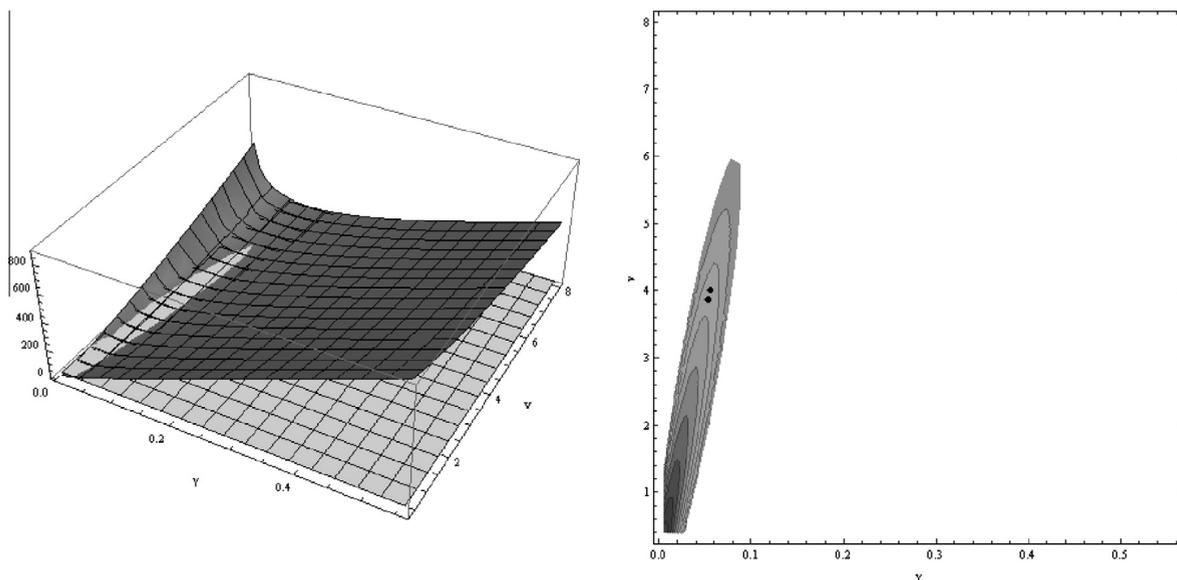


Fig. B.6. As in Fig. B.5 for the ball bearing data set.

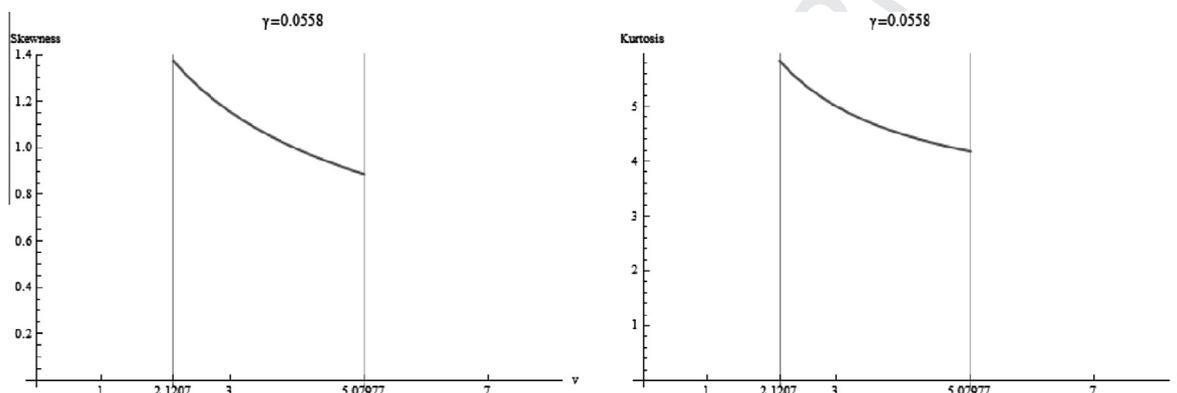


Fig. B.7. The skewness (left plot) and the kurtosis (right plot) of the gamma distribution for the non rejection “credible” region of the shape parameter using the MLE of the gamma scale parameter (0.0558).

453 parameter. The situation can be more severe when we estimate the shape parameter as we have seen in the previous Section.
454 To overcome this obstacle, we propose a different approach in the following subsection.

455 **5.2. Credible regions for the gamma parameters**

456 The application of the proposed test is straightforward when the value of the shape parameter and the scale parameter of
457 the gamma distribution can be assumed to be known. In such cases, we actually test indirectly a third hypothesis concerning
458 the adopted value of the shape parameter and by that point of view the test can be interpreted as a goodness of fit test for the
459 Gamma distribution. Unfortunately, the shape parameter is unknown in most of the cases. Additionally, in most real life
460 applications it is also unlikely to have any information and for the value of the scale parameter that we have to adopt for
461 the null scale hypothesis. As we have already mentioned, a misspecification of the null scale parameter γ and the use of
462 an estimated shape parameter to the $I -$ divergence can lead to a wrong decision. To overcome this problem, we propose
463 to treat the $I -$ divergence as a function of the scale γ and the shape ν parameters and to check if there exists a region of
464 values for the parameter for which none of the null hypotheses are rejected. A similar approach was applied by Economou
465 and Stehlík [7] in order to test the need of frailty modeling in survival and reliability data through homogeneity test.

466 This approach does not only allow us to avoid selecting a single value for the γ (which may be inaccurate) but also pro-
467 vides us with “credible” intervals for the values of both the parameters under which we can not reject the null hypotheses.
468 By this approach, along with the homogeneity and the scale hypotheses, we actually test indirectly a third hypothesis con-
469 cerning the value of the shape parameter.

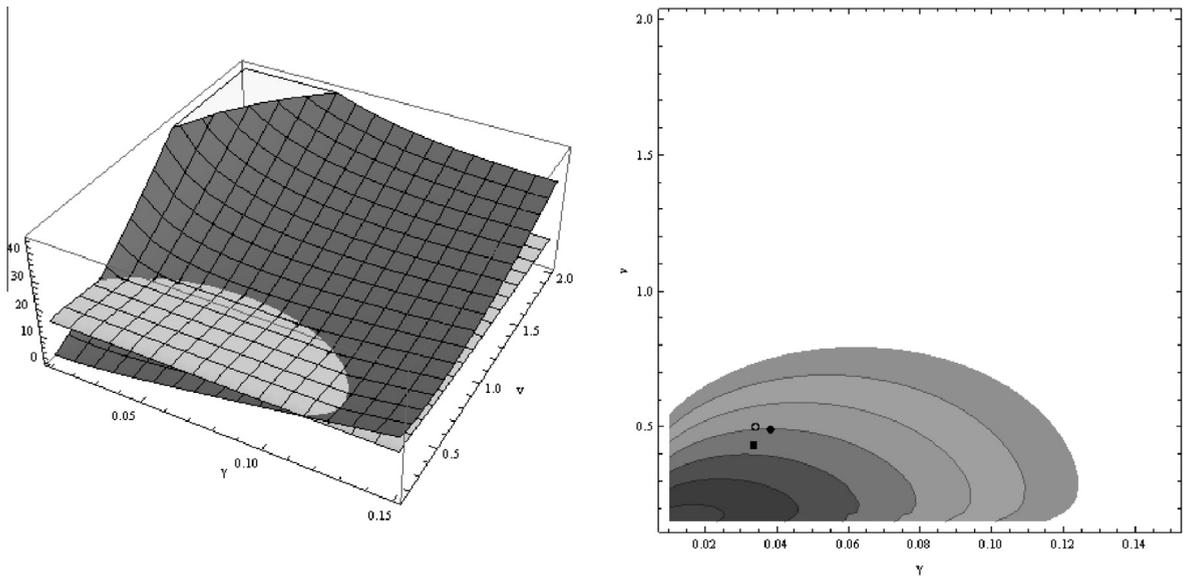


Fig. B.8. The I – divergence as a function of γ and ν , for the first data set from [10], and the surface (in darker gray) determining the critical values (left plot) and the contour plot for the parameters' "credible" regions along with the point estimators obtained by different procedures (right plot).

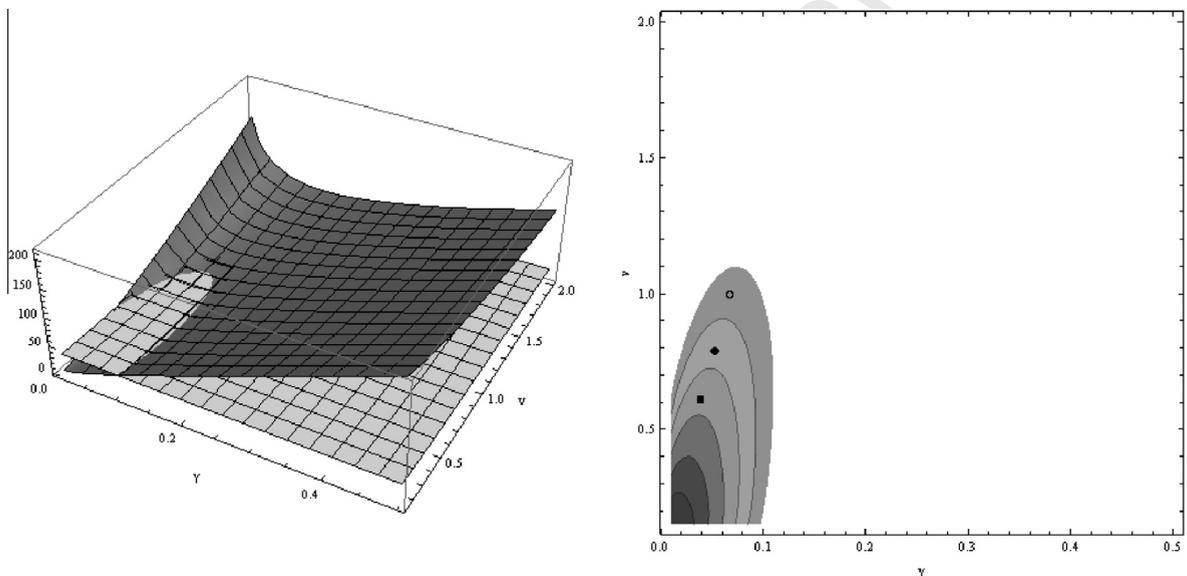


Fig. B.9. The I – divergence as a function of γ and ν for the second data set from [10], and the surface (in darker gray) determining the critical values (left plot) and the contour plot for the parameters' "credible" regions along with the point estimators obtained by different procedures (right plot).

470 It is true that such an approach does not solve directly the hypothesis testing problem but provides us with "credible"
 471 regions for the parameters that can be very useful for a researcher. More specifically, the lack of knowledge of the true value
 472 of the scale and the shape parameter can be overcome by the researchers choice or knowledge or belief on the true value of
 473 the parameters of the population under study by providing him/her a range of values of the parameters that leads to the non
 474 rejection of both the homogeneity and the scale hypotheses.

475 **5.3. Applications**

476 For illustration purposes, the previous described procedure is applied to four real data sets taken from literature and the
 477 aeroplane light indicators data set presented in Section 2. The applications for two of the data sets (taken from [10]) are pre-
 478 sented in Appendix, along with the Mathematica code for the accompanying plots.

479 As a first example, we use the data set presented in Section 2. Since the sample size is extremely small it is naturally to
 480 have little or no confidence on the MLE of the parameters. On the other hand the proposed approach gives some useful

information on the gamma parameters given that both the homogeneity and the scale hypothesis are not rejected. As a consequence, we treat $I_N(y, \gamma)$ not only as a function of γ but also as a function of v , and we construct a three dimensional plot of $I_N(y, \gamma)$ versus the γ and the v parameters. At the same plots, we have embedded also the surface (in darker gray) determined by relationship (3) for the corresponding sample size. These plots are presented on the left part of Fig. B.5. The two null hypotheses are not rejected for any point (γ_0, v_0) satisfying $I_N(y, \gamma_0) < C_{0.05, N, v=v_0}$, determining this way “credible” regions for the parameters γ and v . For these regions, none of the null hypotheses is rejected. At the right parts of Fig. B.5, these “credible” regions are also plotted with the help of a contour plot. The MLE (filled circle) and the estimator obtained by the method of moments (filled square), which all belong to the non rejection area for both the data sets are pointed also in the contour plot. The estimates are very close to each other and this is why the two points cannot be distinguish easily from the plot.

As a last example, we consider the data set presented in [12] and used in [11] for illustrating how their techniques for discriminating between the gamma and log-normal distribution work. The data consist of 23 records on the number of million revolutions before, failure of a ball bearing in a life test. Although, [11] suggest to choose the log-normal distribution rather than the gamma distribution for the data, we prefer to apply the proposed test. This is done because [11] also pointed out that the two fitted distributions (the gamma and the lognormal) are quite close to each other, and so we study the lifetime of a ball bearing under the gamma distribution by determining the gamma parameters’ “credible” regions for which none of the tested null hypotheses are rejected. Additionally, to strengthen this choice, it is worth mentioning that some of the sample descriptive statistics such as the skewness and the kurtosis (0.942 and 3.489 respectively) are closer to the corresponding values predicted by the fitted gamma (0.997 and 4.490) than the fitted lognormal distribution (1.852 and 9.658) (using the maximum likelihood method), making the choice between the two distributions more doubtful.

In Fig. B.6, the two plots along with the point estimations based on the maximum likelihood method and on the method of moments are presented. If we were willing to adopt the assumption of the gamma distribution and not to reject the homogeneity hypothesis we should adopt one point in the gray area of the right plot of Fig. B.6. When adopting, for example, the MLE of the gamma scale parameter (0.0558), we have to adopt a value between the 2.1207 and 5.07977 for the shape parameter. This region for the shape parameter belongs to a range of values for the skewness and the kurtosis for the distribution of the number of million revolutions before, failure of a ball bearing in the life test (see Fig. B.7), which not only can be consider reasonable but are also consistent with the corresponding characteristics of the sample. This should result into consider the gamma distribution at least as a plausible choice for describing the data.

6. Conclusions

Summarizing, we illustrated in this paper the possibility of divergence testing for reliability engineering. In particular, we have illustrated the importance of decomposition of divergences, which may provide us a form of statistical regularization or optimal statistical procedures. We have also derived the exact distribution by geometric integration theory and by convolutions. These approaches shows a high technical complexity of exact distributions, and as a consequence the use of approximation is needed for $N > 4$. Finally, the application of the introduced tests in real data sets was addressed by constructing “credible” regions for the gamma parameters for which none of the homogeneity and the scale hypothesis is rejected.

Acknowledgments

The authors are grateful to the Editor and a Reviewer for their valuable hints.

Appendix A

A.1. Lambert W function

The Lambert W function is defined to be the multivalued inverse of the complex function $f(y) = ye^y$. As the equation $ye^y = z$ has an infinite number of solutions for each (non-zero) value of $z \in \mathbf{C}$, the Lambert W has an infinite number of branches. Exactly one of these branches is analytic at 0. Usually this branch is referred to as the principal branch of the Lambert W and is denoted by W or W_0 . The other branches all have a branch point at 0. These branches are denoted by W_k where $k \in \mathbf{Z} \setminus \{0\}$. The principal branch and the pair of branches W_{-1} and W_1 share an order 2 branch point at $z = -e^{-1}$. A detailed discussion of the branches of the Lambert W can be found in [6]. For more information about the implementation and some computational aspects, see [5].

A.2. Functions forming domain of integration for G_4

If we denote

$$A_{\pm}(x) = \frac{3}{4} - \frac{S}{4\sqrt{2}} \pm \frac{1}{2} \sqrt{\frac{1}{2} - \frac{3e^{-x}}{8g} - \frac{3}{8}g + \frac{1}{\sqrt{2}S}},$$

where

$$S = \sqrt{2 + \frac{3e^{-x}}{g} + 3g}$$

and

$$g = \left(e^{-x} + e^{-3x} \sqrt{-e^{3x} + e^{4x}} \right)^{1/3},$$

then

$$a(x) = A_-(x), \quad b(x) = A_+(x).$$

A.3. Additional applications and Mathematica Code

The two data sets given in the Mathematica code at the end of this section present the empirical variance of the repeated measurements of the glucose levels of diabetic patients. The first data set consists of the variances of the duplicate measurements of 13 diabetic patients and the second data set consists of the variances of triplicate measurements of 31 diabetic patients. In [10], these sample variances were assumed to be distributed according to $\Gamma(1/2, \gamma/2)$ and $\Gamma(1, \gamma)$, respectively, where γ is the within-patients population variance. A pooled estimate for γ given by [10] is 1/14.56.

In a first approach, we test for these data sets the hypothesis

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_N \text{ versus } H_A := \text{non } H_0$$

and the scale hypothesis

$$H_0 : \gamma = \gamma_0 \text{ versus } H_A : \gamma \neq \gamma_0,$$

where $\gamma_0 = 1/14.56$ is the pooled estimate for γ .

For both the data sets, the null hypotheses are not rejected since the value of the I – divergence statistic using the assumed shape parameter is each time smaller than the critical value obtained by relationship (3) in the main text. More specifically, we have for the first data set that

$$I_N(x, \gamma_0/2) = 8.48464 < 14.2447 = C_{0.05, N=13, v=1/2}$$

and for the second one, that

$$I_N(x, \gamma_0) = 23.3075 < 25.843 = C_{0.05, N=31, v=1}.$$

Kimber [10] argued that it is likely to be identify some outliers in these two data sets. Thus, one may have little or no trust to the pooled estimate for the γ parameter. Additionally, there is no theoretical background about the assumed shape parameters. This is why we will additionally apply the proposed approach described in the previous subsection in order to determine “credible” regions for both the parameters for which neither the homogeneity nor the scale hypotheses is rejected.

Thus, we treat $I_N(y, \gamma)$ not only as a function of γ but also as a function of v , and we construct a three dimensional plot of $I_N(y, \gamma)$ versus the γ and the v parameters for each sample. At the same plots, we embedded also the surface (in darker gray) determined by relationship (3) for the corresponding sample size. These plots are presented on the left part of Figs. B.8 and B.9 for the first and the second data set. The two null hypotheses are not rejected for any point (γ_0, v_0) satisfying $I_N(y, \gamma_0) < C_{0.05, N, v=v_0}$ which allows us to determine in this way “credible” regions for the parameters γ and v . For these regions, none of the null hypotheses is rejected. At the right parts of these figures, the “credible” regions are also plotted with the help of a contour plot. The point suggested by [10] (empty circle), the MLE (filled circle) and the estimator obtained by the method of moments (filled square), which all belong to the non rejection area for both the data sets, are pointed also in the contour plot.

The analysis of the data was carried out by the following Mathematica Code.

```

Idiv[_ , v_ , data_] := -Sum[v - vLog[v], {i, 1, Length[data]}] +
  Sum[data[[i]] - vLog[data[[i]]], {i, 1, Length[data]}];
Idivcrit[n_, v_] := 1.28248 + 0.564928n -
  0.496904(n/v)^0.5 - 0.11949/v^2. + 0.864955/v^0.5 + 1.1303Log[n] +
  0.271317(n/v)^0.5Log[n/v]
(*First data set*)
datal={13.005,18.000,0.605,19.845,75.645,1.805,1.125,3.125,
  6.125,1.805,11.045,0.005,13.52}
EstimatedDistribution[datal, GammaDistribution[v, gamma]];
EstimatedDistribution[datal, GammaDistribution[v, gamma],

```

```

ParameterEstimator-> "MethodOfMoments"];
a1 = Plot3D[{Idiv[,v,data1],
  Idivcrit[Length[data2],v]},{.01,.15},{v,0.15,2},
  PlotStyle-> {Gray,Automatic},ClippingStyle-> Opacity[0.15],
  AxesLabel-> {"","v"}];
a2 = DensityPlot[{If[
  Idiv[,v,data1]-Idivcrit[Length[data1],v]< 0,
  Idiv[,v,data2]-Idivcrit[Length[data1],v],]}, {.01,.15},{v,0.15,2},
  ColorFunction-> "AuroraColors",FrameLabel-> Automatic];
a41 = ListPlot[{1/(214.56),1/2}],PlotStyle-> Black,
  PlotMarkers-> "[EmptyCircle]";
a42 = ListPlot[{1/25.94420693337358',0.4911575189180527'}],
  PlotStyle-> Black,PlotMarkers-> "[FilledCircle]";
a43 = ListPlot[{1/29.328669709753576',0.4344790416271278'}],
  PlotStyle-> Black,PlotMarkers-> "[FilledSquare]";
a4 = Show[{a41,a42,a43}];
a3 = ColorConvert[
  Show[GraphicsGrid[{a1,Show[a2,a4]}],ImageSize-> 900],
  "Grayscale"]
(*Second data set*)
data2={29.043,12.653,1.363,22.943,32.363,7.770,28.210,1.963,
  7.413,10.943,7.930,12.463,30.970,0.723,6.040,1.470,0.813,
  1.293,26.080,5.590,29.403,0.790,98.023,5.373,14.573,0.903,
  4.890,13.080,3.610,41.333,3.253};
EstimatedDistribution[data2,GammaDistribution[v,gamma]];
EstimatedDistribution[data2,GammaDistribution[v,gamma],
  ParameterEstimator-> "MethodOfMoments"];
a1 = Plot3D[{Idiv[,v,data2],
  Idivcrit[Length[data1],v]},{.01,.5},{v,0.15,2},
  PlotStyle-> {Gray,Automatic},ClippingStyle-> Opacity[0.15],
  AxesLabel-> {"","v"}];
a2 = DensityPlot[{If[Idiv[,v,data2]-Idivcrit[Length[data2],v]< 0,
  Idiv[,v,data2]-Idivcrit[Length[data2],v],]}, {.01,.5},{v,0.15,2},
  ColorFunction-> "AuroraColors",FrameLabel-> Automatic];
a41 = ListPlot[{1/14.56,1}],PlotStyle-> Black,
  PlotMarkers-> "[EmptyCircle]";
a42 = ListPlot[{1/18.894547252006998',0.7909211358640081'}],
  PlotStyle-> Black,PlotMarkers-> "[FilledCircle]";
a43 = ListPlot[{1/24.313932291118682',0.614631010535975'}],
  PlotStyle-> Black,PlotMarkers-> "[FilledSquare]";
a4 = Show[{a41,a42,a43}];
a3 = ColorConvert[
  Show[GraphicsGrid[{a1,Show[a2,a4]}],ImageSize-> 900],
  "Grayscale"]

```

640

641 **References**

- 642 [1] O. Barndorf-Nielsen, Information and Exponential Families in Statistical Theory, John Wiley & Sons Ltd., New York, 1978.
- 643 [2] M.S. Bartlett, D.G. Kendall, The statistical analysis of variance-heterogeneity and the logarithmic transformation, *Suppl. J. R. Stat. Soc.* 8 (1946) 128–
- 644 138.
- 645 [3] K.C.G. Chan, Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation,
- 646 *Biometrika* 100 (2012) 269–276.
- 647 [4] D. Coit, T. Jin, Gamma distribution parameter estimation for field reliability data with missing failure times, *IIE Trans.* 32 (2000) 1161–1166.
- 648 [5] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, D. Knuth, Lambert's W function in maple, *Maple Tech. Newsl.* 9 (1993) 12–22.
- 649 [6] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, D. Knuth, On the Lambert W function, *Adv. Comput. Math.* 5 (1996) 329–359.
- 650 [7] P. Economou, M. Stehlik, On small samples testing for frailty through homogeneity test, *Commun. Stat. Simul. Comput.* (in press).
- 651 [8] V. Henschel, W.-D. Richter, Geometric generalization of the exponential law, *J. Multi. Anal.* 81 (2002) 189–204.
- 652 [9] A. Karagrigoriou, K. Mattheou, On distributional properties and goodness of fit tests, *Commun. Stat. Theory Methods* 39 (2010) 472–482.
- 653 [10] A.C. Kimber, Trimming in gamma samples, *J. R. Stat. Soc. Ser. C Appl. Stat.* (32) (1983) 7–14.
- 654 [11] D. Kundu, A. Manglick, Discriminating between the gamma and log-normal distributions, *J. Appl. Stat. Sci.* 14 (2005) 175–187.
- 655 [12] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1982.
- 656 [13] L. Pardo, *Statistical Inference Based on Divergence Measures*, Chapman & Hall/CRC, 2006.

- 657 [14] A. Pázman, Nonlinear Statistical Models, Kluwer Acad Publ., Dordrecht, 1993 (Sections 9.1 and 9.2).
658 [15] W.-D. Richter, Continuous $I_{n,p}$ symmetric distributions, Lith. Math. J. 49 (1) (2009) 93–108.
659 [16] W.-D. Richter, Ellipses numbers and geometric measure representation, J. Appl. Anal. 17 (2011) 165–179.
660 [17] W.-D. Richter, Exact distributions under non-standard model assumptions, AIP Conf. Proc. 1479 (2012) 442–445.
661 [18] P. Royston, D. Altman, Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling (with discussion), Appl.
662 Stat. 43 (1994) 429–467.
663 [19] F. Rublík, On optimality of the LR tests in the sense of exact slopes. Part 1: General case, Kybernetika 25 (1989) 13–25.
664 [20] F. Rublík, On optimality of the LR tests in the sense of exact slopes. Part 2, Kybernetika 25 (1989) 117–135.
665 [21] W. Sauerbrei, P. Royston, Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials, J.
666 R. Stat. Soc. Ser. A 162 (1999) 71–94.
667 [22] M. Stehlík, Distributions of exact tests in the exponential family, Metrika 57 (2003) 145–164.
668 [23] M. Stehlík, Homogeneity and scale testing of generalized gamma distribution, Reliab. Eng. Syst. Saf. 93 (2008) 1809–1813.
669 [24] M. Stehlík, Decompositions of information divergences: recent development, open problems and applications, AIP Conf. Proc. 1493 (2012) 972–976.
670 [25] M. Stehlík, Discussion for paper “How to find an appropriate clustering for mixed type variables with application to socio-economic stratification” by
671 Ch. Hennig and T.F. Liao, JRSS-C 62 (2013) 1–25.
672 [26] M. Stehlík, J. Kiseľák, G.A. Ososkov, I-divergence based testing with applications to reliability and physics, Math. Eng. Aerosp. Sci. 3 (2012) 417–432.
673 [27] S.S. Wilks, Mathematical Statistics, John Wiley & Sons Ltd., New York, 1962.
674

UNCORRECTED PROOF