ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 14, No. 4
October 2016

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 845 40 84
    Web site: http://www.ine.pt
    Customer Support Service
    (National network) : 808 201 808
    Other networks: + 351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 14, Number 4, October 2016, 349–394

DENSITY OF A RANDOM INTERVAL CATCH DIGRAPH FAMILY AND ITS USE FOR TESTING UNIFORMITY

Author: Elvan Ceyhan
– Department of Mathematics, Koç University, Istanbul, Turkey
elceyhan@ku.edu.tr

Received: June 2014 Revised: February 2015 Accepted: February 2015

Abstract:

- We consider (arc) density of a parameterized interval catch digraph (ICD) family with random vertices residing on the real line. The ICDs are random digraphs where randomness lies in the vertices and are defined with two parameters, a centrality parameter and an expansion parameter, hence they will be referred as central similarity ICDs (CS-ICDs). We show that arc density of CS-ICDs is a U-statistic for vertices being from a wide family of distributions with support on the real line, and provide the asymptotic (normal) distribution for the (interiors of) entire ranges of centrality and expansion parameters for one dimensional uniform data. We also determine the optimal parameter values at which the rate of convergence (to normality) is fastest. We use arc density of CS-ICDs for testing uniformity of one dimensional data, and compare its performance with arc density of another ICD family and two other tests in literature (namely, Kolmogorov–Smirnov test and Neyman's smooth test of uniformity) in terms of empirical size and power. We show that tests based on ICDs have better power performance for certain alternatives (that are symmetric around the middle of the support of the data).

Key-Words:

- *asymptotic normality; class cover catch digraph; intersection digraph; Kolmogorov–Smirnov test; Neyman's smooth test; proximity catch digraph; random geometric graph; U-statistics.*

AMS Subject Classification:

- 05C80, 05C20, 60D05, 60C05, 62E20.

1. INTRODUCTION

Intersection graphs have received considerable attention in literature since their introduction. The main reasons for this attention are their applications in real life and their "tame" behavior, in the sense that many problems that are NP-hard for graphs in general are solvable in polynomial time for intersection graphs ([Prisner, 1994]). Intersection digraphs are introduced by [Beineke and Zamfirescu, 1982] who called them "connection digraphs". Let V be an index set and (S_v, T_v) be ordered pairs of sets associated with the elements v of V, where S_v is called the *source* and T_v is called the target or *sink* set ([Douglas, 1996]). The intersection digraph associated with this collection of ordered pairs is $D = (V, A)$ which has vertex set V and arc (i.e., directed edge) set A with $(u, v) \in A$ iff $S_u \cap T_v \neq \emptyset$. When the source and sink sets are intervals, we obtain interval digraphs ([Douglas, 1996]). If the set T_v resides in S_v for each $v \in V$, then the ordered pair set is a *nest representation* for the interval digraph, and if T_v is just a point residing in S_v, it is called a *catch representation*. A digraph is called an *interval catch digraph* (ICD), if it is an intersection digraph with a catch representation ([Prisner, 1994]). The set of ordered pairs, $\{S_v, T_v = \{p_v\}\}$, in the catch representation for the ICD is also called a "pointed set" where S_v is a set with base point p_v ([Prisner, 1989]). Equivalently, an ICD is the catch digraph of a family of pointed intervals of T if $(T, \leq)$ is a totally ordered set. Indeed, [Maehara, 1984] provides a simple characterization of ICDs for finite n, for which one can always take $T = \mathbb{R}$.

The ICDs we consider in this article are defined in a randomized setting. Our ICDs are vertex random digraphs in which each vertex corresponds to a random data point from a distribution, and arcs are defined by a bivariate relation using the regions based on these data points. Our ICDs are a special type of proximity graphs which were introduced by [Toussaint, 1980], and are closely related to the class cover problem of [Cannon and Cowen, 2000] and proximity catch digraphs (PCDs) which were introduced recently and have applications in spatial data analysis and statistical pattern classification ([Ceyhan and Priebe, 2005]).

In this article, we define *central similarity (CS) ICDs* for one dimensional data which may also be viewed as one dimensional version of the PCDs considered in [Ceyhan *et al.*, 2007]. We derive the asymptotic distribution of the *arc density* of CS-ICDs for random data points. For undirected simple graphs, the *edge density* (also called *graph density*) is defined as the ratio of number of edges in the graph to the total number of edges possible with the same number of vertices. So the edge density is $2|E|/\big(n(n-1)\big)$ for a graph $G = (V, E)$ with $|V| = n$. The minimal density is 0, which is attained for empty graphs (i.e., for $E = \emptyset$) and the maximal density is 1, which is attained for complete graphs

([Coleman and Moré, 1983]). Based on the graph density concept, 'dense' and 'sparse' graphs are defined. For a dense graph, graph density is close to 1 and for sparse graphs it is close to 0. There are other quantities related to graph density, such as average degree which is defined as $2|E|/n$ ([Goldberg, 1984]); edge density of a graph is also defined as $|E|/n$ in literature (see, e.g., [Grünbaum, 1988]). Notice that both of these quantities are scaled versions of the edge or graph density, $2|E|/\big(n(n-1)\big)$. On the other hand, density of a digraph is the ratio of number of arcs in a given digraph with n vertices to the total number of arcs possible (i.e., to the number of arcs in a complete symmetric digraph of order n). Hence for a simple digraph $D = (V, A)$ with vertex set $|V| = n$ and arc set A, *digraph density* (or *arc density*) is $|A|/\big(n(n-1)\big)$, which is the quantity of interest in this article. Arc density is also referred to as *relative density* in literature. Properly scaled, the arc density of the ICDs is a U-statistic, which yields the asymptotic normality by the general central limit theory of U-statistics ([Lehmann, 2004]). Our ICDs can also be viewed as a generalization of class cover catch digraphs (CCCDs) which was introduced by [Priebe *et al.*, 2001]. CS-ICDs have two defining parameters, a centrality and an expansion parameter. Here, we derive the explicit form of the asymptotic normal distribution of the arc density of the CS-ICDs for the (interiors of) entire ranges of these parameters for uniform one dimensional data from a class whose support being partitioned by points from another class. We investigate the arc density of CS-ICDs for uniform data in one interval (in $\mathbb{R}$) and the analysis is generalized to uniform data in multiple intervals (see Remark 4.1). We determine the optimal parameters for the rate of convergence to normality and show that arc density of CS-ICDs has a faster rate than that of the respective optimal parameter values of another ICD family called proportional-edge (PE) ICDs which were introduced in [Ceyhan, 2012] (and therein referred to as proportional-edge proximity catch digraphs). We employ the arc density of CS-ICDs for testing uniformity of one dimensional data and compare its performance with two prevalent tests in literature (namely, Kolmogorov–Smirnov test and Neyman's smooth test) in terms of size and power as well as arc density of the PE-ICDs. Testing uniformity of one-dimensional data is of substantial importance in various fields, e.g., for assessing the goodness-of-fit problems ([Marhuenda *et al.*, 2005]). For this purpose, some graph theoretical tools are used in literature although not so commonly; e.g., minimum spanning trees are employed for testing uniformity of two-dimensional data ([Jain *et al.*, 2002]). However, to the best of author's knowledge, arc density is not previously employed for testing uniformity of one-dimensional data. The tests based on the arc density of the ICD families have been shown to have better power performance for certain types of alternatives (which are symmetric around the midpoint of the support of the distribution) against uniformity. CS-ICDs can also be used for testing spatial patterns between (two or more) classes of data points.

We define the ICDs and describe the random ICDs and CS-ICDs in Section 2, define their arc density and provide preliminary results in Section 3,

provide the distribution of the arc density for uniform data in one interval in Section 4, present the size and power analysis and comparison with other tests as well as some consistency results in Section 5, and discussion and conclusions in Section 6. Shorter proofs are given in the main body of the article; while longer proofs are deferred to the Appendix.

2. RANDOM INTERVAL CATCH DIGRAPHS

Let $(\Omega, \mathcal{F}, P_x)$ be a probability space equipped with a metric $d\colon \Omega\times\Omega \to [0,\infty)$. Our random catch digraphs will be defined in a randomized setting where vertices are randomly generated in Ω and the associated metric distance will be taken to be the Euclidean distance. Let $\mathcal{X}_n = \{X_1, X_2, ..., X_n\}$ and $\mathcal{Y}_m = \{Y_1, Y_2, ..., Y_m\}$ be two sets of Ω-valued random variables from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively, whose joint probability distribution is $F_{X,Y}$ with marginals F_X and F_Y, respectively. Our random catch digraph will be based on $\mathcal{X}_n$ and $\mathcal{Y}_m$. More specifically, we choose $\mathcal{X}$ points to be the vertices and put an arc from X_i to X_j, based on a binary relation which measures the relative allocation of X_i and X_j with respect to $\mathcal{Y}$ points. In particular, in our setting, the $\mathcal{Y}$ points will be used to partition the support set Ω, and the relative position of X_i and X_j with respect to $\mathcal{Y}$ points will be determined by the Euclidean distances between X_i, X_j, and the $\mathcal{Y}$ points. Notice that the randomness is only on the vertices, hence our catch digraphs are *vertex random*. Given $\mathcal{Y}_m \subseteq \Omega$, let $\mathcal{P}(\Omega)$ represent the power set of Ω, then *proximity map* $N_{\mathcal{Y}}\colon \Omega \to \mathcal{P}(\Omega)$ maps each point $x \in \Omega$ to a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$. A *vertex random catch digraph* has the vertex set $\mathcal{V} = \mathcal{X}_n$ and arc set $\mathcal{A}$ defined by $(X_i, X_j) \in \mathcal{A}$ if $X_j \in N_{\mathcal{Y}}(X_i)$ for $i \neq j$. Hence the binary relation defining the digraph is based on the proximity region, $N_{\mathcal{Y}}$, which indicates the relative allocation of $\mathcal{X}$ points with respect to $\mathcal{Y}$ points. Notice also that arcs of the form (X_i, X_i) (i.e., loops) are not allowed in our catch digraph definition. If loops were allowed, the corresponding digraph would have been called a *pseudodigraph* according to some authors (see, e.g., [Chartrand *et al.*, 2010]). We also define arc probability, denoted $p_a(i,j)$, between two vertices X_i and X_j as $p_a(i,j) := P((X_i, X_j) \in \mathcal{A})$ for all $i \neq j$, $i, j = 1, 2, ..., n$. If $\mathcal{X}_n$ is a random sample from F_X, then $p_a(i,j) = p_a$ for all $i \neq j$, $i, j = 1, 2, ..., n$. For calculations leading to the distribution of arc density of ICDs, we also need a concept which is dual to proximity regions. For a set $B \subseteq \Omega$, the Γ_1-*region* is the image of the map $\Gamma_1(\cdot, N_{\mathcal{Y}})\colon \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ that assigns the region $\Gamma_1(B, N_{\mathcal{Y}}) := \{z \in \Omega\colon B \subseteq N_{\mathcal{Y}}(z)\}$ to the set B. For a point $x \in \Omega$, we denote $\Gamma_1(\{x\}, N_{\mathcal{Y}})$ as $\Gamma_1(x, N_{\mathcal{Y}})$. The concept of Γ_1-region is introduced in [Ceyhan and Priebe, 2005] and is associated with another graph invariant called domination number (which is denoted as γ). In a proximity graph, if a vertex falls in the Γ_1-region, then the domination number would equal to 1. For brevity, we drop the subscript $\mathcal{Y}$ in the notation, $N_{\mathcal{Y}}$, henceforth.

2.1. Central Similarity ICDs

For one dimensional data, we have $\Omega = \mathbb{R}$, then there is a natural partitioning of the real line based on $\mathcal{Y}$ points. Let $Y_{(i)}$ be the i^{th} order statistic of $\mathcal{Y}_m$ for $i = 1, 2, ..., m$, with the extension that $-\infty =: Y_{(0)}$ and $Y_{(m+1)} := \infty$ and assume $Y_{(i)}$ values are distinct (which happens with probability one for continuous distributions). The $Y_{(i)}$ values partition $\mathbb{R}$ into $(m+1)$ intervals, with $\left(-\infty, Y_{(1)}\right)$ and $\left(Y_{(m)}, \infty\right)$ being the *end intervals*, and $\left(Y_{(i-1)}, Y_{(i)}\right)$ for $i = 2, ..., m$ being the *middle intervals*. For one dimensional data sets, $\mathcal{X}_n$ and $\mathcal{Y}_m$, we define the CS-ICD with expansion parameter $\tau > 0$ and centrality parameter $c \in (0, 1)$ as follows. For $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ (i.e., for x in a middle interval) with $i \in \{2, ..., m\}$ and $M_{c,i} = Y_{(i-1)} + c\left(Y_{(i)} - Y_{(i-1)}\right) \in \left(Y_{(i-1)}, Y_{(i)}\right)$, that is $c \times 100\,\%$ of $\left(Y_{(i)} - Y_{(i-1)}\right)$ is to the left of $M_{c,i}$, we define the CS proximity region as follows:

$$N(x, \tau, c) = \begin{cases} \left(x - \tau\left(x - Y_{(i-1)}\right),\ x + \frac{\tau(1-c)}{c}\left(x - Y_{(i-1)}\right)\right) \cap \left(Y_{(i-1)}, Y_{(i)}\right) & \text{if } x \in \left(Y_{(i-1)}, M_{c,i}\right), \\ \left(x - \frac{c\,\tau}{1-c}\left(Y_{(i)} - x\right),\ x + \tau\left(Y_{(i)} - x\right)\right) \cap \left(Y_{(i-1)}, Y_{(i)}\right) & \text{if } x \in \left(M_{c,i}, Y_{(i)}\right). \end{cases} \tag{2.1}$$

Notice that dependence on $\mathcal{Y}$ points is explicit in the definition of the CS proximity region. Furthermore, the Euclidean distance is implicit in the terms $\left(x - Y_{(i-1)}\right)$ and $\left(Y_{(i)} - x\right)$, where the former is $d\left(x, Y_{(i-1)}\right)$ and the latter is $d\left(x, Y_{(i)}\right)$. This definition yields two types of regions for $N(x, \tau, c)$, one with $\tau \in (0, 1]$ and the other with $\tau > 1$. For $\tau \in (0, 1]$, we have

$$N(x, \tau, c) = \begin{cases} \left(x - \tau\left(x - Y_{(i-1)}\right),\ x + \frac{\tau(1-c)}{c}\left(x - Y_{(i-1)}\right)\right) & \text{if } x \in \left(Y_{(i-1)}, M_{c,i}\right), \\ \left(x - \frac{c\,\tau}{1-c}\left(Y_{(i)} - x\right),\ x + \tau\left(Y_{(i)} - x\right)\right) & \text{if } x \in \left(M_{c,i}, Y_{(i)}\right), \end{cases} \tag{2.2}$$

and with $\tau > 1$, we have

$$N(x, \tau, c) = \begin{cases} \left(Y_{(i-1)}, x + \frac{\tau(1-c)}{c}\left(x - Y_{(i-1)}\right)\right) & \text{if } x \in \left(Y_{(i-1)}, \frac{c\,Y_{(i)} + \tau(1-c)\,Y_{(i-1)}}{c + \tau(1-c)}\right), \\ \left(Y_{(i-1)}, Y_{(i)}\right) & \text{if } x \in \left(\frac{c\,Y_{(i)} + \tau(1-c)\,Y_{(i-1)}}{c + \tau(1-c)}, \frac{(1-c)\,Y_{(i-1)} + c\,\tau\,Y_{(i)}}{1 - c + c\,\tau}\right), \\ \left(x - \frac{c\,\tau}{1-c}\left(Y_{(i)} - x\right), Y_{(i)}\right) & \text{if } x \in \left(\frac{(1-c)\,Y_{(i-1)} + c\,\tau\,Y_{(i)}}{1 - c + c\,\tau}, Y_{(i)}\right). \end{cases} \tag{2.3}$$

For an illustration of $N(x, \tau, c)$ in the middle interval case, see Figure 1 (left) where $\mathcal{Y}_2 = \{y_1, y_2\}$ with $y_1 = 0$ and $y_2 = 1$ (hence $M_{c,2} = c$).

Additionally, for x in an end interval, i.e., $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ with $i \in \{1, m+1\}$, the central similarity proximity region depends on the expansion parameter only.

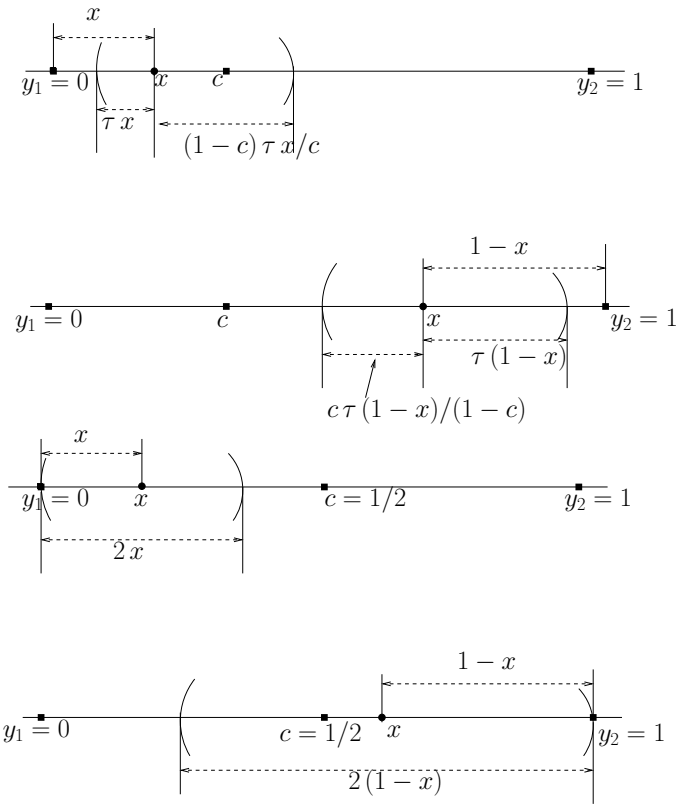

Figure 1: Plotted in the top two rows are illustrations of the construction of central similarity proximity regions, $N(x, \tau, c)$ with $\tau \in (0, 1]$, $\mathcal{Y}_2 = \{y_1, y_2\}$ with $y_1 = 0$ and $y_2 = 1$ (hence $M_{c,2} = c$) and $x \in (0, c)$ (first row) and $x \in (c, 1)$ (second row); and in the bottom two rows are the proximity regions associated with CCCD, i.e., $N(x, \tau = 1, c = 1/2)$ for an $x \in (0, 1/2)$ (third row) and $x \in (1/2, 1)$ (fourth row).

So we denote the central similarity proximity region for an x in an end interval as $N_e(x, \tau)$. Then with $\tau \in (0, 1]$, we have

$$N_e(x,\tau) = \begin{cases} \left(x - \tau\left(Y_{(1)} - x\right),\, x + \tau\left(Y_{(1)} - x\right)\right) & \text{if } x < Y_{(1)}\,, \\ \left(x - \tau\left(x - Y_{(m)}\right),\, x + \tau\left(x - Y_{(m)}\right)\right) & \text{if } x > Y_{(m)}\,, \end{cases} \tag{2.4}$$

and with $\tau > 1$, we have

$$N_e(x,\tau) = \begin{cases} \left(x - \tau\left(Y_{(1)} - x\right),\, Y_{(1)}\right) & \text{if } x < Y_{(1)}\,, \\ \left(Y_{(m)},\, x + \tau\left(x - Y_{(m)}\right)\right) & \text{if } x > Y_{(m)}\,. \end{cases} \tag{2.5}$$

If $x \in \mathcal{Y}_m$, then we define $N(x, \tau, c) = \{x\}$ and $N_e(x, \tau) = \{x\}$ for all $\tau > 0$, and if $x = M_{c,i}$, then in Equation (2.1), we arbitrarily assign $N(x, \tau, c)$ to be one of

the two defining intervals. For X from a continuous distribution, these special cases in the construction of central similarity proximity region — $X \in \mathcal{Y}_m$ and $X = M_{c,i}$ — occur with probability zero. Notice that $\tau > 0$ implies $x \in N(x, \tau, c)$ for all $x \in \left[Y_{(i-1)}, Y_{(i)}\right]$ with $i \in \{2, ..., m\}$ and $x \in N_e(x, \tau)$ for all $x \in \left[Y_{(i-1)}, Y_{(i)}\right]$ with $i \in \{1, m+1\}$. Furthermore, $\lim_{\tau \to \infty} N(x, \tau, c) = \left(Y_{(i-1)}, Y_{(i)}\right)$ for all $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ with $i \in \{2, ..., m\}$, so we define $N(x, \infty, c) = \left(Y_{(i-1)}, Y_{(i)}\right)$ for all such x. Similarly, $\lim_{\tau \to \infty} N_e(x, \tau) = \left(Y_{(i-1)}, Y_{(i)}\right)$ for all $x \in \left(Y_{(i-1)}, Y_{(i)}\right)$ with $i \in \{1, m+1\}$, so we define $N_e(x, \infty) = \left(Y_{(i-1)}, Y_{(i)}\right)$ for all such x. In the special case of $c = 1/2$ and $\tau = 1$, central similarity proximity region $N(x, \tau, c)$ is identical to the proportional edge proximity region with centrality parameter 1/2 and expansion parameter 2 (see [Ceyhan, 2012]).

In a vertex random CS-ICD, the vertex set is $\mathcal{X}_n$ and arc set $\mathcal{A}$ is defined by $(X_i, X_j) \in \mathcal{A} \iff X_j \in N(X_i, \tau, c)$ for X_i, X_j with $i \neq j$ in the middle intervals and $(X_i, X_j) \in \mathcal{A} \iff X_j \in N_e(X_i, \tau)$ for X_i, X_j with $i \neq j$ in the end intervals. We denote such digraphs as $\mathbf{D}_{n,m}(\tau, c)$. When $\tau = 1$ and $c = 1/2$ $\left(\text{i.e., } M_{c,i} = \left(Y_{(i-1)} + Y_{(i)}\right)/2\right)$ we have $N(x, 1, 1/2) = B(x, r(x))$ for an x in a middle interval and $N_e(x, 1) = B(x, r(x))$ for an x in an end interval where $r(x) = d(x, \mathcal{Y}_m) = \min_{y \in \mathcal{Y}_m} d(x, y)$ and the corresponding ICD is the CCCD of [Priebe *et al.*, 2001] or the proportional-edge PCD (PE-PCD) of [Ceyhan, 2012] with expansion parameter 2 and centrality parameter 1/2. See also Figure 1 (right).

3. ARC DENSITY OF CS-ICDS

For a digraph $D_n = (\mathcal{V}, \mathcal{A})$ with vertex set $\mathcal{V}$ and arc set $\mathcal{A}$, the arc density of D_n which is of order $|\mathcal{V}| = n \geq 2$, denoted $\rho(D_n)$, is defined as $\rho(D_n) = \frac{|\mathcal{A}|}{n(n-1)}$ ([Janson *et al.*, 2000]) where $|\cdot|$ stands for the set cardinality function. So $\rho(D_n)$ is the ratio of the number of arcs in the digraph D_n to the number of arcs in the complete symmetric digraph of order n, which is $n(n-1)$. For $n \leq 1$, we set $\rho(D_n) = 0$.

Let $\mathbb{I}_{ij} = \mathbb{I}\big((X_i, X_j) \in \mathcal{A}\big) = \mathbb{I}\big(X_j \in N(X_i)\big)$. Then for an ICD (hence for a CS-ICD), we can write the arc density as

$$\rho(D_n) \;=\; \frac{2}{n(n-1)} \sum_{i<j} h_{ij}$$

where $h_{ij} := (\mathbb{I}_{ij} + \mathbb{I}_{ji})/2$ Since the digraph is based on a relation that is not symmetric, h_{ij} is defined as half of the number of arcs between X_i and X_j in order to produce a symmetric kernel with finite variance ([Lehmann, 2004]). Notice that

$$\mathbf{E}\big[\rho(D_n)\big] \;=\; \mathbf{E}\left[h_{12}\right] \;=\; p_a$$

and

$$0 \le \mathbf{Var}\big[\rho(D_n)\big] = \frac{2}{n(n-1)}\mathbf{Var}\left[h_{12}\right] + \frac{4(n-2)}{n(n-1)}\mathbf{Cov}\left[h_{12},h_{13}\right] \le 1/4$$

where

$$\mathbf{Var}[h_{ij}] = \mathbf{Var}\left[h_{12}\right] = \frac{1}{4}\mathbf{Var}\big[\mathbb{I}_{12}+\mathbb{I}_{21}\big] = (p_a+p_{sa})/2-(1-p_a)^2 ,$$

where $p_{sa} = P\big(\{(X_i,X_j),(X_j,X_i)\}\subset\mathcal{A}\big)$ is the symmetric arc probability and

$$\mathbf{Cov}\left[h_{12},h_{13}\right] = \mathbf{E}\left[h_{12}h_{13}\right] - p_a^2 ,$$

with

$$\begin{aligned} 4\,\mathbf{E}\left[h_{12}h_{13}\right] &= 4\,\mathbf{E}\big[(\mathbb{I}_{12}+\mathbb{I}_{21})(\mathbb{I}_{13}+\mathbb{I}_{31})\big] \\ &= P\Big(\{X_2,X_3\}\subset N(X_1)\Big) + 2\,P\Big(X_2\in N(X_1), X_3\in\Gamma_1(X_1,N)\Big) \\ &\quad + P\Big(\{X_2,X_3\}\subset\Gamma_1(X_1,N)\Big) . \end{aligned}$$

See [Ceyhan, 2012] for the derivations. Since $\rho(D_n)$, is a one-sample U-statistic of degree 2 and is an unbiased estimator of the arc probability p_a, a CLT for U-statistics ([Lehmann, 2004]) yields $\sqrt{n}\left[\rho(D_n)-p_a\right] \xrightarrow{\mathcal{L}} \mathbb{N}(0,4\,\nu)$ as $n\to\infty$, where $\xrightarrow{\mathcal{L}}$ stands for convergence in law and $\mathbb{N}(\mu,\sigma^2)$ stands for the normal distribution with mean μ and variance σ^2 provided $\nu = \mathbf{Cov}\left[h_{ij},h_{ik}\right] > 0$ for all $i\neq j\neq k,\ i,j,k\in\{1,2,...,n\}$.

Since $\mathbf{E}\left[|h_{ij}|^3\right]\le 1$, for $\nu>0$, the sharpest rate of convergence in the asymptotic normality of $\rho(D_n)$ is

$$\begin{aligned} \sup_{t\in\mathbb{R}}\left|P\left(\frac{\sqrt{n}\left(\rho(D_n)-p_a\right)}{\sqrt{4\,\nu}}\le t\right)-\Phi(t)\right| &\le 8\,K\,p_a\,(4\nu)^{-3/2}\,n^{-1/2} \\ &= K\,\frac{p_a}{\sqrt{n\,\nu^3}} , \end{aligned} \tag{3.1}$$

where K is a constant and $\Phi(t)$ is the distribution function for the standard normal distribution ([Callaert and Janssen, 1978]).

3.1. Distribution of the arc density of CS-ICDs

We consider CS-ICDs for which $\mathcal{X}_n$ and $\mathcal{Y}_m$ are random samples from F_X and F_Y, respectively, so that the joint distribution of X,Y is $F_{X,Y}\in\mathcal{F}(\mathbb{R})$ where

$$\mathcal{F}(\mathbb{R}) := \Big\{F_{X,Y} \text{ on } \mathbb{R} \text{ with } P(X=Y)=0 \text{ and the marginals, } F_X \text{ and } F_Y, \text{ are non-atomic}\Big\} .$$

Then the order statistics of $\mathcal{X}_n$ and $\mathcal{Y}_m$ are distinct with probability one. We denote such digraphs as $\mathbf{D}_{n,m}(F,\tau,c)$ and focus on the random variable $\rho_{n,m}(F,\tau,c) := \rho\big(\mathbf{D}_{n,m}(F,\tau,c)\big)$. Clearly $0 \leq \rho_{n,m}(F,\tau,c) \leq 1$, and $\rho_{n,m}(F,\tau,c) > 0$ for nontrivial digraphs.

We first partition the real line based on $\mathcal{Y}$ points. Along this line, we let $\mathcal{Y}_{[i]} := \big\{Y_{(i-1)}, Y_{(i)}\big\}$, $\mathcal{I}_i := \big(Y_{(i-1)}, Y_{(i)}\big)$, and $\mathcal{X}_{[i]} := \mathcal{X}_n \cap \mathcal{I}_i$ for $i = 1, 2, ..., (m+1)$. Let $\mathbf{D}_{[i]}(F,\tau,c)$ be the component of the random CS-ICD induced by the vertices in $\mathcal{X}_{[i]}$ (and based on $\mathcal{Y}_{[i]}$). Then we have a disconnected digraph with subdigraphs, each might be null or itself disconnected and denoted as $\mathbf{D}_{[i]}(F,\tau,c)$ for $i = 1, 2, ..., (m+1)$. Let $\mathcal{A}_{[i]}$ be the arc set of $\mathbf{D}_{[i]}(F,\tau,c)$, and $\rho_{[i]}(F,\tau,c)$ denote the arc density of $\mathbf{D}_{[i]}(F,\tau,c)$; $n_i := \big|\mathcal{X}_{[i]}\big|$, and F_i be the distribution F_X restricted to $\mathcal{I}_i$ for $i \in \{1, 2, ..., m+1\}$. Furthermore, let $M_{c,i} \in \mathcal{I}_i$ be the point so that it divides the interval $\mathcal{I}_i$ in ratios c and $1-c$. Since we have at most $m+1$ subdigraphs $\mathbf{D}_{[i]}(F,\tau,c)$ each of which having at most $n_i(n_i-1)$ arcs, it follows that we can have at most $n_{_T} := \sum_{i=1}^{m+1} n_i(n_i-1)$ arcs in the digraph $\mathbf{D}_{n,m}(F,\tau,c)$. We adjust the arc density for the entire digraph as

$$\widetilde{\rho}_{n,m}(F,\tau,c) := \frac{|\mathcal{A}|}{n_{_T}} = \frac{\sum_{i=1}^{m+1} |\mathcal{A}_{[i]}|}{n_{_T}} = \frac{1}{n_{_T}} \sum_{i=1}^{m+1} \big(n_i(n_i-1)\big)\, \rho_{[i]}(F,\tau,c)\,. \tag{3.2}$$

Hence, $\widetilde{\rho}_{n,m}(F,\tau,c)$ is called as the *adjusted arc density* and is a mixture of the $\rho_{[i]}(F,\tau,c)$ values, since $\frac{n_i(n_i-1)}{n_{_T}} \geq 0$ for each i and $\sum_{i=1}^{m+1} \frac{n_i(n_i-1)}{n_{_T}} = 1$. We first focus on the simpler random variable $\rho_{[i]}(F,\tau,c)$. The almost sure (a.s.) results follow from the marginal distributions F_X and F_Y being non-atomic in the rest of this section.

Lemma 3.1. *For $i \in \{1, (m+1)\}$ (i.e., in the end intervals) if $n_i \leq 1$, then $\rho_{[i]}(F,\tau,c) = 0$ for all $\tau > 0$. Moreover, if $n_i > 1$, then $\rho_{[i]}(F,\tau,c) \geq 1/2$ a.s. for all $\tau > 1$.*

Proof: By symmetry, distribution of $\rho_{[i]}(F,\tau,c)$ is same for $i = 1, m+1$. So we only consider $i = m+1$ (i.e., the right end interval). If $n_{m+1} \leq 1$, then by definition $\rho_{[m+1]}(\tau,c) = 0$. So, assume $n_{m+1} > 1$ and let $\mathcal{X}_{[m+1]} = \{Z_1, Z_2, ..., Z_{n_{m+1}}\}$ with $Z_{(j)}$ being the corresponding order statistics. Then there is an arc from $Z_{(j)}$ to each $Z_{(k)}$ for $k < j$, with $j, k \in \{1, 2, ..., n_{m+1}\}$ (and possibly to some other Z_l) for all $\tau > 1$, since $N_e\big(Z_{(j)}, \tau\big) = \big(Y_{(m)}, Z_{(j)} + \tau\,\big(Z_{(j)} - Y_{(m)}\big)\big)$ and so $Z_{(k)} \in N_e\big(Z_{(j)}, \tau\big)$. This implies that there are at least $0 + 1 + 2 + \cdots + n_{m+1} - 1 = n_{m+1}(n_{m+1}-1)/2$ arcs in $D_{[m+1]}(\tau,c)$. Then $\rho_{[m+1]}(\tau,c) \geq \big(n_{m+1}(n_{m+1}-1)/2\big)/\big(n_{m+1}(n_{m+1}-1)\big) = 1/2$. □

Let $\mathbf{D}_{n,m}(F,\tau,c)$ be a CS-ICD with $n > 0$ and $m > 0$. Then we obtain the following lower bound for $\rho_{n,m}(F,\tau,c)$ with $\tau > 1$.

Theorem 3.1. *Let k_1 and k_2 be two natural numbers defined as $k_1 := \sum_{i=2}^{m}\big(n_{i,\ell}(n_{i,\ell}-1)/2 + n_{i,r}(n_{i,r}-1)/2\big)$ and $k_2 := \sum_{i\in\{1,m+1\}} n_i(n_i-1)/2$, where $n_{i,\ell} := \big|\mathcal{X}_n \cap \big(Y_{(i-1)}, M_{c,i}\big)\big|$ and $n_{i,r} := \big|\mathcal{X}_n \cap \big(M_{c,i}, Y_{(i)}\big)\big|$. Then for $\tau > 1$, we have $(k_1+k_2)/n_{_T} \le \rho_{n,m}(F,\tau,c) \le 1$ a.s.*

Proof: We have k_2 as in Lemma 3.1 for the end intervals (i.e., for $i \in \{1,(m+1)\}$). In the middle intervals, i.e., for $i \in \{2,3,...,m\}$, let $\mathcal{X}_{i,\ell} := \mathcal{X}_{[i]} \cap \big(Y_{(i-1)}, M_{c,i}\big) = \{U_1, U_2, ..., U_{n_{i,\ell}}\}$, and $\mathcal{X}_{i,r} := \mathcal{X}_{[i]} \cap \big(M_{c,i}, Y_{(i)}\big) = \{V_1, V_2, ..., V_{n_{i,r}}\}$. Furthermore, let $U_{(j)}$ and $V_{(k)}$ be the corresponding order statistics. For $\tau > 1$, there is an arc from $U_{(j)}$ to $U_{(k)}$ and possibly to some other U_l for $k < j$ with $j,k,l \in \{1,2,...,n_{i,\ell}\}$, and similarly there is an arc from $V_{(j)}$ to $V_{(k)}$ and possibly to some other V_l for $k > j$ with $j,k,l \in \{1,2,...,n_{i,r}\}$. Therefore, we have $\rho_{n,m}(F,\tau,c) \ge (k_1+k_2)/n_{_T}$, since there are at least $n_{i,\ell}(n_{i,\ell}-1)/2 + n_{i,r}(n_{i,r}-1)/2$ arcs in $\mathbf{D}_{[i]}(F,\tau,c)$. □

Theorem 3.2. *When the expansion parameter is infinity (i.e., $\tau = \infty$), we have $\rho_{_{[i]}}(\tau=\infty,c) = \mathbb{I}(n_i > 1)$ and $\rho_{n,m}(\tau=\infty,c) = 1$ a.s. for $i = 1,2,3,...,m+1$ and $n_i > 1$.*

Proof: For $\tau = \infty$, if $n_i \le 1$, then $\rho_{_{[i]}}(\tau=\infty,c) = 0$. So we assume $n_i > 1$ and let $i = m+1$. Then $N_e(x,\infty) = \big(Y_{(m)}, \infty\big)$ for all $x \in \big(Y_{(m)}, \infty\big)$. Hence $D_{[m+1]}(\infty,c)$ is a complete symmetric digraph of order n_{m+1}, which implies $\rho_{_{[m+1]}}(\tau=\infty,c) = 1$. By symmetry, the same holds for $i = 1$. For $i \in \{2,3,...,m\}$ and $n_i > 1$, we have $N(x,\infty,c) = \mathcal{I}_i$ for all $x \in \mathcal{I}_i$, hence $D_{[i]}(\infty,c)$ is a complete symmetric digraph of order n_i, which implies $\rho_{_{[i]}}(\infty,c) = 1$. Then $\rho_{n,m}(\infty,c) = \sum_{i=1}^{m+1} \frac{n_i(n_i-1)\,\rho_{_{[i]}}(\infty,c)}{n_{_T}} = 1$, since when $n_i \le 1$, n_i has no contribution to $n_{_T}$, and when $n_i > 1$, we have $\rho_{_{[i]}}(\infty,c) = 1$. □

4. DISTRIBUTION OF THE ARC DENSITY OF CS-ICDS FOR UNIFORM DATA

Let $\mathcal{X}_n = \{X_1, X_2, ..., X_n\}$ be a random sample from $F_X = \mathcal{U}(\delta_1,\delta_2)$, the uniform distribution on the bounded interval (δ_1,δ_2), and let $\mathcal{Y}_m$ be a random sample from non-atomic F_Y with support $\mathcal{S}(F_Y) \subseteq (\delta_1,\delta_2)$. Then $F_{X,Y} \in \mathcal{F}(\mathbb{R})$. Suppose we have a realization of $\mathcal{Y}_m$ as $\mathcal{Y}_m = \{y_1, y_2, ..., y_m\}$ with the order statistics satisfying $\delta_1 < y_{(1)} < y_{(2)} < \cdots < y_{(m)} < \delta_2$, with the extension that $y_{(0)} := \delta_1$ and $y_{(m+1)} := \delta_2$. Then the distribution of X_i restricted to $\mathcal{I}_i$ is $F_X|_{\mathcal{I}_i} = \mathcal{U}(\mathcal{I}_i)$. We provide the distribution of the arc density of $\mathbf{D}_{n,m}(\tau,c)$ for the whole range of the parameters τ and c. The following "scale invariance" for CS-ICDs will

allow us to consider the special case of the unit interval $(0,1)$ as the support of $\mathcal{X}$ points, thereby simplifying the computations in our subsequent analysis.

Theorem 4.1 (Scale Invariance Property). *Let $\mathcal{Y}_m$ be a set of m distinct $\mathcal{Y}$ points in a bounded interval (δ_1,δ_2) and $\mathcal{X}_n$ be random sample from $\mathcal{U}(\delta_1,\delta_2)$. Then the distribution of $\rho_{[i]}(\tau,c)$ is independent of $\mathcal{Y}_{[i]}$ (and hence independent of the restricted support interval $\mathcal{I}_i$) for all $i\in\{1,2,...,m+1\}$, $\tau>0$, and $c\in(0,1)$.*

Proof: Let δ_1 and δ_2 and $\mathcal{Y}_m$ be as in the hypothesis. Any $\mathcal{U}(\delta_1,\delta_2)$ random variable can be transformed into a $\mathcal{U}(0,1)$ random variable by $\phi(x)=(x-\delta_1)/(\delta_2-\delta_1)$, which maps intervals $(t_1,t_2)\subseteq(\delta_1,\delta_2)$ to intervals $\big(\phi(t_1),\phi(t_2)\big)\subseteq(0,1)$. That is, if $X\sim\mathcal{U}(\delta_1,\delta_2)$, then we have $\phi(X)\sim\mathcal{U}(0,1)$ and $P\big(X\in(t_1,t_2)\big)=P\big(\phi(X)\in\big(\phi(t_1),\phi(t_2)\big)\big)$ for all $(t_1,t_2)\subseteq(\delta_1,\delta_2)$. The distribution of $\rho_{[i]}(\tau,c)$ is obtained by calculating such probabilities. So, without loss of generality, we can assume $\mathcal{X}_{[i]}$ is a set of iid (independent identically distributed) random variables from the $\mathcal{U}(0,1)$ distribution. That is, the distribution of $\rho_{[i]}(\tau,c)$ does not depend on $\mathcal{Y}_{[i]}$ and hence does not depend on the restricted support interval $\mathcal{I}_i$. □

For $\tau=\infty$, we have $\rho_{[i]}(\tau=\infty,c)=1$ a.s. for any non-atomic F_X with support in (δ_1,δ_2), hence the scale invariance of $\rho_{[i]}(\tau=\infty,c)$ holds for all $\mathcal{X}_n$ from any such F_X. Based on Theorem 4.1, we may assume each $\mathcal{I}_i$ as the unit interval $(0,1)$ for uniform data. If $x\in\mathcal{I}_i$ for $i\in\{2,...,m\}$ (i.e., in the middle intervals), when transformed to $(0,1)$, the central similarity proximity region for $x\in(0,1)$ with parameters $c\in(0,1)$ and $\tau>0$ is

$$N(x,\tau,c)=\begin{cases}\Big((1-\tau)x,\Big(1+\frac{(1-c)}{c}\tau\Big)x\Big)\cap(0,1) & \text{if } x\in(0,c)\,,\\ \Big(x-\frac{c\,\tau}{(1-c)}(1-x),\,x+(1-x)\tau\Big)\cap(0,1) & \text{if } x\in(c,1)\,.\end{cases}\tag{4.1}$$

In particular, for $\tau\in(0,1]$, we have

$$N(x,\tau,c)=\begin{cases}\Big((1-\tau)x,\Big(1+\frac{(1-c)}{c}\tau\Big)x\Big) & \text{if } x\in(0,c)\,,\\ \Big(x-\frac{c\,\tau}{(1-c)}(1-x),\,x+(1-x)\tau\Big) & \text{if } x\in(c,1)\,,\end{cases}\tag{4.2}$$

and for $\tau>1$, we have

$$N(x,\tau,c)=\begin{cases}\Big(0,\Big(1+\frac{(1-c)}{c}\tau\Big)x\Big) & \text{if } x\in\Big(0,\frac{c}{c+(1-c)\,\tau}\Big),\\ (0,1) & \text{if } x\in\Big(\frac{c}{c+(1-c)\,\tau},\frac{c\,\tau}{1-c+c\,\tau}\Big),\\ \Big(x-\frac{c\,\tau}{(1-c)}(1-x),1\Big) & \text{if } x\in\Big(\frac{c\,\tau}{1-c+c\,\tau},1\Big),\end{cases}\tag{4.3}$$

and $N(x=c,\tau,c)$ is arbitrarily taken to be one of the two defining intervals above. But the case of "$X=c$" happens with probability zero for uniform X.

Furthermore, when transformed to $(0,1)$, if x is in the left end interval (i.e., $x \in \mathcal{I}_1$), we have $N_e(x,\tau) = \big(\max\big(0, x-\tau(1-x)\big), \min\big(1, x+\tau(1-x)\big)\big)$; and if x is in the right end interval (i.e., $x \in \mathcal{I}_{m+1}$), we have $N_e(x,\tau) = \big(\max\big(0, (1-\tau)x\big), \min\big(1, (1+\tau)x\big)\big)$.

Each subdigraph $D_{[i]}(\tau,c)$ is itself a random CS-ICD (for brevity of notation, we suppress the dependence on the uniform distribution). The distribution of the arc density of $D_{[i]}(\tau,c)$ is given in the following theorem.

Theorem 4.2. *Let $\rho_{[i]}(\tau,c)$ be the arc density of subdigraph $D_{[i]}(\tau,c)$ of the CS-ICD based on $\mathcal{U}(\delta_1,\delta_2)$ data and $\mathcal{Y}_m$ be a set of m distinct $\mathcal{Y}$ points in (δ_1,δ_2). Then, as $n_i \to \infty$, for $\tau \in (0,\infty)$ we have,*

(**i**) $\sqrt{n_i}\left[\rho_{[i]}(\tau,c) - p_a(\tau,c)\right] \xrightarrow{\mathcal{L}} \mathbb{N}(0, 4\,\nu(\tau,c))$, *where* $p_a(\tau,c) = \mathbf{E}\Big[\rho_{[i]}(\tau,c)\Big]$ *is the arc probability and* $\nu(\tau,c) = \mathbf{Cov}[h_{12}, h_{13}]$ *in the middle intervals (i.e., for* $i \in \{2,...,m\}$*), and*

(**ii**) $\sqrt{n_i}\left[\rho_{[i]}(\tau,c) - p_a^e(\tau,c)\right] \xrightarrow{\mathcal{L}} \mathbb{N}(0, 4\,\nu_e(\tau))$, *where* $p_a^e(\tau,c) = \mathbf{E}\Big[\rho_{[i]}(\tau,c)\Big]$ *is the arc probability and* $\nu_e(\tau) = \mathbf{Cov}[h_{12}, h_{13}]$ *in the end intervals (i.e., for* $i \in \{1, m+1\}$*).*

Proof: By Theorem 1 of [Ceyhan, 2012], arc density of CS-ICDs is a U-statistic, and hence the proofs follow by the asymptotic normality of U-statistics provided the asymptotic variance is positive. In particular, in (i) by the scale invariance for uniform data (see Theorem 4.1), a middle interval can be assumed to be the unit interval $(0,1)$. Then

$$\mathbf{E}\Big[\rho_{[i]}(\tau,c)\Big] \;=\; \mathbf{E}[h_{12}] \;=\; P\Big(X_2 \in N(X_1,\tau,c)\Big) \;=\; p_a(\tau,c)$$

which is the arc probability. Similarly in (ii) we have $\mathbf{E}\Big[\rho_{[i]}(\tau,c)\Big] = \mathbf{E}[h_{12}] = P\big(X_2 \in N_e(X_1,\tau)\big) = p_a^e(\tau,c)$.

Furthermore, in (i), for $\tau \in (0,\infty)$, h_{12} and h_{13} tend to be high (resp. low) together, if the proximity region $N(X_1,\tau,c)$ is large (resp. small), since $2h_{12} = \mathbb{I}\big(X_2 \in N(X_1,\tau,c)\big) + \mathbb{I}\big(X_1 \in N(X_2,\tau,c)\big)$ is the number of arcs between X_1 and X_2 in the ICDs. Hence the asymptotic variance of $\rho_{[i]}(\tau,c)$, $\mathbf{Cov}[h_{12}, h_{13}] = 4\,\nu(\tau,c) > 0$. The same holds for end intervals in (ii) as well. □

For middle intervals, the asymptotic variance in Theorem 4.2 can be written as

$$\mathbf{Cov}[h_{12}, h_{13}] \;=\; \frac{1}{4}\left(P_{2N} + 2\,P_{NG} + P_{2G}\right) - p_a(\tau,c)^2\,,$$

where

$$P_{2N} \;:=\; P\Big(\{X_2, X_3\} \subset N(X_1,\tau,c)\Big)\,,$$

$$P_{NG} \;:=\; P\Big(X_2 \in N(X_1,\tau,c),\, X_3 \in \Gamma_1(X_1,\tau,c)\Big)\,,$$

and

$$P_{2G} := P\Big(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, c)\Big).$$

Similarly, for end intervals

$$\mathbf{Cov}[h_{12}, h_{13}] = \frac{1}{4}\left(P_{2N,e} + 2\,P_{NG,e} + P_{2G,e}\right) - p_a^e(\tau, c)^2,$$

where

$$P_{2N,e} := P\Big(\{X_2, X_3\} \subset N_e(X_1, \tau)\Big),$$
$$P_{NG,e} := P\Big(X_2 \in N_e(X_1, \tau),\, X_3 \in \Gamma_{1,e}(X_1, \tau)\Big),$$
and
$$P_{2G,e} := P\Big(\{X_2, X_3\} \subset \Gamma_{1,e}(X_1, \tau)\Big),$$

with $\Gamma_{1,e}(x, \tau)$ being the Γ_1-region corresponding to $N_e(x, \tau)$ in the end intervals. Furthermore, for $\tau = \infty$, $\mathbf{E}\Big[\rho_{[i]}(\infty, c)\Big] = \mathbf{E}[h_{12}] = \mu(\infty, c) = P\big(X_2 \in N(X_1, \infty, c)\big) = P(X_2 \in \mathcal{I}_i) = 1$ and $\nu(\infty, c) = 0$. Thus, $\rho_{[i]}(\tau = \infty, c) = 1$ a.s. and the CLT result does not hold for $\tau = \infty$.

4.1. Distribution of the arc density of $\mathbf{D}_{n,2}(\tau, c)$

In this section, we find the distribution of the arc density of $\mathbf{D}_{n,2}(\tau, c)$ for $\tau > 0$ and $c \in (0, 1)$. For the special case of $m = 2$, we have $\mathcal{Y}_2 = \{y_1, y_2\}$ and $\delta_1 = y_1 < y_2 = \delta_2$, and only one middle interval and the two end intervals are empty. By Theorems 4.1 and 4.2, the asymptotic distribution of any $\rho_{[i]}(\tau, c)$ for the middle intervals with $m > 2$ will be same as the asymptotic distribution of density of the CS-ICD based on $\mathcal{U}(0, 1)$ data.

For $\tau \in (0, 1]$, the proximity region is defined as in Equation (4.2) and for $\tau > 1$, the proximity region is as in Equation (4.3).

Theorem 4.3. *For $\tau \in (0, \infty)$, we have $\sqrt{n}\left[\rho_{n,2}(\tau, c) - p_a(\tau, c)\right] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(\tau, c)\big)$, as $n \to \infty$, where*

$$p_a(\tau, c) = \begin{cases} \frac{\tau}{2} & \text{if } 0 < \tau < 1, \\ \frac{\tau\,(1+2\,c\,(\tau-1)(1-c))}{2\,(c\,\tau-c+1)(\tau+c-c\,\tau)} & \text{if } \tau > 1, \end{cases} \tag{4.4}$$

and

$$4\,\nu_1(\tau, c) = \kappa_1(\tau, c)\,\mathbb{I}(0 < \tau < 1) + \kappa_2(\tau, c)\,\mathbb{I}(\tau > 1)$$

where

$$\kappa_1(\tau, c) = \frac{\tau^2\left(c^2\,\tau^3 - 3\,c^2\,\tau^2 - c\,\tau^3 + 2\,c^2\,\tau + 3\,c\,\tau^2 - c^2 - 2\,c\,\tau - \tau^2 + c + \tau\right)}{3\,(c\,\tau - c + 1)\,(c + \tau - c\,\tau)},$$

and

$$\begin{aligned}\kappa_2(\tau,c) \;=\; \Big[\, c\,(1-c)\,\Big(&2\,c^4\tau^5 - 7\,c^4\tau^4 - 4\,c^3\tau^5 + 8\,c^4\tau^3 + 14\,c^3\tau^4 + 3\,c^2\tau^5 \\ &- 2\,c^4\tau^2 - 16\,c^3\tau^3 - 7\,c^2\tau^4 - c\,\tau^5 - 2\,c^4\tau + 4\,c^3\tau^2 + 12\,c^2\tau^3 \\ &+ c^4 + 4\,c^3\tau - 6\,c^2\tau^2 - 4\,c\,\tau^3 - 2\,c^3 - 3\,c^2\tau + 4\,c\,\tau^2 \\ &+ c^2 + c\,\tau - \tau^2\Big)\Big] \Big/ \Big[3\,\big(c\,\tau - c + 1\big)^3\big(c\,\tau - c - \tau\big)^3\Big]\,.\end{aligned}$$

The proof is provided in the Appendix. Notice that $p_a(\tau, c)$ is independent of the centrality parameter c for $\tau \in (0, 1]$. See Figure 2 for the surface plots of $p_a(\tau, c)$ and $4\,\nu(\tau, c)$. Observe that $\lim_{\tau\to\infty} \nu(\tau, c) = 0$, so the CLT result fails for $\tau = \infty$ and $\lim_{\tau\to 0} \nu_1(\tau, c) = 0$, but CS-ICD is not defined for $\tau = 0$.

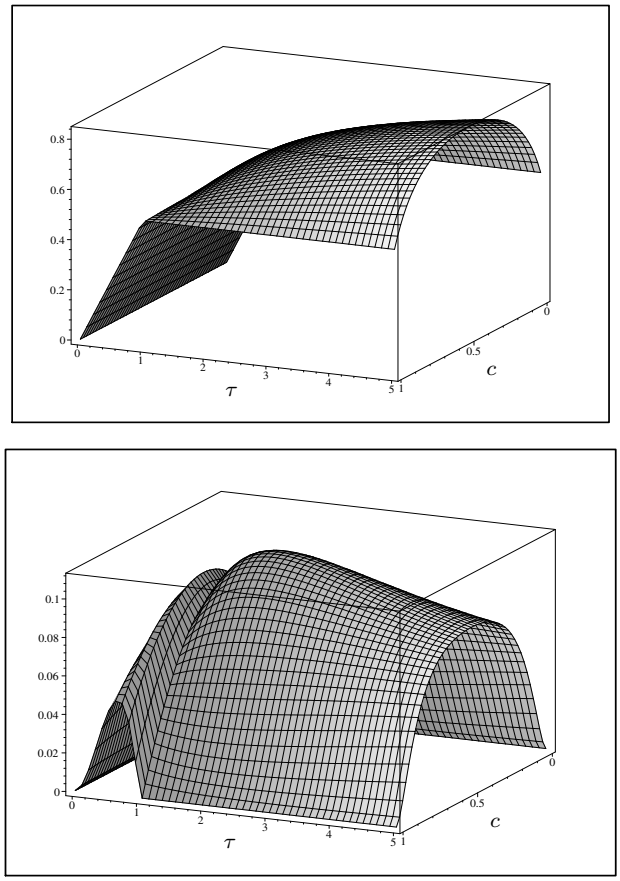

Figure 2: The surface plots of the asymptotic mean $p_a(\tau, c)$ (top) and the variance $4\,\nu(\tau, c)$ (bottom) as a function of τ and c for $\tau \in (0, 5]$ and $c \in (0, 1)$, respectively.

The sharpest rate of convergence in Theorem 4.3 is $K\frac{p_a(\tau,c)}{\sqrt{n\,\nu(\tau,c)^3}}$ (the explicit form not presented) and is minimized at $\tau \approx 1.55$ and $c = 1/2$ which is found by setting

the first order partial derivatives of this convergence rate with respect to τ and c to zero and solving for τ and c numerically and verified by the surface plot. Surface plots for the convergence rates $f^c_{CS}(\tau, c)$ and $f^c_{PE}(\tau, c)$ are presented in Figure 3. At optimal parameters within their entire ranges, the convergence rate for the arc density of CS-ICDs is faster than that of the PE-PCDs.

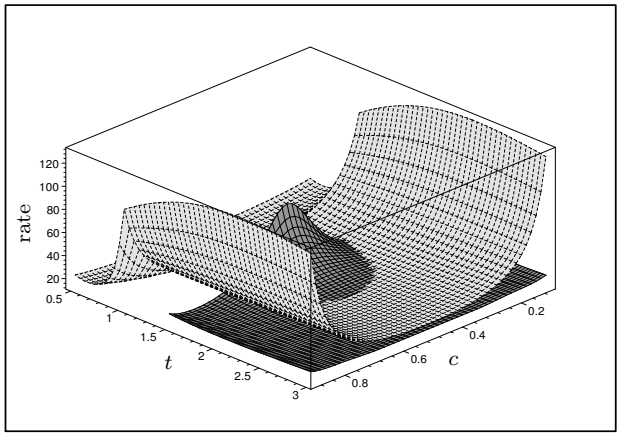

Figure 3: The surface plots of the rates of convergence to normality for PE- and CS-ICDs for the entire ranges of expansion parameter, t, and centrality parameter, c. The rate for CS-ICD is plotted in light gray, while that for PE-PCDs is plotted in dark gray.

Each of the following special cases follows as a corollary of Theorem 4.3.

Special Cases:

Case (i): $\tau > 0$ and $c = 1/2$.

As $n \to \infty$, we have $\sqrt{n}\left[\rho_{n,2}(\tau, 1/2) - p_a(\tau, 1/2)\right] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(\tau, 1/2)\big)$, where

$$p_a(\tau, 1/2) = \begin{cases} \tau/2 & \text{if } 0 < \tau < 1\,, \\ \tau/(\tau+1) & \text{if } \tau > 1\,, \end{cases} \tag{4.5}$$

and

$$4\,\nu(\tau, 1/2) = \begin{cases} \frac{\tau^2\,(1+2\,\tau-\tau^2-\tau^3)}{3\,(\tau+1)^2} & \text{if } 0 < \tau \leq 1\,, \\ \frac{2\,\tau-1}{3\,(\tau+1)^2} & \text{if } \tau > 1\,. \end{cases} \tag{4.6}$$

Case (ii): $\tau = 1$ and $c \in (0, 1)$.

As $n \to \infty$, we have $\sqrt{n}\left[\rho_{n,2}(1, c) - p_a(1, c)\right] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(1, c)\big)$, where $p_a(1, c) = 1/2$ and $4\,\nu(1, c) = c\,(1-c)/3$.

Case (iii): $\tau = 1$ and $c = 1/2$:

As $n \to \infty$, we have $\sqrt{n}\left[\rho_n(1,1/2) - p_a(1,1/2)\right] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(1,1/2)\big)$, where $p_a(1,1/2) = 1/2$ and $4\,\nu(1,1/2) = 1/12$.

4.2. Arc density in the case of end intervals (for $\mathcal{U}\left(\delta_1, y_{(1)}\right)$ or $\mathcal{U}\left(y_{(m)}, \delta_2\right)$ data)

With $m \geq 1$, we have the end intervals, $\mathcal{I}_1 = \big(\delta_1, y_{(1)}\big)$ and $\mathcal{I}_{m+1} = \big(y_{(m)}, \delta_2\big)$. In these intervals, the proximity and Γ_1-regions are only dependent on x and τ (but not on c). Let $D_{[i]}(1,c)$ be the subdigraph of the CS-ICD based on uniform data in (δ_1, δ_2) where $\delta_1 < \delta_2$ and $\mathcal{Y}_m$ be a set of m distinct $\mathcal{Y}$ points in (δ_1, δ_2). By scale invariance of Theorem 4.1, we can assume that each of the end intervals is $(0,1)$.

For $\tau \in (0,1]$ and x in the right end interval, the proximity region is

$$N_e(x,\tau) = \begin{cases} \big((1-\tau)\,x,\, (1+\tau)\,x\big) & \text{if } x \in \big(0, 1/(1+\tau)\big)\,, \\ \big((1-\tau)\,x, 1\big) & \text{if } x \in \big(1/(1+\tau), 1\big)\,, \end{cases} \tag{4.7}$$

and for $\tau > 1$ and x in the right end interval, the proximity region is

$$N_e(x,\tau) = \begin{cases} \big(0, (1+\tau)\,x\big) & \text{if } x \in \big(0, 1/(1+\tau)\big)\,, \\ (0,1) & \text{if } x \in \big(1/(1+\tau), 1\big)\,. \end{cases} \tag{4.8}$$

Theorem 4.4. *For $i \in \{1, m+1\}$ and $\tau \in (0,\infty)$, as $n_i \to \infty$, we have $\sqrt{n_i}\left[\rho_{[i]}(\tau) - p_a^e(\tau)\right] \xrightarrow{\mathcal{L}} \mathbb{N}(0, 4\,\nu_e(\tau))$, where*

$$p_a^e(\tau, c) = \begin{cases} \frac{\tau\,(\tau+2)}{2\,(\tau+1)} & \text{if } 0 < \tau < 1\,, \\ \frac{1+2\,\tau}{2\,(\tau+1)} & \text{if } \tau > 1\,, \end{cases} \tag{4.9}$$

and

$$4\,\nu_e(\tau) = \begin{cases} \frac{\tau^2\left(4\,\tau+4-2\,\tau^4-4\,\tau^3-\tau^2\right)}{3\,(\tau+1)^3} & \text{if } 0 < \tau < 1\,, \\ \frac{\tau^2}{3\,(\tau+1)^3} & \text{if } \tau > 1\,. \end{cases} \tag{4.10}$$

The proof is provided in the Appendix. See Figure 4 for the plots of $p_a^e(\tau)$ and $4\,\nu_e(\tau)$ with $\tau \in (0, 10]$. Notice that $\lim_{\tau\to\infty} \nu_e(\tau) = 0$, so the CLT result fails for $\tau = \infty$ and $\lim_{\tau\to 0} \nu_e(\tau) = 0$. The sharpest rate of convergence in Theorem 4.4

is $K\frac{p_a^e(\tau)}{\sqrt{n_i\,\nu_e(\tau)^3}}$ (explicit form not presented) for $i \in \{1, m+1\}$ and is minimized at $\tau \approx 0.58$ which is found numerically as before and verified by the plot of $p_a^e(\tau)/\sqrt{\nu_e(\tau)^3}$.

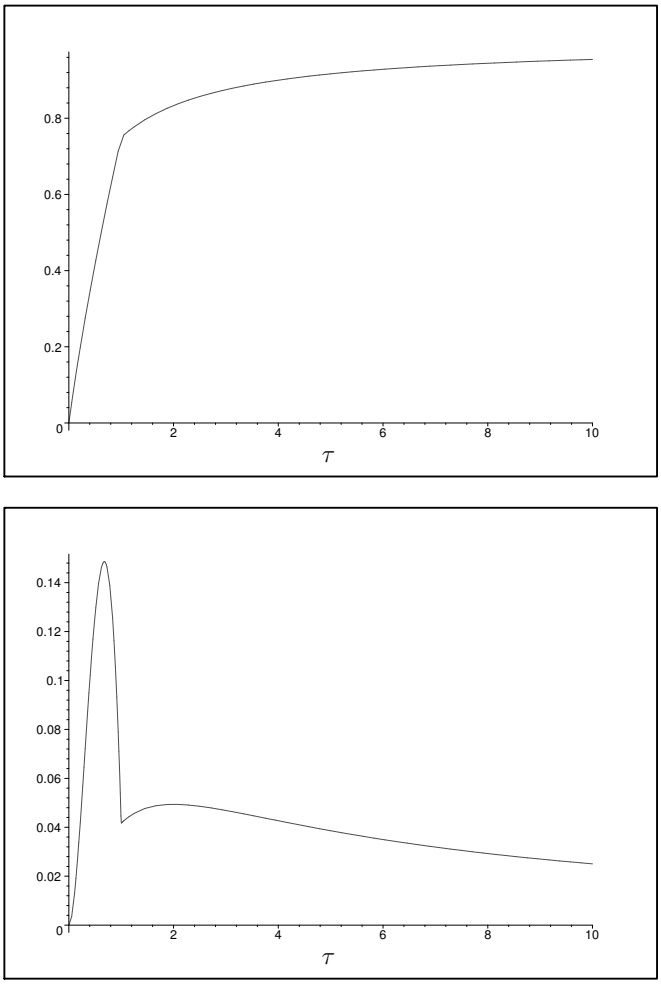

Figure 4: The plots of the asymptotic mean $p_a^e(\tau, c)$ (top) and the variance $4\,\nu_e(\tau)$ (bottom) for the end intervals as a function of τ for $\tau \in (0, 10]$.

The distribution for the special case of $\tau = 1$ follows as a corollary to Theorem 4.4: For x in the right end interval, $N_e(x, 1) = \big(0, \min(1, 2x)\big)$ for $x \in (0, 1)$. For $i \in \{1, m+1\}$ (i.e., in the end intervals), as $n_i \to \infty$, we have $\sqrt{n_i}\Big[\rho_{[i]}(1) - p_a^e(1)\Big] \xrightarrow{\mathcal{L}} \mathbb{N}(0, 4\,\nu_e(1))$, where $p_a^e(1) = 3/4$ and $4\,\nu_e(1) = 1/24$.

Remark 4.1. Multiple Interval Case: In the case of $m > 2$, we have two versions of arc density. One is defined as the (adjusted) arc density as in Equation

(3.2). The asymptotic distribution of $\rho_{n,m}(\tau, c)$ is the same as given in Theorem 11 of [Ceyhan, 2012]. As for the other one, if we consider the entire data set $\mathcal{X}_n$, then we have n vertices. So we can also consider the arc density as $\rho_{n,m}(\tau, c) = |\mathcal{A}| / \big(n\,(n-1)\big)$. The asymptotic distribution for $\rho_{n,m}(\tau, c)$ is as in Theorem 12 of [Ceyhan, 2012].

Remark 4.2. Use of Arc Density for Testing Multi-Class Spatial Interactions: Arc density of CS-ICDs can be employed in testing two-class spatial point patterns of one-dimensional data, as was done in [Ceyhan *et al.*, 2007] for two-dimensional data. In particular, for two classes of points, $\mathcal{X}$ and $\mathcal{Y}$, whose support is in a compact interval in $\mathbb{R}$, we assume some form of *complete spatial randomness* of $\mathcal{X}$ points (i.e., $\mathcal{X}$ points having uniform distribution in the support interval irrespective of the distribution of the $\mathcal{Y}$ points) as our null hypothesis. The alternative patterns of interest are *segregation* of $\mathcal{X}$ from $\mathcal{Y}$ points or *association* of $\mathcal{X}$ points with $\mathcal{Y}$ points. Association is the pattern in which the points from the two classes tend to occur close to each other, while segregation is the pattern in which the points from the same class tend to cluster together. In our context, association implies that $\mathcal{X}$ points are clustered around $\mathcal{Y}$ points, while segregation implies that $\mathcal{X}$ points are clustered away from the $\mathcal{Y}$ points. The use of arc density of CS-ICDs requires number of $\mathcal{X}$ points to be much larger compared to the number of $\mathcal{Y}$ points. Furthermore, the power comparisons are possible for data from large families of distributions to obtain the optimal parameters against segregation and association alternatives.

Remark 4.3. Extension of Central Similarity Proximity Regions to Higher Dimensions: In this article, we discuss the construction of CS-ICDs for one-dimensional data and asymptotic distribution of their arc density (for uniform data). The CS-ICDs in this article can be viewed as the one-dimensional version of the PCDs introduced in [Ceyhan *et al.*, 2007], which also contains the extension to higher dimensions.

5. TESTING UNIFORMITY WITH THE ARC DENSITY OF CS-ICDS

We can employ the arc density of the CS-PCDs for testing uniformity based on its asymptotic normality. Let $X_i \overset{iid}{\sim} F$ for $i = 1, 2, ..., n$ where F has a bounded interval support (a, b) in $\mathbb{R}$. Then our null hypothesis is $H_o\colon F = \mathcal{U}(a, b)$. For testing this hypothesis, we use the arc density $\rho_{n,2}(\tau, c)$ whose asymptotic distribution is provided in Theorem 4.3 for uniform data. By the scale invariance property of the distribution of $\rho_{n,2}(\tau, c)$ (see Theorem 4.1), without loss of generality, we can assume $(a, b) = (0, 1)$. In this approach, for each choice of (τ, c),

we compute the arc density, $\rho_{n,2}(\tau, c)$, and standardize it as

$$R_n(\tau, c) := \sqrt{n}\big(\rho_{n,2}(\tau, c) - p_a(\tau, c)\big)\big/\sqrt{4\nu(\tau, c)}$$

and use the standardized version as our test statistic. The critical values for the one- and two-sided alternatives are based on the standard normal distribution, e.g., the level α critical value for the left-sided alternative is z_α, the $\alpha\times 100^{\text{th}}$ percentile of the standard normal distribution.

For comparative purposes, we employ the arc density of PE-ICDs introduced by [Ceyhan, 2012]. In particular, the defining regions for the PE-ICD are

$$N_{PE}(x, r, c) = \begin{cases} (0, r\,x) \cap (0,1) & \text{if } x \in (0, c)\,, \\ \big(1 - r\,(1-x), 1\big) \cap (0,1) & \text{if } x \in (c, 1)\,. \end{cases} \tag{5.1}$$

The asymptotic distribution of the arc density of PE-ICDs for uniform data was provided in [Ceyhan, 2012]. Furthermore, we also employ Kolmogorov–Smirnov (KS) test for uniform distribution and Neyman's smooth test of uniformity since the former is one of the most commonly used tests and latter is recommended for a large family of alternatives for testing uniformity by [Marhuenda *et al.*, 2005].

5.1. Empirical size analysis

We first perform an extensive size analysis to determine for which parameter values the arc density of the ICDs have the appropriate size at specific sample sizes in testing $H_o\colon F = \mathcal{U}(0,1)$. For this purpose, we partition the ranges of τ and c for the CS-ICD as follows. We take $c = .01, .02, ..., .99$ and $\tau = .01, .02, ..., 10.00$, and consider each (τ, c) combination on a 99×1000 grid with $n = 20, 50, 100$. Similarly, we partition the ranges of r and c for the PE-ICD as follows. We use the same partition above for c and take $r = 1.01, ..., 10.00$, and consider each (r, c) combination on a 99×900 grid with $n = 20, 50, 100$. For each (τ, c) (and (r, c)) combination, we generate $N_{mc} = 10000$ samples of size n iid from $\mathcal{U}(0,1)$ distribution. Then for each sample generated, we compute the arc densities and use their standardized versions as approximate test statistics. Empirical size is estimated as the frequency of number of times p-value is significant at $\alpha = .05$ level divided by $N_{mc} = 10000$. We also estimate the empirical sizes for KS and Neyman's smooth tests with $n = 20$ and $N_{mc} = 10000$. With $N_{mc} = 10000$, empirical size estimates larger than .0536 (resp. less than .0464) are deemed liberal (resp. conservative). These bounds are determined using a binomial test for the proportions with $n = 10000$ trials at .05 level of significance. The size estimates for KS and Neyman's smooth test are found to be about the nominal level (i.e., between .0464 and .0536).

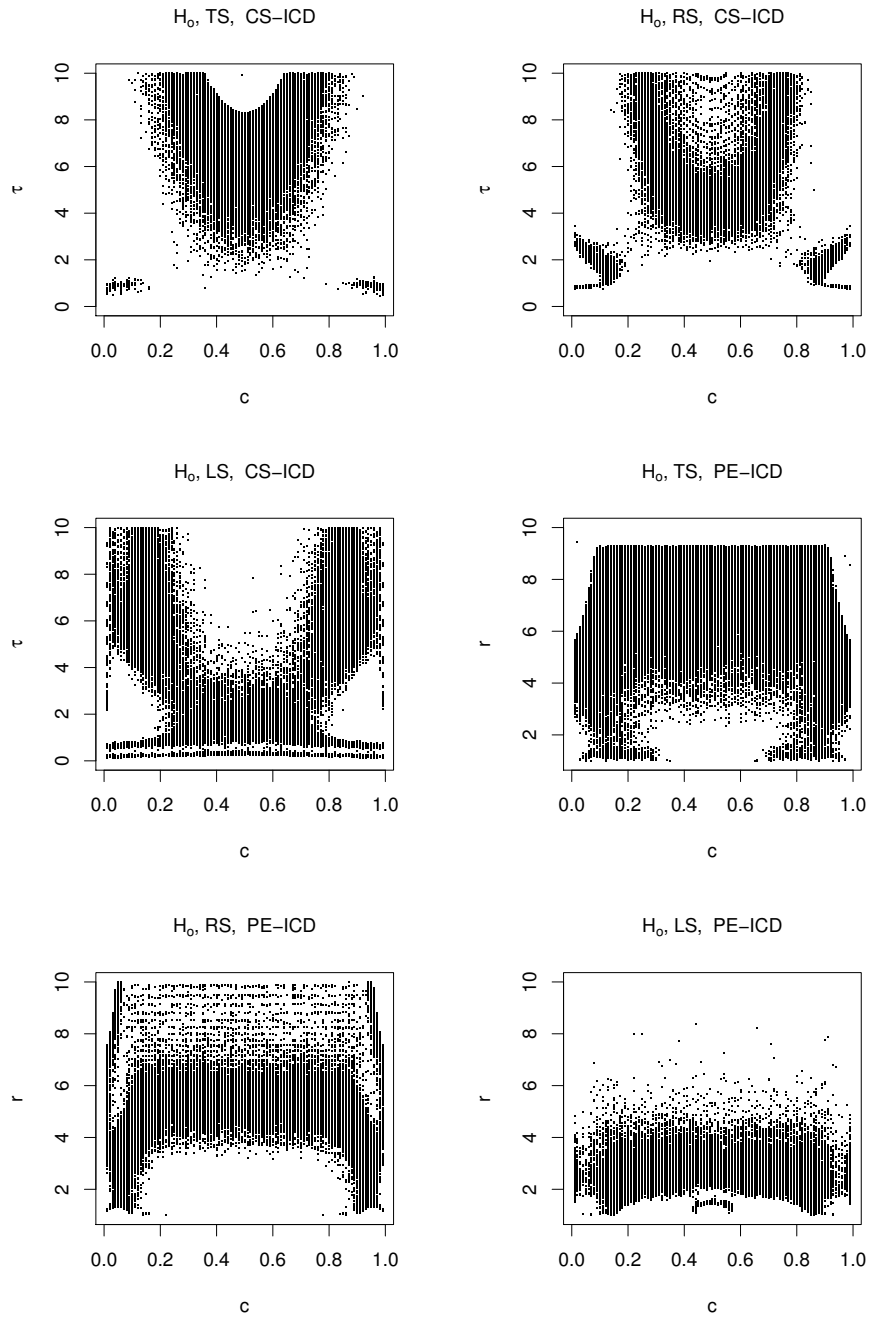

Figure 5: Two-level (i.e., black and white) image plots for the empirical size estimates for the arc density of CS-ICD and PE-PCD based on $n = 20$ and $N_{mc} = 10000$ the two-sided (TS), right-sided (RS) and left-sided (LS) alternatives. The empirical sizes not significantly different from 0.05 are represented with black dots, and others are represented with white dots. For CS-ICD, we take $\tau = .01, .02, ..., 10.00$ and for PE-ICD, we take $r = 1.01, 1.02, ..., 10.00$ and for both ICDs, we take $c = .01, .02, ..., .99$ with $N_{mc} = 10000$ Monte Carlo replications.

We present the empirical size estimates in two-level image plots (with empirical sizes not significantly different from 0.05 in black, and others in white) for the two-, right- and left-sided alternatives for the CS-ICD with $n = 20$, $c = .01, .02, ..., .99$ and $\tau = .01, .02, ..., 10.00$ and for the PE-ICD with $n = 20$, $c = .01, .02, ..., .99$ and $r = 1.01, 1.02, ..., 10.00$ in Figure 5. The size estimates for $n = 50$ and 100 have the similar trend with sizes closer to nominal level for more parameter combinations (not presented). Notice that there is symmetry in size estimates around $c = 1/2$. For the one-sided alternatives, the regions at which size estimates are appropriate are somewhat complementary, in the sense that, the sizes are appropriate for the parameter combinations in one region for left-sided alternative and mostly in its complement for the right-sided alternative for each ICD family. Notice also that arc density of PE-ICD has appropriate size for the two-sided alternative for more parameter combinations, and arc density of CS-ICD has appropriate size for the left-sided alternative for more parameter combinations.

5.2. Empirical power analysis

We perform power analysis to determine which tests have higher power under various alternatives against uniformity. For the alternatives, we use three families of non-uniform distributions with support in $(0, 1)$ which are proposed by [Stephens, 1974]:

$$\begin{aligned}
&\textbf{(I)} \quad F_1(x,\delta) = \left(1-(1-x)^{\delta}\right) \mathbb{I}(0 \le x < 1) + \mathbb{I}(x \ge 1)\,,\\
&\textbf{(II)} \quad F_2(x,\delta) = \left(2^{\delta-1}x^{\delta}\right) \mathbb{I}(0 \le x < 1/2) + \left(1-2^{\delta-1}(1-x)^{\delta}\right) \mathbb{I}(1/2 \le x < 1)\\
&\qquad\qquad + \mathbb{I}(x \ge 1)\,,\\
&\textbf{(III)} \quad F_3(x,\delta) = \left(1/2-2^{\delta-1}(1/2-x)^{\delta}\right) \mathbb{I}(0 \le x < 1/2)\\
&\qquad\qquad + \left(1/2+2^{\delta-1}(x-1/2)^{\delta}\right) \mathbb{I}(1/2 \le x < 1) + \mathbb{I}(x \ge 1)\,.
\end{aligned}$$

That is,

$$\begin{aligned}
H_a^I &: \quad F = F_1(x,\delta) \ \text{ with } \delta > 1\,,\\
H_a^{II} &: \quad F = F_2(x,\delta) \ \text{ with } \delta > 1\,,\\
H_a^{III} &: \quad F = F_3(x,\delta) \ \text{ with } \delta > 1\,.
\end{aligned}$$

The corresponding pdfs for the distributions in the alternatives are

$$\begin{aligned}
&\textbf{(I)} \quad f_1(x) = \left(\delta(1-x)^{\delta-1}\right) \mathbb{I}(0<x<1)\,,\\
&\textbf{(II)} \quad f_2(x) = \left(\delta\, 2^{\delta-1}x^{\delta-1}\right) \mathbb{I}(0<x\le 1/2) + \left(\delta\, 2^{\delta-1}(1-x)^{\delta-1}\right) \mathbb{I}(1/2<x<1)\,,\\
&\textbf{(III)} \quad f_3(x) = \left(\delta(1-2x)^{\delta-1}\right) \mathbb{I}(0<x\le 1/2) + \left(\delta(2x-1)^{\delta-1}\right) \mathbb{I}(1/2<x<1)\,.
\end{aligned}$$

See Figure 6 for sample plots of the pdfs with various parameters under types I–III alternatives. Notice that in all the alternatives, $\delta = 1$ corresponds to $\mathcal{U}(0,1)$ distribution. Under type I alternatives, with increasing $\delta > 1$, the pdf of the distribution is more clustered around 0 and less clustered around 1; under type II alternatives, with increasing $\delta > 1$, the pdf of the distribution gets more clustered around 1/2 (and less clustered around the end points, 0 and 1); and under type III alternatives, with increasing $\delta > 1$, the pdf of the distribution is more clustered around the end points, 0 and 1, and less clustered around 1/2. Under the type II and III alternatives, the pdfs are symmetric around 1/2.

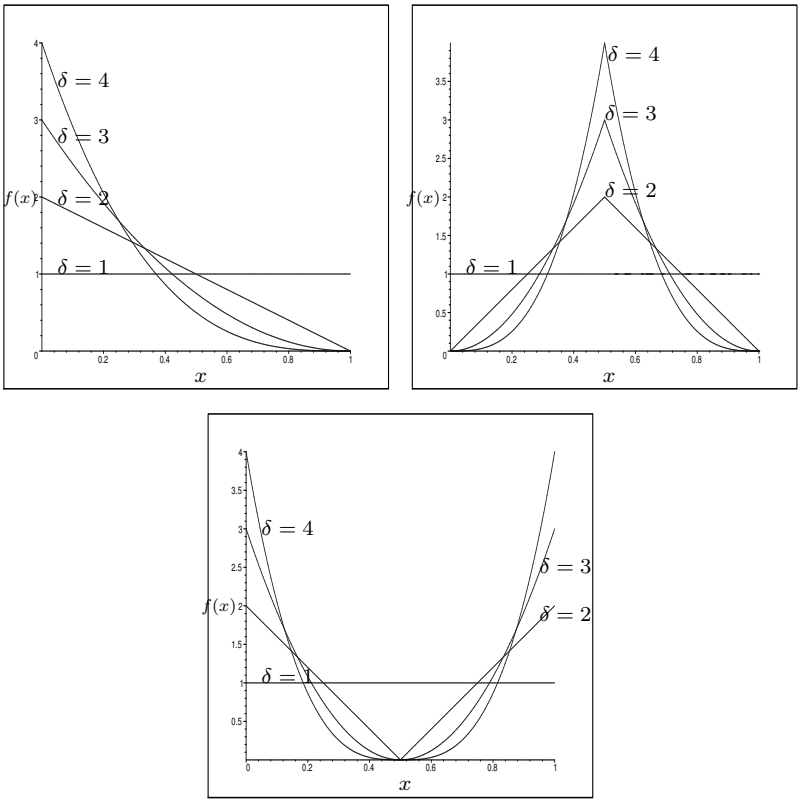

Figure 6: Sample plots for the pdfs of the alternative types I (top left), II (top right), and III (bottom) with $\delta = 2, 3, 4$. The horizontal line at 1 indicates the pdf for $\mathcal{U}(0,1)$ distribution (with $\delta = 1$).

We generate $n = 20$ points according to the specified alternatives with various parameters. In particular, for each of H_a^I–H_a^{III}, we consider $\delta = 2, 3, 4$. With CS-ICDs, we use (τ, c) for $\tau = .01, .02, ..., 10.00$ and $c = .01, .02...., .99$ and with PE-ICDs, we use (r, c) for $r = 1.01, .02, ..., 10.00$ and $c = .01, .02...., .99$. With CS-ICDs, for each (τ, c) and δ combination, and with PE-ICDs, for each (r, c) and δ combination, we replicate the sample generation $N_{mc} = 10000$ times. We compute the power using the asymptotic critical values based on the normal approximation.

Table 1: The maximum power estimates for the one-sided alternatives unadjusted (the first entry) and adjusted (the second entry) for size. In the size adjusted version, only the parameter combinations at which the tests have appropriate level are kept. RS: right-sided, LS: left-sided alternatives.

alternative	CS-ICD		PE-ICD	
	RS	LS	RS	LS
H_a^I	0.86, .73	.65, .30	.93, .75	.41, .41
H_a^{II}	0.93, .90	.29, .00	.91, .90	.60, .00
H_a^{III}	0.41, .18	.81, .81	.27, .13	.81, .81

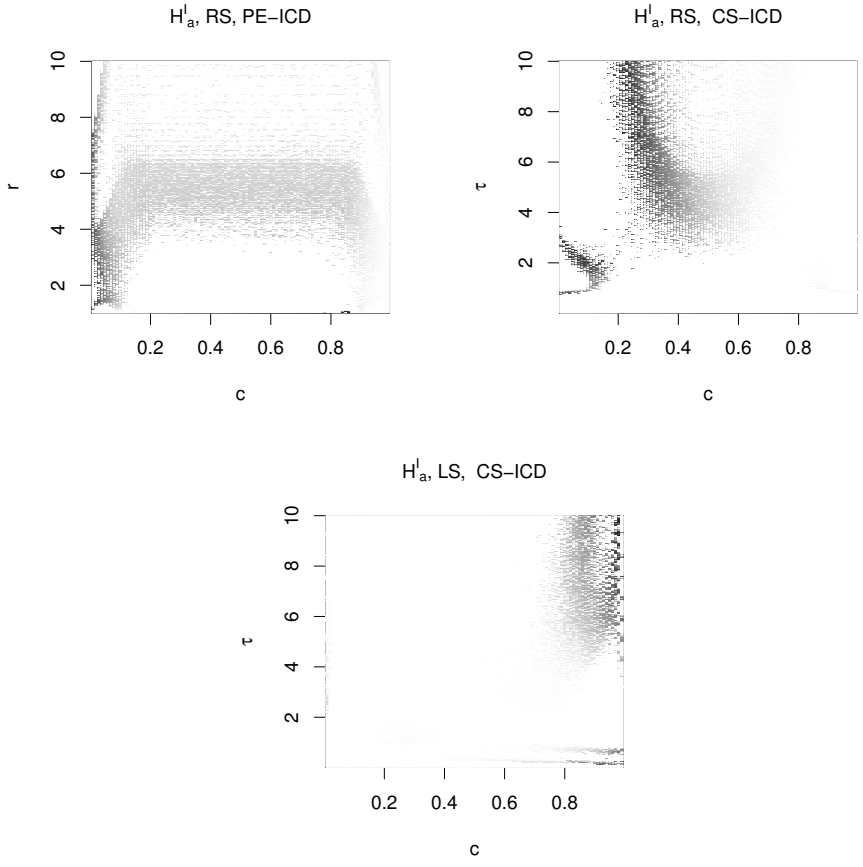

Figure 7: Image plots for the power estimates for PE-ICD with $r \in (1, 10)$ and $c \in (0, 1)$ and CS-ICD with $\tau \in (0, 10)$ and $c \in (0, 1)$ under $H_a^I : \delta = 2$, with $n = 20$, $N_{mc} = 10000$. The intensity of the gray level increases as the power increases, and the same darkness scale is used for each image plot. RS stands for right-sided, LS stands for left-sided alternatives.

We only keep the parameter combinations ((r, c) for PE-ICDs and (τ, c) for CS-ICDs) at which the tests have the appropriate level (of .05), i.e., if the test is conservative or liberal for the one-sided version in question, we ignore that parameter combination in our power estimation, as they would yield unreliable results. We call this procedure the "size adjustment" for the power estimation. The maximum values of the power estimates under the one-sided alternatives adjusted and unadjusted for the correct size are provided in Table 1. Observe that the size adjustment has a substantial effect on the highest power values (and tends to reduce the highest power estimates). Furthermore, under the alternatives H_a^I and H_a^{II}, the ICDs yield higher power for the right-sided alternative, while under H_a^{III} the ICDs yield higher power for the left-sided alternative. In particular, PE-ICDs have high power for the right-sided alternative under H_a^I and H_a^{II}, and left-sided alternative under H_a^{III} with virtually zero power for the opposite direction under these alternatives. On the other hand, CS-ICDs tend to have a

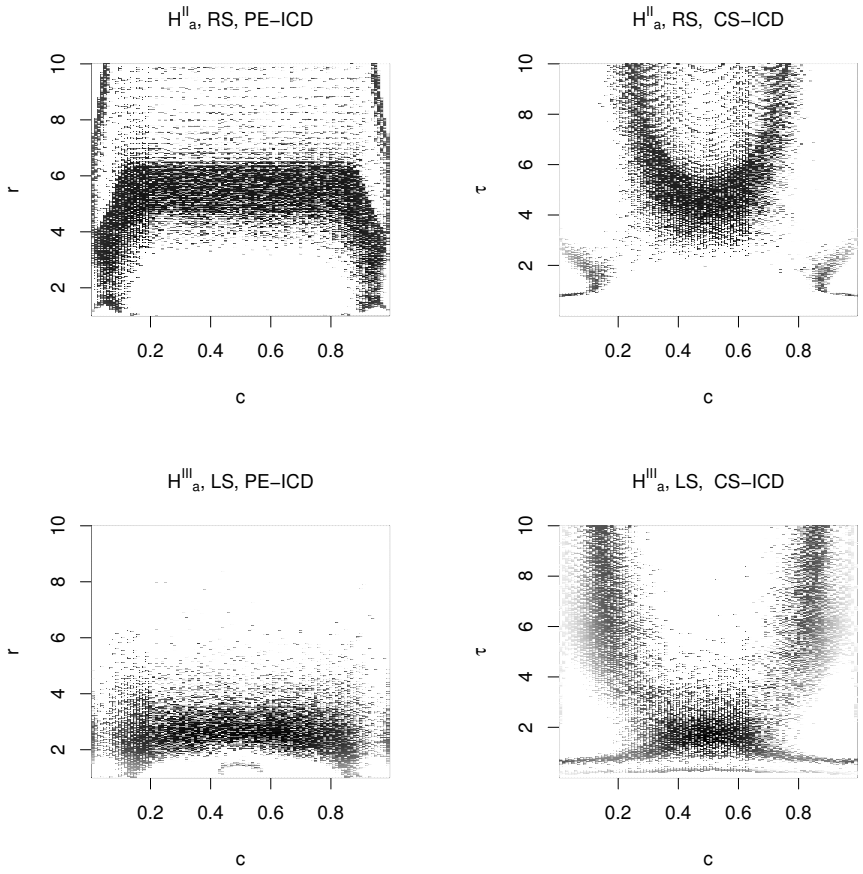

Figure 8: Image plots for the power estimates for PE-ICD with $r \in (1, 10)$ and $c \in (0, 1)$ and CS-ICD with $\tau \in (0, 10)$ and $c \in (0, 1)$ under H_a^{II} and H_a^{III} with $\delta = 2$, $n = 20$, $N_{mc} = 10000$. The gray level intensity and alternative labeling are as in Figure 7.

similar trend, but the power estimates for the direction of the one-sided version depends on the parameters. That is, e.g., under H_a^I, CS-PCD has high power estimates for the right-sided alternative at some (τ, c) combinations, and for the left-sided alternative at some other (τ, c) values. The gray-scale image plots of the power estimates under H_a^I are presented in Figure 7 and under H_a^{II} and H_a^{III} in Figure 8 (with the higher power estimates are represented with darker gray level). Notice that the power estimates are symmetric around $c = 1/2$ under H_a^{II} and H_a^{III}, which is in agreement with the symmetry in the corresponding pdfs (around $c = 1/2$).

The maximum power estimates and at which parameters of the ICDs they occur are presented in Table 2. We also plot the histograms of the power estimates (normalized to have unit area) under the alternatives in Figure 9. Under H_a^I, although the maximum power estimate for the right-sided alternative is attained by PE-ICD test at $(r, c) = (1.02, .78)$, the CS-ICD test tends to have higher power estimates. Among the competitors, the power estimate is .50 for Neyman's smooth test and .82 for KS test (with the right-sided alternative), and the ICD tests have lower power compared to KS-test. Likewise, under H_a^{II}, although the maximum power estimate for the right-sided alternative is attained by CS-ICD at $(\tau, c) = (1.96, .49)$, the PE-ICD test tends to have higher power estimates. Among the competitors, the power estimate is .39 for Neyman's test and .14 for KS test (with the right-sided alternative), and the ICD tests have higher power compared to Neyman's test. Finally, under H_a^{III}, the PE-ICD test tends to have higher power estimates. Among the competitors, the power estimate is .59

Table 2: The maximum power estimates and the parameter combinations at which they occur. RS: right-sided, LS: left-sided alternatives and $\widehat{\beta}$ stands for empirical power estimates.

For CS-ICDs						
	H_a^I		H_a^{II}		H_a^{III}	
	RS	LS	RS	LS	RS	LS
$\widehat{\beta}$	0.65–.73	.20–.29	.85–.90	—	.15–.18	.75–.80
τ	(7,9)	(6.5,10)	(2.75,4)	—	(2.5,3)	(1,2.5)
c	$\approx .2$	(.96,1)	(.35,.65)	—	$(0, .04) \cup (.96, 1)$	(.4,.6)

For PE-ICDs with RS alternatives			
	H_a^I	H_a^{II}	H_a^{III}
$\widehat{\beta}$	0.65–.75	.88–.89	.80–.81
r	≈ 1	≈ 3.8	≈ 2.5
c	$\approx .86$	(.2,.8)	(.33,.67)

for Neyman's test and .23 for KS test (with the left-sided alternative), and the ICD tests have higher power compared to Neyman's test for most parameter combinations.

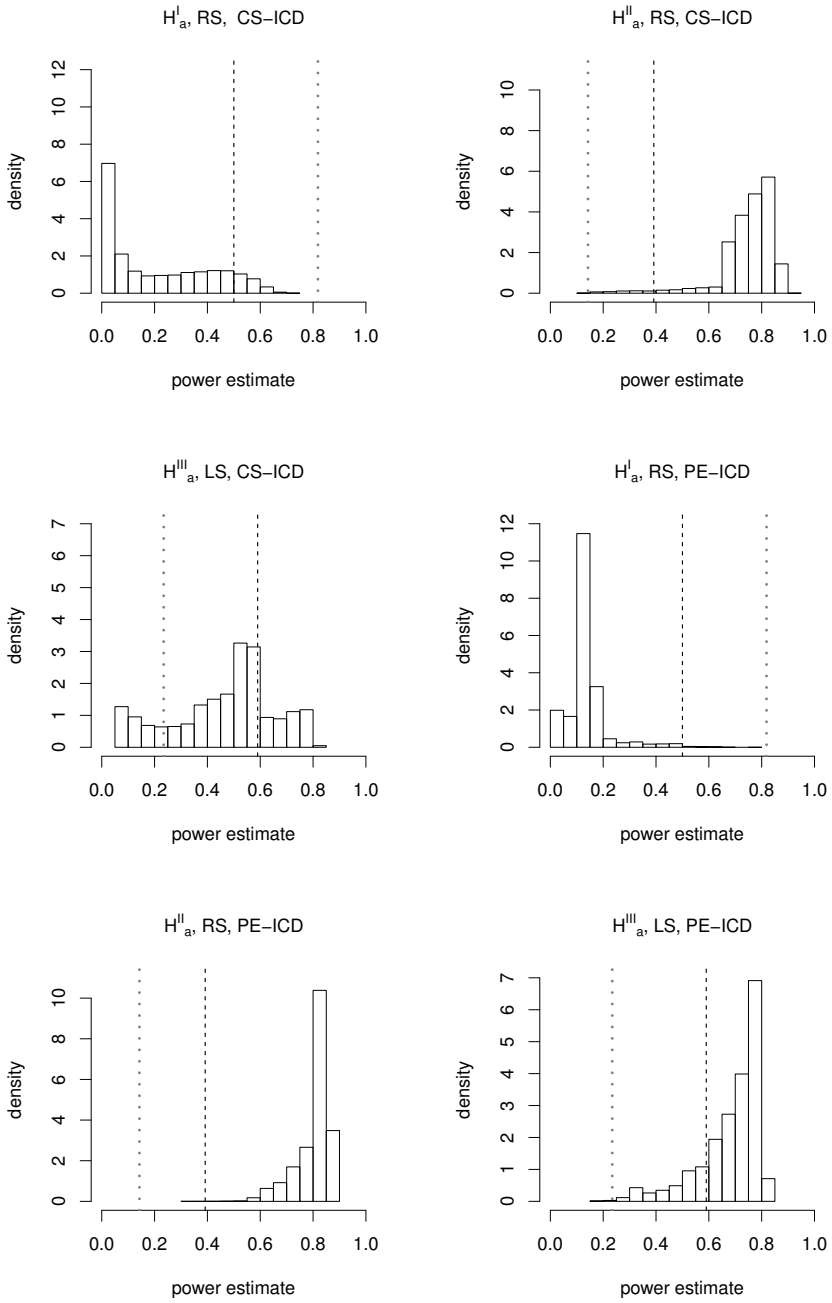

Figure 9: Histograms of the power estimates the alternatives H_a^I–H_a^{III} for the appropriate one-sided alternatives for CS-ICDs and PE-ICDs. The vertical lines are the power estimates for KS (dotted lines) and Neyman's tests (dashed lines). LS: left-sided, RS: right-sided alternatives.

5.3. Consistency of the tests based on the density of ICDs

Any reasonable test should have higher power under the alternatives as the sample size increases, and this property is reflected in the concept of *consistency*. We will prove consistency of the tests under the alternatives based on the arc density of ICDs in a general framework, and then extend the results to our alternative types for certain parameter combinations. Let $H_o: F = \mathcal{U}(0,1)$ and the alternative $H_a: F \neq \mathcal{U}(0,1)$ is parameterized by δ so that $\delta = \delta_o$ corresponds to the null hypothesis and with increasing $\delta > \delta_o$, arc probability tends to increase or decrease.

Theorem 5.1 (Consistency). *Let $\rho_{n,2}(\delta)$ be the arc density of the ICD based on data from F parameterized by δ and $p_a(\delta)$ be the corresponding arc probability. Moreover, suppose $4\nu(\delta)$ be the covariance term $\mathbf{Cov}[h_{12}, h_{13}]$. If the arc probability increases as δ increases (resp. decreases), the test against $H_a: F \neq \mathcal{U}(0,1)$ which rejects for $R_n > z_{1-\alpha}$ (resp. for $R_n < z_\alpha$) are consistent.*

Proof: Under H_o (i.e., for $\mathcal{X}_n$ being a random sample from $\mathcal{U}(0,1)$), the arc density is $\rho_{n,2}(\delta_o)$, and arc probability is $p_a(\delta_o)$ and $\mathbf{Cov}(h_{12}, h_{13})$ is $\nu(\delta_o)$. Similarly, under H_a (i.e., for $\mathcal{X}_n$ being a random sample from F) these quantities are denoted similarly with δ_o being replaced with δ. Suppose arc probability increases as $\delta > \delta_o$ increases. Then $p_a(\delta) > p_a(\delta_o)$ and the asymptotic variances $4\nu(\delta_o)/n$ and $4\nu(\delta)/n$ tends to zero as $n \to \infty$. As standardized arc density, R_n, tends to standard normal distribution or is degenerate with unit mass at $p_a(\delta_o)$ or $p_a(\delta)$ under both null and alternative hypotheses, respectively, the power under H_a tends to 1 as n goes to infinity, and hence consistency follows. The consistency for the alternative under which arc probability increases as δ decreases is similar. □

The alternatives, H_a^I–H_a^{III}, are parameterized with δ so that $\delta_o = 1$. Under H_a^I and H_a^{II} with $F = F_i(x, \delta)$ for $i = 1, 2$ the test based on CS-ICD and PE-ICD which rejects for $R_n > z_{1-\alpha}$ is consistent for most of the parameter combinations. In particular, let $\rho_{n,2}(F, \tau, c)$ be the arc density, $p_a(F, \tau, c)$ and $\nu(F, \tau, c)$ be the arc probability and $\mathbf{Cov}(h_{12}, h_{13})$ for CS-ICD with $\mathcal{X}_n$ being a random sample from F. Then under H_a^I, $p_a(F_1, \tau, c) > p_a(\mathcal{U}, \tau, c)$ for $c \leq 1/2$, since under F_1, X_i are more likely to be in $(0, 1/2)$ and hence more likely to be closer to c and hence the $N(\tau, c)$ regions are more likely to be larger which implies higher arc probability compared to the null case. Moreover, for c closer to 1 and large τ (say $\tau \geq 5$), under F_1, the $N(\tau, c)$ regions are more likely to be smaller which implies lower arc probability compared to the null case. Under H_a^{II}, $p_a(F_1, \tau, c) > p_a(\mathcal{U}, \tau, c)$ for all c away from 0 and 1 and $\tau > 0$, since under F_1, X_i are more likely to be

closer to 1/2 and hence the $N(\tau, c)$ regions are more likely to be larger which implies higher arc probability compared to the null case. Moreover, for c closer to 1 and large τ (say $\tau \geq 5$), under F_1, the $N(\tau, c)$ regions are more likely to be smaller which implies lower arc probability compared to the null case. Similarly, under H_a^{II}, $p_a(F_2, \tau, c) < p_a(\mathcal{U}, \tau, c)$ for all c away from 0 and 1 and $\tau > 0$. Hence consistency follows for these one-sided tests for such parameter combinations. In fact, with careful bookkeeping one can determine the parameter ranges for which consistency holds for each of the one-sided alternatives. For example, under $H_a^I : \delta = 2$ with $c \in (0, 1)$, for $\tau \in (0, 1)$, $p_a(F_1, \tau, c) >$ (resp. $<$) $p_a(\mathcal{U}, \tau, c)$ for $\tau >$ (resp. $<$) $\frac{2c^2-6c+3}{(2c-1)(2c-3)}$, hence consistency for the right-sided (resp. left-sided) alternative follows; likewise, for $\tau > 1$, $p_a(F_1, \tau, c) >$ (resp. $<$) $p_a(\mathcal{U}, \tau, c)$ for $\tau <$ (resp. $>$) $\frac{2c+1-\sqrt{3}}{2c-3+\sqrt{3}}$, hence consistency for the right-sided (resp. left-sided) alternative follows. The corresponding three dimensional figure to illustrate these regions of consistency for the one-sided alternatives are plotted in Figure 10 in the Appendix. Under $H_a^{II} : \delta = 2$, $p_a(F_2, \tau, c) > p_a(\mathcal{U}, \tau, c)$ for all $c \in (0, 1)$ and $\tau > 0$, hence consistency for the right-sided alternative follows; and under $H_a^{III} : \delta = 2$, $p_a(F_3, \tau, c) < p_a(\mathcal{U}, \tau, c)$ for all $c \in (0, 1)$ and $\tau > 0$, hence consistency for the left-sided alternative follows. The actual ranges of (τ, c) for the one-sided alternatives with other specific δ values can also be determined by careful calculations.

Similarly, let $\rho_{n,2}^{PE}(F, r, c)$ be the arc density, $p_a^{PE}(F, r, c)$ and $\nu_{PE}(F, r, c)$ be the arc probability and $\mathbf{Cov}(h_{12}, h_{13})$ for PE-ICD with $\mathcal{X}_n$ being a random sample from F, respectively. Then under H_a^I, $p_a^{PE}(F_1, r, c) > p_a^{PE}(\mathcal{U}, r, c)$ for c close to 0, since for F_1, X_i are more likely to be around 0 and hence the $N_{PE}(r, c)$ regions are more likely to be larger which implies higher arc probability compared to the null case. Under H_a^{II} (resp. H_a^{III}), $p_a^{PE}(F_2, r, c) >$ (resp. $<$) $p_a^{PE}(\mathcal{U}, r, c)$ for c around 1/2, since for F_2 (resp. F_3), X_i are more likely to be closer to 1/2 (resp. 0 and 1) and hence the $N_{PE}(r, c)$ regions are more likely to be larger (resp. smaller) which implies higher (resp. lower) arc probability compared to the null case. Hence consistency follows for the right-sided (resp. left-sided) tests for such parameter combinations. In fact, under $H_a^I : \delta = 2$:

- With $c \in (0, 1/2)$:
  - $p_a^{PE}(F_1, r, c) >$ (resp. $<$) $p_a^{PE}(\mathcal{U}, r, c)$, for $1 < r < 1/(1-c)$ and $r <$ (resp. $>$) $\frac{8c^3-6c^2-6c+3}{6c^4-16c^3+18c^2-12c+3}$;
  - $p_a^{PE}(F_1, r, c) >$ (resp. $<$) $p_a^{PE}(\mathcal{U}, r, c)$, for $1/(1-c) < r < 1/c$ and (r, c) is in (resp. outside) the region bounded by $r = 1/(1-c)$ and the implicit curve $3r^4c^4 - 4r^4c^3 - 8r^3c^3 + 9r^3c^2 + 6r^2c^2 - 6cr^2 + 3r - 3 = 0$;
  - $p_a^{PE}(F_1, r, c) > p_a^{PE}(\mathcal{U}, r, c)$ for $r \geq 1/c$.

  Hence consistency for the right-sided (resp. left-sided) alternative follows for these parameter values.

- With $c \in (1/2, 1)$:
  - $p_a^{PE}(F_1, r, c) > p_a^{PE}(\mathcal{U}, r, c)$ for $1 < r < 1/c$ hence consistency for the right-sided alternative follows;
  - $p_a^{PE}(F_1, r, c) >$ (resp. $<$) $p_a^{PE}(\mathcal{U}, r, c)$, for $1/c < r < 1/(1-c)$ and (r, c) is in (resp. outside) the region bounded by $c = 1$ and the implicit curve $3\,r^4c^4 - 12\,r^4c^3 + 18\,c^2r^4 - 3\,r^3c^2 - 12\,cr^4 - 6\,r^2c^2 + 6\,cr^3 + 3\,r^4 + 6\,cr^2 - 3\,r^3 - r + 1 = 0$;
  - $p_a^{PE}(F_1, r, c) > p_a^{PE}(\mathcal{U}, r, c)$ for $r \geq 1/(1-c)$.

  Hence consistency for the right-sided (resp. left-sided) alternative follows for these parameter values.

These regions of consistency for the one-sided alternatives are plotted in Figure 11 in the Appendix. Under $H_a^{II}: \delta = 2$, $p_a^{PE}(F_2, r, c) > p_a^{PE}(\mathcal{U}, r, c)$ for all $c \in (0, 1)$ and $r > 1$, hence consistency for the right-sided alternative follows; and under $H_a^{III}: \delta = 2$, $p_a^{PE}(F_3, r, c) < p_a^{PE}(\mathcal{U}, r, c)$ for all $c \in (0, 1)$ and $r > 1$ (except (r, c) inside a region that is part of $[1, 1.4] \times ([9.98, 1] \cup [0, .02])$ where the inequality is reversed), hence consistency for the left-sided (resp. right-sided) alternative follows. These regions of consistency for the one-sided alternatives are presented in Figure 12 in the Appendix. The actual ranges of (r, c) for the one-sided alternatives with other specific δ values can also be determined by careful calculations.

5.4. Extension of the methodology to test non-uniform distributions

We can modify the CS-ICD approach to test any distribution in a bounded interval in $\mathbb{R}$. Since any bounded interval (a, b) with $a < b$ can be mapped to $(0, 1)$, we can assume the support for the distribution in question to be $(0, 1)$. First we prove the below result which is instrumental for this purpose.

Proposition 5.1. *Let X_i be iid from an absolutely continuous distribution F with support $(0, 1)$ and let $\mathcal{X}_n = \{X_1, X_2, ..., X_n\}$. Define the proximity map $N_F(x, \tau, c) := F^{-1}\big(N\big(F(x), \tau, c\big)\big)$. More specifically for $\tau \in (0, 1]$,*

$$N_F(x, \tau, c) = \tag{5.2}$$

$$= \begin{cases} \left(F^{-1}\big((1-\tau)F(x)\big), F^{-1}\Big(\Big(1+\frac{(1-c)}{c}\tau\Big)F(x)\Big)\right) & \text{if } x \in \big(0, F^{-1}(c)\big), \\ \left(F^{-1}\Big(F(x) - \frac{c\,\tau}{(1-c)}\,\big(1-F(x)\big)\Big), F^{-1}\Big(F(x) + \big(1-F(x)\big)\,\tau\Big)\right) & \text{if } x \in \big(F^{-1}(c), 1\big), \end{cases}$$

and for $\tau > 1$,

(5.3) $N_F(x,\tau,c) =$

$$= \begin{cases} \left(0, F^{-1}\Big(\Big(1+\frac{(1-c)}{c}\,\tau\Big)F(x)\Big)\right) & \text{if } x \in \left(0, F^{-1}\big(\frac{c}{c+(1-c)\,\tau}\big)\right), \\ (0,1) & \text{if } x \in \left(F^{-1}\big(\frac{c}{c+(1-c)\,\tau}\big), F^{-1}\big(\frac{c\,\tau}{1-c+c\,\tau}\big)\right), \\ \left(F^{-1}\Big(F(x)-\frac{c\,\tau}{(1-c)}\,\big(1-F(x)\big)\Big), 1\right) & \text{if } x \in \left(F^{-1}\big(\frac{c\,\tau}{1-c+c\,\tau}\big), 1\right). \end{cases}$$

Then the arc density of the ICD based on N_F *and* $\mathcal{X}_n$ *has the same distribution as* $\rho_{n,2}(\mathcal{U},\tau,c)$ *(provided in Theorem 4.3).*

Proof: Let $U_i := F(X_i)$ for $i = 1,2,...,n$ and $\mathcal{U}_n := \{U_1, U_2, ..., U_n\}$. Hence, by probability integral transform, $U_i \overset{iid}{\sim} \mathcal{U}(0,1)$. So the image of $N_F(x,r,c)$ under F is $F\big(N_F(x,r,c)\big) = N\big(F(x),r,c\big)$ for (almost) all $x \in (0,1)$. Then $F\big(N_F(X_i,r,c)\big) = N\big(F(X_i),r,c\big) = N(U_i,r,c)$ for $i = 1,2,...,n$. Since $U_i \overset{iid}{\sim} \mathcal{U}(0,1)$, the distribution of the arc density of the ICD based on $N(\cdot,\tau,c)$ and $\mathcal{U}_n$ is given in Theorem 4.3. Observe that for any j, $X_j \in N_F(X_i,\tau,c)$ iff $X_j \in F^{-1}\big(N\big(F(X_i),\tau,c\big)\big)$ iff $F(X_j) \in N\big(F(X_i),\tau,c\big)$ iff $U_j \in N(U_i,\tau,c)$ for $i = 1,2,...,n$. Hence the desired result follows. □

A similar construction is available for the PE-ICDs.

In Proposition 5.1, we have shown that if the defining proximity region for our ICD is defined as $N_F(x,\tau,c) := F^{-1}\big(N\big(F(x),\tau,c\big)\big)$ where F is an increasing function in (a,b) with $a < b$, the exact (and asymptotic) distribution of the arc density based on the ICD for N_F is the same as $\rho_{n,2}(\mathcal{U},\tau,c)$. Hence we can test whether the distribution of any data set is from F or not with the methodology proposed in this article. For example, to test a "data set is from $F(x) = x^2$ with $\mathcal{S}(F) = (0,1)$" (so the inverse is $F^{-1}(x) = \sqrt{x}$ and the corresponding pdf is $f(x) = 2x\,\mathbb{I}(0 < x < 1)$), we need to compute the arc density for the ICD based on the following proximity region: For $\tau \in (0,1]$,

(5.4)

$$N_F(x,\tau,c) = \begin{cases} \left(x\sqrt{1-\tau},\ x\sqrt{1+\frac{(1-c)}{c}\,\tau}\right) & \text{if } x \in \left(0,\sqrt{c}\right), \\ \left(\sqrt{x^2-\frac{c\,\tau}{(1-c)}\,(1-x^2)},\ \sqrt{x^2+(1-x^2)\,\tau}\right) & \text{if } x \in \left(\sqrt{c},1\right), \end{cases}$$

and for $\tau > 1$,

(5.5)

$$N_F(x,\tau,c) = \begin{cases} \left(0,\ x\sqrt{1+\frac{(1-c)}{c}\,\tau}\right) & \text{if } x \in \left(0,\sqrt{\frac{c}{c+(1-c)\,\tau}}\right), \\ (0,1) & \text{if } x \in \left(\sqrt{\frac{c}{c+(1-c)\,\tau}},\sqrt{\frac{c\,\tau}{1-c+c\,\tau}}\right), \\ \left(\sqrt{x^2-\frac{c\,\tau}{(1-c)}\,(1-x^2)},\ 1\right) & \text{if } x \in \left(\sqrt{\frac{c\,\tau}{1-c+c\,\tau}},1\right). \end{cases}$$

Then the arc density for the ICD based on $N_F(\cdot, \tau, c)$ will have the same distribution as $\rho_{n,2}(\mathcal{U}, \tau, c)$ and hence can be used for testing data is from F or not with the procedure discussed in Section 5.

6. DISCUSSION AND CONCLUSIONS

We consider the central similarity interval catch digraphs (CS-ICDs) based on one dimensional data. The CS-ICDs are defined with two parameters: an expansion parameter $\tau > 0$ and a centrality parameter $c \in (0, 1)$. We study the arc density of CS-ICDs, and using its U-statistics property, we derive its asymptotic (normal) distribution for uniform data for the (interiors of) entire ranges of τ and c. Along this process, we also determine the parameters τ and c for which the rate of convergence to normality is the fastest. We also consider the arc density of proportional-edge ICD (PE-ICD) for comparative purposes. We demonstrate that convergence rate of arc density of CS-ICDs is faster than that of PE-PCDs at their respective optimal parameters, which implies that distribution of arc density of CS-ICDs is closer to normality at smaller sample sizes, compared to the arc density of PE-PCDs.

We use the arc density of the ICDs for testing uniformity (i.e., for testing H_o "data set is a random sample from $\mathcal{U}(0, 1)$"), and show that under type I alternatives in which pdf of the data points is larger around one of the end points (0 or 1) CS-ICD test has higher power compared to PE-ICD test, but under the types II and III alternatives in which pdf is larger around 1/2 or around both end points, then PE-ICD test tends to have higher power compared to CS-ICDs. We also compare the ICD tests with two well known tests in literature (namely, Kolmogorov–Smirnov (KS) test and Neyman's smooth test of uniformity). Under type I alternatives, KS test tends to have higher power compared to the ICD tests and Neyman's smooth test, Neyman's smooth test has higher power compared to PE-ICD test, but lower power compared to CS-ICD tests for some parameter combinations. Under type II (resp. type III) alternatives, ICD tests have higher power than KS and Neyman's smooth test for almost all (resp. most) parameter values which have appropriate size. The recommended parameter combinations for the ICDs are provided in Table 2.

The CS-ICDs for one dimensional data can also be used in testing spatial interaction between multiple classes whose support is one-dimensional (see Remark 4.2). The arc density approach is easily adaptable to testing nonuniform distributions as well (see Section 5.4 for more detail). Furthermore, the study of arc density of CS-ICDs in the one dimensional case will provide insight for and form the foundation of related catch digraph extensions in higher dimensions.

APPENDIX

A. PRELIMINARIES

In the proofs below, we can, without loss of generality, assume that the support of the uniform distribution is $(0, 1)$ based on Theorem 4.1.

A.1. Proof of Theorem 4.3

There are two cases for τ, namely $0 < \tau \leq 1$ and $\tau > 1$.

For $\tau \in (0, 1]$, the proximity region is defined as in Equation (4.2) and the Γ_1-region is

$$\Gamma_1(x, \tau, c) = \begin{cases} \left(\frac{cx}{c+(1-c)\tau}, \frac{x}{1-\tau}\right) & \text{if } x \in \left(0, c(1-\tau)\right], \\ \left(\frac{cx}{c+(1-c)\tau}, \frac{(1-c)x+c\tau}{1-c+c\tau}\right) & \text{if } x \in \left(c(1-\tau), c(1-\tau)+\tau\right], \\ \left(\frac{x-\tau}{1-\tau}, \frac{(1-c)x+c\tau}{1-c+c\tau}\right) & \text{if } x \in \left(c(1-\tau)+\tau, 1\right). \end{cases} \tag{A.1}$$

For $\tau > 1$, the proximity region is as in Equation (4.3) and the Γ_1-region is

$$\Gamma_1(x, \tau, c) = \left(\frac{cx}{c+(1-c)\tau}, \frac{(1-c)x+c\tau}{1-c+c\tau}\right). \tag{A.2}$$

Case 1: $0 < \tau \leq 1$: In this case depending on the location of x_1, the following are the different types of the combinations of $N(x_1, \tau, c)$ and $\Gamma_1(x_1, \tau, c)$. Let

$$a_1 := (1-\tau)x_1, \qquad a_2 := x_1\left(1 + \frac{(1-c)\tau}{c}\right),$$

$$a_3 := x_1 - \frac{c\tau(1-x_1)}{1-c}, \qquad a_4 := x_1 + (1-x_1)\tau,$$

and

$$g_1 := \frac{cx_1}{c+(1-c)\tau}, \qquad g_2 := \frac{x_1}{1-\tau},$$

$$g_3 := \frac{x_1-\tau}{1-\tau}, \qquad g_4 := \frac{x_1(1-c)+c\tau}{1-c+c\tau}.$$

Then

(**i**) for $0 < x_1 \leq c(1-\tau)$, we have $N(x_1, \tau, c) = (a_1, a_2)$ and $\Gamma_1(x_1, \tau, c) = (g_1, g_2)$,

(**ii**) for $c\,(1-\tau) < x_1 \leq c$, we have $N(x_1,\tau,c) = (a_1,a_2)$ and $\Gamma_1(x_1,\tau,c) = (g_1,g_4)$,

(**iii**) for $c < x_1 \leq c\,(1-\tau)+\tau$, we have $N(x_1,\tau,c) = (a_3,a_4)$ and $\Gamma_1(x_1,\tau,c) = (g_1,g_4)$,

(**iv**) for $c\,(1-\tau)+\tau < x_1 < 1$, we have $N(x_1,\tau,c) = (a_3,a_4)$ and $\Gamma_1(x_1,\tau,c) = (g_3,g_4)$.

Then

$$p_a(\tau,c) \,=\, P\Big(X_2 \in N(X_1,\tau,c)\Big) \,=\, \int_0^c (a_2-a_1)\,dx_1 + \int_c^1 (a_4-a_3)\,dx_1 \,=\, \tau/2\,.$$

For $\mathbf{Cov}(h_{12},h_{13})$, we need to calculate P_{2N}, P_{NG}, and P_{2G}.

$$\begin{aligned} P_{2N} \,&=\, P\Big(\{X_2,X_3\} \subset N(X_1,\tau,c)\Big) \\ &=\, \int_0^c (a_2-a_1)^2\,dx_1 + \int_c^1 (a_4-a_3)^2\,dx_1 \,=\, \tau^2/3\,. \end{aligned}$$

$$\begin{aligned} P_{NG} \,&=\, P\Big(X_2 \in N(X_1,\tau,c),\, X_3 \in \Gamma_1(X_1,\tau,c)\Big) \\ &=\, \int_0^{c\,(1-\tau)} (a_2-a_1)\,(g_2-g_1)\,dx_1 + \int_{c\,(1-\tau)}^{c} (a_2-a_1)\,(g_4-g_1)\,dx_1 \\ &\quad + \int_c^{c\,(1-\tau)+\tau} (a_4-a_3)\,(g_4-g_1)\,dx_1 + \int_{c\,(1-\tau)+\tau}^{1} (a_4-a_3)\,(g_4-g_3)\,dx_1 \\ &=\, \frac{\tau^2\left(c^2\tau^3 - 5\,c^2\tau^2 - c\,\tau^3 + 4\,c^2\tau + 5\,c\,\tau^2 - 2\,c^2 - 4\,c\,\tau - \tau^2 + 2\,c + 2\,\tau\right)}{6\,(c\,\tau - c + 1)\,(c + \tau - c\,\tau)}\,. \end{aligned}$$

Finally,

$$\begin{aligned} P_{2G} \,&=\, P\Big(\{X_2,X_3\} \subset \Gamma_1(X_1,\tau,c)\Big) \\ &=\, \int_0^{c\,(1-\tau)} (g_2-g_1)^2\,dx_1 + \int_{c\,(1-\tau)}^{c\,(1-\tau)+\tau} (g_4-g_1)^2\,dx_1 + \int_{c\,(1-\tau)+\tau}^{1} (g_4-g_3)^2\,dx_1 \\ &=\, \frac{\left(2\,c^2\tau - c^2 - 2\,c\,\tau + c + \tau\right)\tau^2}{3\,(c\,\tau - c + 1)\,(c + \tau - c\,\tau)}\,. \end{aligned}$$

Therefore

$$\begin{aligned} 4\,\mathbf{E}[h_{12}h_{13}] \,&= \\ &=\, P_{2N} + 2\,P_{NG} + P_{2G} \\ &=\, \frac{\tau^2\left(c^2\tau^3 - 6\,c^2\tau^2 - c\,\tau^3 + 8\,c^2\tau + 6\,c\,\tau^2 - 4\,c^2 - 8\,c\,\tau - \tau^2 + 4\,c + 4\,\tau\right)}{3\,(c\,\tau - c + 1)\,(c + \tau - c\,\tau)}\,. \end{aligned}$$

Hence

$$\begin{aligned} 4\,\mathbf{Cov}[h_{12},h_{13}] \,&= \\ &=\, \frac{\tau^2\left(c^2\tau^3 - 3\,c^2\tau^2 - c\,\tau^3 + 2\,c^2\tau + 3\,c\,\tau^2 - c^2 - 2\,c\,\tau - \tau^2 + c + \tau\right)}{3\,(c\,\tau - c + 1)\,(c + \tau - c\,\tau)}\,. \end{aligned}$$

Case 2: $\tau > 1$: In this case depending on the location of x_1, the following are the different types of the combinations of $N(x_1, \tau, c)$ and $\Gamma_1(x_1, \tau, c)$.

(**i**) for $0 < x_1 \leq \frac{c}{c+(1-c)\tau}$, we have $N(x_1, \tau, c) = (0, a_2)$ and $\Gamma_1(x_1, \tau, c) = (g_1, g_4)$,

(**ii**) for $\frac{c}{c+(1-c)\tau} < x_1 \leq \frac{c\tau}{1-c+c\tau}$, we have $N(x_1, \tau, c) = (0, 1)$ and $\Gamma_1(x_1, \tau, c) = (g_1, g_4)$,

(**iii**) for $\frac{c\tau}{1-c+c\tau} < x_1 < 1$, we have $N(x_1, \tau, c) = (a_3, 1)$ and $\Gamma_1(x_1, \tau, c) = (g_1, g_4)$.

Then

$$\begin{aligned} p_a(\tau, c) &= P\Big(X_2 \in N(X_1, \tau, c)\Big) \\ &= \int_0^{\frac{c}{c+(1-c)\tau}} a_2 \, dx_1 + \int_{\frac{c}{c+(1-c)\tau}}^{\frac{c\tau}{1-c+c\tau}} 1 \, dx_1 + \int_{\frac{c\tau}{1-c+c\tau}}^{1} (1 - a_3) \, dx_1 \\ &= \frac{\tau\left(2\,c^2\tau - 2\,c^2 - 2\,c\tau + 2\,c - 1\right)}{2\,(c\tau - c + 1)\,(c\tau - c - \tau)} \, . \end{aligned}$$

Next

$$\begin{aligned} P_{2N} &= P\Big(\{X_2, X_3\} \subset N(X_1, \tau, c)\Big) \\ &= \int_0^{\frac{c}{c+(1-c)\tau}} a_2^2 \, dx_1 + \int_{\frac{c}{c+(1-c)\tau}}^{\frac{c\tau}{1-c+c\tau}} 1 \, dx_1 + \int_{\frac{c\tau}{1-c+c\tau}}^{1} (1 - a_3)^2 \, dx_1 \\ &= \frac{3\,c^2\tau^2 - 2\,c^2\tau - 3\,c\tau^2 - c^2 + 2\,c\tau + c - \tau}{3\,(c\tau - c + 1)\,(c\tau - c - \tau)} \, . \end{aligned}$$

$$\begin{aligned} P_{NG} &= P\Big(X_2 \in N(X_1, \tau, c),\, X_3 \in \Gamma_1(X_1, \tau, c)\Big) \\ &= \int_0^{\frac{c}{c+(1-c)\tau}} a_2\,(g_4 - g_1) \, dx_1 + \int_{\frac{c}{c+(1-c)\tau}}^{\frac{c\tau}{1-c+c\tau}} (g_4 - g_1) \, dx_1 \\ &\quad + \int_{\frac{c\tau}{1-c+c\tau}}^{1} (1 - a_3)\,(g_4 - g_1) \, dx_1 \\ &= \Big[\tau^2\Big(6\,c^6\tau^4 - 24\,c^6\tau^3 - 18\,c^5\tau^4 + 36\,c^6\tau^2 + 72\,c^5\tau^3 \\ &\quad + 18\,c^4\tau^4 - 24\,c^6\tau - 108\,c^5\tau^2 - 84\,c^4\tau^3 - 6\,c^3\tau^4 + 6\,c^6 \\ &\quad + 72\,c^5\tau + 132\,c^4\tau^2 + 48\,c^3\tau^3 - 18\,c^5 - 92\,c^4\tau - 84\,c^3\tau^2 \\ &\quad - 12\,c^2\tau^3 + 26\,c^4 + 64\,c^3\tau + 30\,c^2\tau^2 - 22\,c^3 - 26\,c^2\tau - 6\,c\tau^2 \\ &\quad + 10\,c^2 + 6\,c\tau - 2\,c - \tau\Big)\Big] \Big/ \Big[6\,(c\tau - c + 1)^3\,(c\tau - c - \tau)^3\Big] \, . \end{aligned}$$

Finally,

$$\begin{aligned}
P_{2G} &= P\Big(\{X_2, X_3\} \subset \Gamma_1(X_1, \tau, c)\Big) \\
&= \int_0^1 (g_4 - g_1)^2 \, dx_1 \\
&= \Big[\tau^2\Big(3\,c^4\tau^2 - 6\,c^4\tau - 6\,c^3\tau^2 + 3\,c^4 + 12\,c^3\tau + 3\,c^2\tau^2 - 6\,c^3 \\
&\quad - 9\,c^2\tau + 7\,c^2 + 3\,c\tau - 4\,c + 1\Big)\Big] \Big/ \Big[3\,(c\tau - c + 1)^2\,(c\tau - c - \tau)^2\Big] .
\end{aligned}$$

Therefore

$$\begin{aligned}
4\,\mathbf{E}[h_{12}h_{13}] &= P_{2N} + 2\,P_{NG} + P_{2G} \\
&= \Big[12\,c^6\tau^6 - 50\,c^6\tau^5 - 36\,c^5\tau^6 + 79\,c^6\tau^4 + 150\,c^5\tau^5 + 36\,c^4\tau^6 \\
&\quad - 56\,c^6\tau^3 - 237\,c^5\tau^4 - 175\,c^4\tau^5 - 12\,c^3\tau^6 + 14\,c^6\tau^2 \\
&\quad + 168\,c^5\tau^3 + 297\,c^4\tau^4 + 100\,c^3\tau^5 + 2\,c^6\tau - 42\,c^5\tau^2 \\
&\quad - 220\,c^4\tau^3 - 199\,c^3\tau^4 - 25\,c^2\tau^5 - c^6 - 6\,c^5\tau + 58\,c^4\tau^2 \\
&\quad + 160\,c^3\tau^3 + 75\,c^2\tau^4 + 3\,c^5 + 7\,c^4\tau - 46\,c^3\tau^2 \\
&\quad - 70\,c^2\tau^3 - 15\,c\tau^4 - 3\,c^4 - 4\,c^3\tau + 20\,c^2\tau^2 + 18\,c\tau^3 \\
&\quad + c^3 + c^2\tau - 4\,c\tau^2 - 3\,\tau^3\Big] \Big/ \Big[3\,(c\tau - c + 1)^3\,(c\tau - c - \tau)^3\Big] .
\end{aligned}$$

Hence

$$\begin{aligned}
4\,\mathbf{Cov}[h_{12}, h_{13}] &= \Big[c\,(1-c)\,\Big(2\,c^4\tau^5 - 7\,c^4\tau^4 - 4\,c^3\tau^5 + 8\,c^4\tau^3 + 14\,c^3\tau^4 \\
&\quad + 3\,c^2\tau^5 - 2\,c^4\tau^2 - 16\,c^3\tau^3 - 7\,c^2\tau^4 - c\tau^5 - 2\,c^4\tau + 4\,c^3\tau^2 \\
&\quad + 12\,c^2\tau^3 + c^4 + 4\,c^3\tau - 6\,c^2\tau^2 - 4\,c\tau^3 - 2\,c^3 - 3\,c^2\tau + 4\,c\tau^2 \\
&\quad + c^2 + c\tau - \tau^2\Big)\Big] \Big/ \Big[3\,(c\tau - c + 1)^3\,(c\tau - c - \tau)^3\Big] . \quad \square
\end{aligned}$$

A.1.1. Special Case (i) $\tau > 0$ and $c = 1/2$

For $x \in (0, 1/2)$, the proximity region for $\tau \in (0, 1]$ is

$$N(x, \tau, 1/2) = \begin{cases} \big((1-\tau)\,x,\ (1+\tau)\,x\big) & \text{if } x \in (0, 1/2)\,, \\ \big(x - (1-x)\,\tau,\ x + (1-x)\,\tau\big) & \text{if } x \in (1/2, 1)\,, \end{cases} \tag{A.3}$$

and for $\tau > 1$

$$N(x, \tau, 1/2) = \begin{cases} \big(0, (1+\tau)\,x\big) & \text{if } x \in \big(0, 1/(1+\tau)\big)\,, \\ (0, 1) & \text{if } x \in \big(1/(1+\tau),\ \tau/(1+\tau)\big)\,, \\ \big(x - (1-x)\,\tau, 1\big) & \text{if } x \in \big(\tau/(1+\tau), 1\big)\,. \end{cases} \tag{A.4}$$

Corollary A.1. *For $\tau \in (0,\infty)$ and $c = 1/2$, we have $\sqrt{n}\big[\rho_{n,2}(\tau,1/2) - p_a(\tau,1/2)\big] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(\tau,1/2)\big)$ as $n \to \infty$, where*

$$p_a(\tau,1/2) = \begin{cases} \tau/2 & \text{if } 0 < \tau < 1\,, \\ \tau/(\tau+1) & \text{if } \tau > 1\,, \end{cases} \tag{A.5}$$

and

$$4\,\nu(\tau,1/2) = \begin{cases} \frac{\tau^2\,(1+2\tau-\tau^2-\tau^3)}{3\,(\tau+1)^2} & \text{if } 0 < \tau \le 1\,, \\ \frac{2\tau-1}{3\,(\tau+1)^2} & \text{if } \tau > 1\,. \end{cases} \tag{A.6}$$

See Figure 10 for the plots of $p_a(\tau,1/2)$ and $4\,\nu(\tau,1/2)$ with $\tau \in (0,5]$.

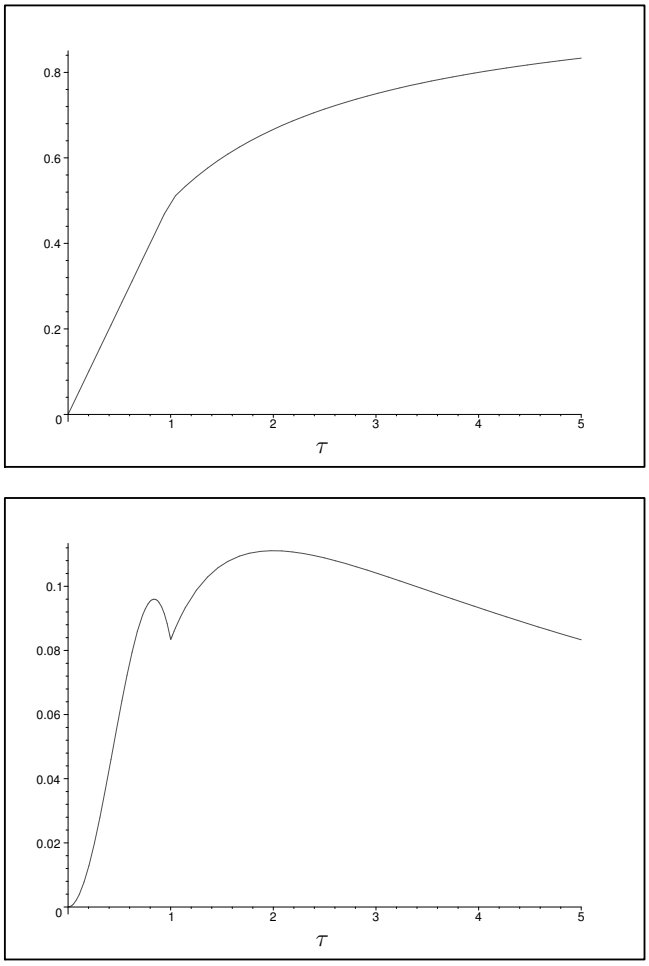

Figure 10: The plots of the asymptotic mean $p_a(\tau,1/2)$ (top) and the variance $4\,\nu(\tau,1/2)$ (bottom) as a function of τ for $\tau \in (0,5]$.

The sharpest rate of convergence in Corollary A.1 is $\frac{K}{\sqrt{n}} f_{CS}^{c}(\tau, 1/2)$ where

$$f_{CS}^{c}(\tau,1/2) = \begin{cases} \frac{27\tau}{2}\left(\frac{(6\tau+3-3\tau^3-3\tau^2)\tau^2}{(\tau+1)^2}\right)^{-3/2} & \text{if } 0<\tau\leq 1\,, \\ \frac{3\sqrt{3}\tau}{\tau+1}\left(\frac{2\tau-1}{(\tau+1)^2}\right)^{-3/2} & \text{if } \tau>1\,, \end{cases} \tag{A.7}$$

and is minimized at $\tau \approx .73$ which is found by using simple calculus and numerical methods.

The plot of $p_a(\tau, 1/2)/\sqrt{\nu(\tau, 1/2)^3}$ also indicates that this is where the global minimum occurs. Convergence rates for PE- and CS-ICDs are presented in Figure 11 (bottom) for $c = 1/2$ as a function of expansion parameter. See [Ceyhan, 2012] for the explicit form of $f_{PE}^{c}(r, 1/2)$. Notice that at the optimal expansion parameters, convergence rate of CS-ICDs is faster with $c = 1/2$.

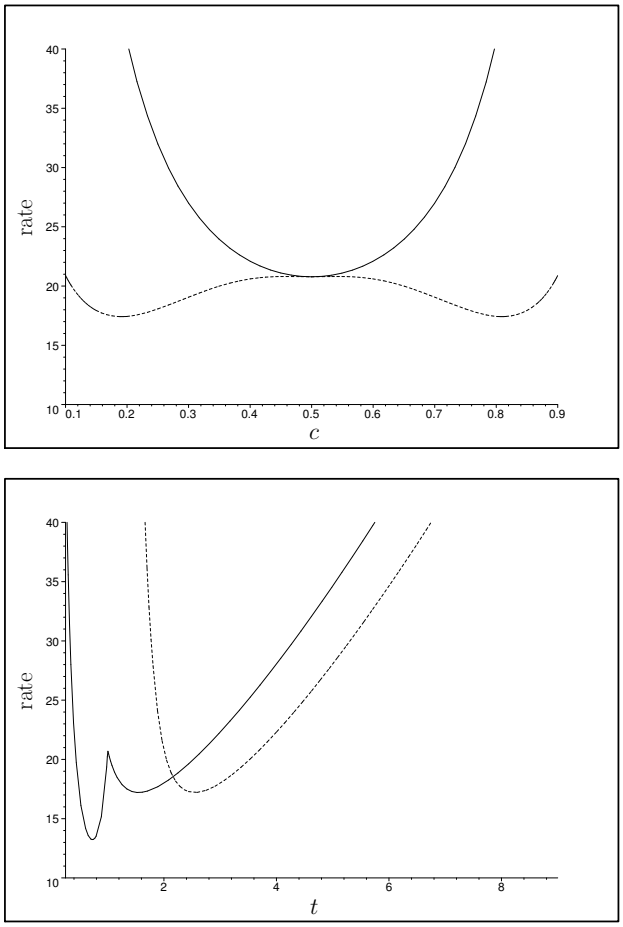

Figure 11: The plots of the rates of convergence to normality for PE- and CS-ICDs. Plotted in the top are $f_{CS}^{c}(1, c)$ (solid line) and $f_{PE}^{c}(2, c)$ (dashed line); and in the bottom are $f_{CS}^{c}(t, 1/2)$ (solid) and $f_{PE}^{c}(t, 1/2)$ (dashed).

A.1.2. Special Case (ii) $\tau = 1$ and $c \in (0,1)$

For $x \in (0,1)$, the proximity region has the following form:

$$N(x,1,c) = \begin{cases} (0, x/c) & \text{if } x \in (0,c)\,, \\ \big((x-c)/(1-c), 1\big) & \text{if } x \in (c,1)\,. \end{cases} \tag{A.8}$$

Corollary A.2. *As $n \to \infty$, for $c \in (0,1)$, we have $\sqrt{n}\,[\rho_{n,2}(1,c) - p_a(1,c)] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(1,c)\big)$, where $p_a(1,c) = 1/2$ and $4\nu(1,c) = c\,(1-c)/3$.*

Observe that $p_a(1,c)$ is constant (i.e., independent of c) and $\nu(1,c)$ is symmetric around $c = 1/2$ with $\nu(1,c) = \nu(1,1-c)$. Let $\frac{K}{\sqrt{n}}\, f^c_{CS}(\tau,c)$ be the rate of convergence to normality for CS-ICDs. Then the sharpest rate of convergence in Corollary A.2 is $\frac{K}{\sqrt{n}}\, f^c_{CS}(1,c)$ where

$$f^c_{CS}(1,c) \;=\; \frac{3\sqrt{3}}{2\sqrt{c^3\,(1-c)^3}}\,. \tag{A.9}$$

Convergence rate is minimized at $c = 1/2$ (which can be verified by simple calculus). Also, let $\frac{K}{\sqrt{n}}\, f^c_{PE}(r,c)$ be the rate of convergence to normality for PE-ICDs (see [Ceyhan, 2012] for its explicit forms). Then we have $f^c_{PE}(2,c) \le f^c_{CS}(1,c)$ for all $c \in (0,1)$ with equality holding only at $c = 1/2$ (see also Figure 11 (top)). Thus at these specific centrality parameters, convergence rate to normality is faster for PE-PCDs.

A.1.3. Special Case (iii) $\tau = 1$ and $c = 1/2$

In this case we have $N(x,1,1/2) = B(x, r(x))$ where $r(x) = \min(x, 1-x)$ for $x \in (0,1)$. Hence CS-ICD based on $N(x,1,1/2)$ is equivalent to the CCCD of [Priebe *et al.*, 2001] and the PE-ICD with expansion parameter 2 and centrality parameter 1/2 of [Ceyhan, 2012].

Corollary A.3. *As $n \to \infty$, we have $\sqrt{n}\,[\rho_n(1,1/2) - p_a(1,1/2)] \xrightarrow{\mathcal{L}} \mathbb{N}\big(0, 4\,\nu(1,1/2)\big)$, where $p_a(1,1/2) = 1/2$ and $4\nu(1,1/2) = 1/12$ with the sharpest rate of convergence being $K\frac{p_a(1,1/2)}{\sqrt{n\,\nu(1,1/2)^3}} = 12\sqrt{3}\,\frac{K}{\sqrt{n}}$.*

A.2. Proof of Theorem 4.4

There are two cases for τ, namely, $0 < \tau < 1$ and $\tau > 1$.

For $\tau \in (0,1]$ and x in the right end interval, the proximity region is

$$N_e(x,\tau) = \begin{cases} \big((1-\tau)\,x,\ (1+\tau)\,x\big) & \text{if } x \in \big(0, 1/(1+\tau)\big)\,, \\ \big((1-\tau)\,x, 1\big) & \text{if } x \in \big(1/(1+\tau), 1\big)\,, \end{cases} \tag{A.10}$$

and the Γ_1-region is

$$\Gamma_{1,e}(x,\tau) = \begin{cases} \Big(\frac{x}{1+\tau}, \frac{x}{1-\tau}\Big) & \text{if } x \in (0, 1-\tau)\,, \\ \Big(\frac{x}{1+\tau}, 1\Big) & \text{if } x \in (1-\tau, 1)\,. \end{cases} \tag{A.11}$$

For $\tau > 1$ and x in the right end interval, the proximity region is

$$N_e(x,\tau) = \begin{cases} \big(0, (1+\tau)\,x\big) & \text{if } x \in \big(0, 1/(1+\tau)\big)\,, \\ (0,1) & \text{if } x \in \big(1/(1+\tau), 1\big)\,, \end{cases} \tag{A.12}$$

and the Γ_1-region is $\Gamma_{1,e}(x,\tau) = \big(x/(1+\tau), 1\big)$.

Case 1: $0 < \tau \leq 1$: For $x_1 \in (0,1)$, depending on the location of x_1, the following are the different types of the combinations of $N_e(x_1,\tau)$ and $\Gamma_{1,e}(x_1,\tau)$.

- (**i**) for $0 < x_1 \leq 1-\tau$, we have $N_e(x_1,\tau) = \big((1-\tau)\,x_1, (1+\tau)\,x_1\big)$ and $\Gamma_{1,e}(x_1,\tau) = \big(x_1/(1+\tau), x_1/(1-\tau)\big)$,
- (**ii**) for $1-\tau < x_1 \leq 1/(1+\tau)$, we have $N_e(x_1,\tau) = \big((1-\tau)\,x_1, (1+\tau)\,x_1\big)$ and $\Gamma_{1,e}(x_1,\tau) = \big(x_1/(1+\tau), 1\big)$,
- (**iii**) for $1/(1+\tau) < x_1 < 1$, we have $N_e(x_1,\tau) = \big((1-\tau)\,x_1, 1\big)$ and $\Gamma_{1,e}(x_1,\tau) = \big(x_1/(1+\tau), 1\big)$.

Then

$$\begin{aligned} p_a^e(\tau,c) &= P\Big(X_2 \in N_e(X_1,\tau)\Big) \\ &= \int_0^{1/(1+\tau)} \big((1+\tau)\,x_1 - (1-\tau)\,x_1\big)\,dx_1 + \int_{1/(1+\tau)}^1 \big(1-(1-\tau)\,x_1\big)\,dx_1 \\ &= \int_0^{1/(1+\tau)} (2\,\tau\,x_1)\,dx_1 + \int_{1/(1+\tau)}^1 (1-x_1+x_1\,\tau)\,dx_1 \;=\; \frac{\tau\,(\tau+2)}{2\,(\tau+1)}\,. \end{aligned}$$

For $\mathbf{Cov}(h_{12}, h_{13})$, we need to calculate $P_{2N,e}$, $P_{NG,e}$, and $P_{2G,e}$.

$$\begin{aligned} P_{2N,e} &= P\Big(\{X_2, X_3\} \subset N_e(X_1,\tau)\Big) \\ &= \int_0^{1/(1+\tau)} (2\,\tau\,x_1)^2\,dx_1 + \int_{1/(1+\tau)}^1 (1-x_1+x_1\,\tau)^2\,dx_1 \\ &= \frac{\tau^2\,(\tau^2+3\,\tau+4)}{3\,(\tau+1)^2}\,. \end{aligned}$$

$$\begin{aligned} P_{NG,e} &= P\Big(X_2 \in N_e(X_1,\tau), X_3 \in \Gamma_{1,e}(X_1,\tau)\Big) \\ &= \int_0^{1-\tau} (2\,\tau\,x_1)\left(\frac{2\,\tau\,x_1}{1-\tau^2}\right) dx_1 + \int_{1-\tau}^{1/(1+\tau)} (2\,\tau\,x_1)\left(1-\frac{x_1}{1+\tau}\right) dx_1 \\ &\quad + \int_{1/(1+\tau)}^{1} \big(1-(1-\tau)\,x_1\big)\left(1-\frac{x_1}{1+\tau}\right) dx_1 \\ &= \frac{\big(7\,\tau^2+14\,\tau+8-2\,\tau^4-2\,\tau^3\big)\,\tau^2}{6\,(\tau+1)^3}\,. \end{aligned}$$

Finally,

$$\begin{aligned} P_{2G,e} &= P\Big(\{X_2,X_3\} \subset \Gamma_{1,e}(X_1,\tau)\Big) \\ &= \int_0^{1-\tau}\left(\frac{2\,\tau\,x_1}{1-\tau^2}\right)^2 dx_1 + \int_{1-\tau}^{1}\left(1-\frac{x_1}{1+\tau}\right)^2 dx_1 \;=\; \frac{\tau^2\,(3\,\tau+4)}{3\,(\tau+1)^2}\,. \end{aligned}$$

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N,e} + 2\,P_{NG,e} + P_{2G,e} = \frac{\tau^2\left(2\tau^2+5\tau+4\right)\left(2\tau+4-\tau^2\right)}{3\,(\tau+1)^3}$. Hence

$$4\,\mathbf{Cov}[h_{12},h_{13}] \;=\; \frac{\tau^2\left(4\,\tau+4-2\,\tau^4-4\,\tau^3-\tau^2\right)}{3\,(\tau+1)^3}\,.$$

Case 2: $\tau > 1$: For $x_1 \in (0,1)$, depending on the location of x_1, the following are the different types of the combinations of $N_e(x_1,\tau)$ and $\Gamma_{1,e}(x_1,\tau)$.

- (**i**) for $0 < x_1 \le 1/(1+\tau)$, we have $N_e(x_1,\tau) = \big(0,(1+\tau)\,x_1\big)$ and $\Gamma_{1,e}(x_1,\tau) = \big(x_1/(1+\tau),1\big)$,
- (**ii**) for $1/(1+\tau) < x_1 < 1$, we have $N_e(x_1,\tau) = (0,1)$ and $\Gamma_{1,e}(x_1,\tau) = \big(x_1/(1+\tau),1\big)$.

Then

$$\begin{aligned} p_a^e(\tau,c) &= P\Big(X_2 \in N_e(X_1,\tau)\Big) \\ &= \int_0^{1/(1+\tau)} (1+\tau)\,x_1\,dx_1 + \int_{1/(1+\tau)}^{1} 1\,dx_1 \;=\; \frac{1+2\,\tau}{2\,(\tau+1)}\,. \end{aligned}$$

Next,

$$\begin{aligned} P_{2N,e} &= P\Big(\{X_2,X_3\} \subset N_e(X_1,\tau)\Big) \\ &= \int_0^{1/(1+\tau)} \big((1+\tau)\,x_1\big)^2\,dx_1 + \int_{1/(1+\tau)}^{1} 1\,dx_1 \;=\; \frac{1+3\,\tau}{3\,(\tau+1)}\,, \end{aligned}$$

$$\begin{aligned} P_{NG,e} &= P\Big(X_2 \in N_e(X_1,\tau), X_3 \in \Gamma_{1,e}(X_1,\tau)\Big) \\ &= \int_0^{1/(1+\tau)} \big((1+\tau)\,x_1\big)\left(1-\frac{x_1}{1+\tau}\right) dx_1 + \int_{1/(1+\tau)}^{1}\left(1-\frac{x_1}{1+\tau}\right) dx_1 \\ &= \frac{6\,\tau^3+12\,\tau^2+6\,\tau+1}{6\,(\tau+1)^3}\,. \end{aligned}$$

Finally,

$$P_{2G,e} = P\Big(\{X_2, X_3\} \subset \Gamma_{1,e}(X_1, \tau)\Big)$$
$$= \int_0^1 \left(1 - \frac{x_1}{1+\tau}\right)^2 dx_1 = \frac{3\,\tau^2 + 3\,\tau + 1}{3\,(\tau+1)^2}\,.$$

Therefore $4\,\mathbf{E}[h_{12}h_{13}] = P_{2N,e} + 2\,P_{NG,e} + P_{2G,e} = \frac{12\,\tau^3+25\,\tau^2+15\,\tau+3}{3\,(\tau+1)^3}$. Hence

$$4\,\mathbf{Cov}[h_{12}, h_{13}] = \frac{\tau^2}{3\,(\tau+1)^3}\,. \quad \square$$

A.3. Figures for the consistency results in Section 5.3

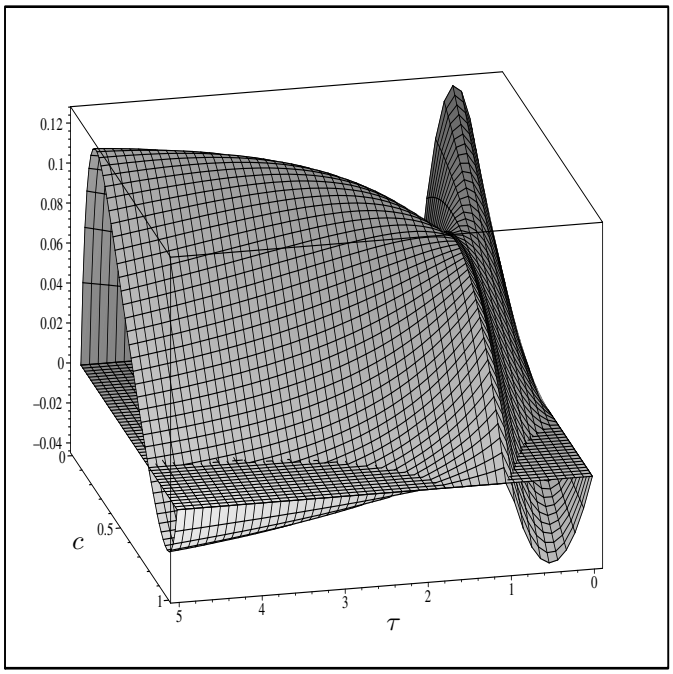

Figure 12: The three dimensional plot of the difference between arc probability of CS-ICD under $H_a^I\colon \delta = 2$ and the null hypothesis $p_a(F_1, \tau, c) - p_a(\mathcal{U}, \tau, c)$ for $c \in (0, 1)$ and $\tau \in (0, 5)$. The horizontal plane is at $z = 0$ and is used to determine the sign changes in the difference.

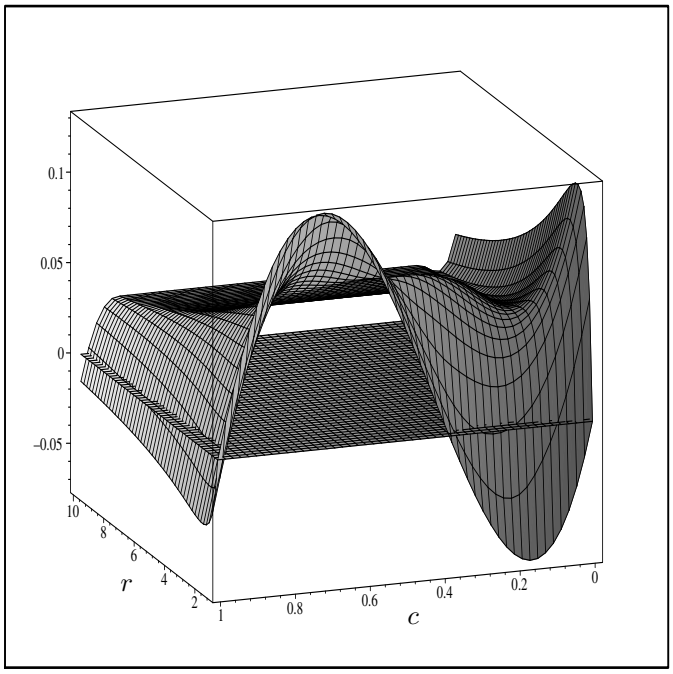

Figure 13: The three dimensional plot of the difference between arc probability of PE-ICD under $H_a^I : \delta = 2$ and the null hypothesis $p_a^{PE}(F_1, r, c) - p_a^{PE}(\mathcal{U}, r, c)$ for $c \in (0, 1)$ and $r \in (1, 10)$. The horizontal plane is at $z = 0$ and is used to determine the sign changes in the difference.

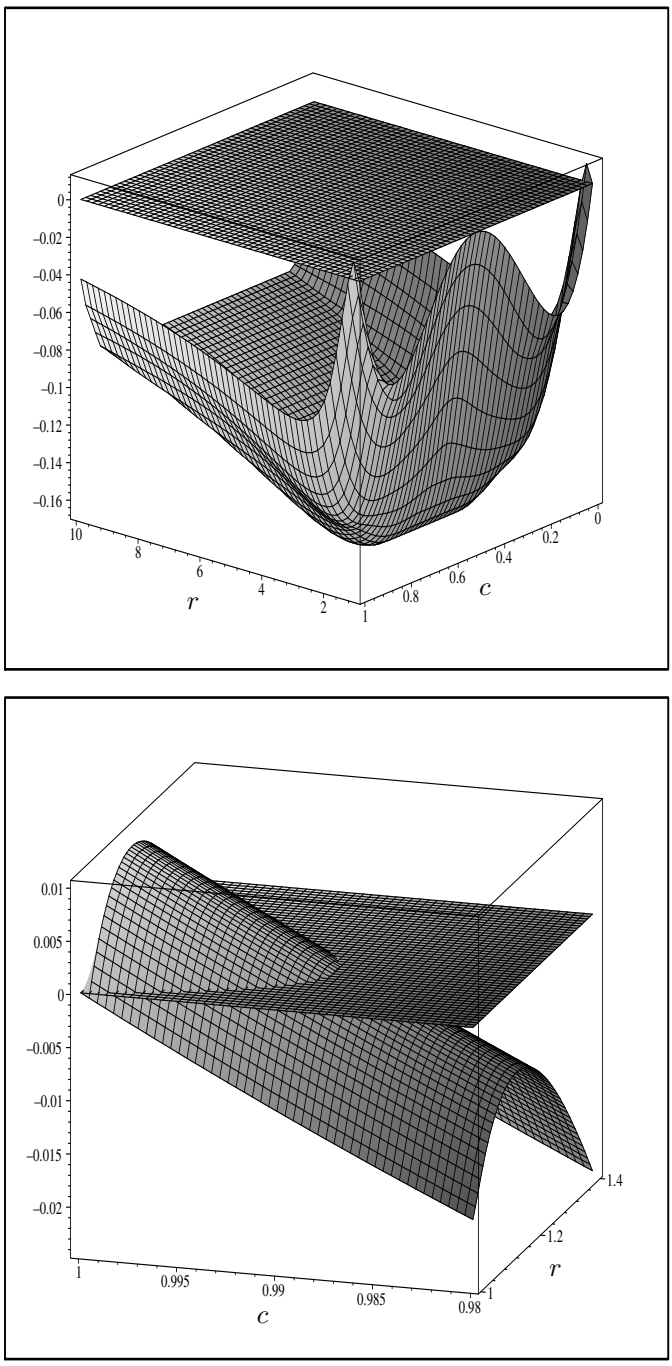

Figure 14: The three dimensional plot of the difference between arc probability of PE-ICD under $H_a^{III}: \delta = 2$ and the null hypothesis $p_a^{PE}(F_3, r, c) - p_a^{PE}(\mathcal{U}, r, c)$. The top plot is with $c \in (0, 1)$ and $r \in (1, 10)$ and the bottom plot is with $c \in (.98, 1)$ and $r \in (1, 1.4)$ (to better visualize the region of positive difference around $(r, c) = (1, 1)$). The horizontal planes at $z = 0$ are used to determine the sign changes in the difference.

ACKNOWLEDGMENTS

I would like to thank the anonymous referee, whose constructive comments and suggestions greatly improved the presentation and flow of this article. This research was supported by the European Commission under the Marie Curie International Outgoing Fellowship Programme via Project # 329370 titled PRinHDD.

REFERENCES

[1] Beineke, L.W. and Zamfirescu, C.M. (1982). Connection digraphs and second order line graphs, *Discrete Mathematics*, **39**, 237–254.

[2] Callaert, H. and Janssen, P. (1978). The Berry–Esseen theorem for U-statistics, *Annals of Statistics*, **6**, 417–421.

[3] Cannon, A. and Cowen, L. (2000). *Approximation algorithms for the class cover problem*. In "Proceedings of the 6th International Symposium on Artificial Intelligence and Mathematics", January 5–7, 2000, Fort Lauderdale, Florida.

[4] Ceyhan, E. (2012). The distribution of the relative arc density of a family of interval catch digraph based on uniform data, *Metrika*, **75**(6), 761–793.

[5] Ceyhan, E. and Priebe, C.E. (2005). The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association, *Statistics & Probability Letters*, **73**, 37–50.

[6] Ceyhan, E.; Priebe, C.E. and Marchette, D.J. (2007). A new family of random graphs for testing spatial segregation, *Canadian Journal of Statistics*, **35**(1), 27–50.

[7] Chartrand, G.; Lesniak, L. and Zhang, P. (2010). *Graphs & Digraphs*, Chapman and Hall/CRC 5th Edition, Boca Raton, Florida.

[8] Coleman, T.F. and Moré, J.J. (1983). Estimation of sparse Jacobian matrices and graph coloring problems, *SIAM Journal on Numerical Analysis*, **20**(1), 187–209.

[9] Douglas, B.W. (1996). Short proofs for interval digraphs, *Discrete Math*, **178**, 287–292.

[10] Goldberg, A.V. (1984). Finding a maximum density subgraph, Technical Report UCB/CSD-84-171, EECS Department, University of California, Berkeley.

[11] Grünbaum, B. (1988). The edge-density of 4-critical planar graphs, *Combinatorica*, **8**(1), 137–139.

[12] Jain, A.K.; Xu, X.; Ho, T.K. and Xiao, F. (2002). *Uniformity testing using minimal spanning tree*. In "Proceedings of the 16th International Conference on Pattern Recognition (ICPRŠ02)". 04:40281.

[13] Janson, S.; Łuczak, T. and Ruciński, A. (2000). *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York.

[14] Lehmann, E.L. (2004). *Elements of Large Sample Theory*, Springer, New York.

[15] Maehara, H. (1984). A digraph represented by a family of boxes or spheres, *Journal of Graph Theory*, **8**(3), 431–439.

[16] Marhuenda, Y.; Morales, D. and Pardo, M.C. (2005). A comparison of uniformity tests, *Statistics*, **39**(4), 315–327.

[17] Priebe, C.E.; DeVinney, J.G. and Marchette, D.J. (2001). On the distribution of the domination number of random class cover catch digraphs, *Statistics & Probability Letters*, **55**, 239–246.

[18] Prisner, E. (1989). A characterization of interval catch digraphs, *Discrete Mathematics*, **73**, 285–289.

[19] Prisner, E. (1994). Algorithms for interval catch digraphs, *Discrete Applied Mathematics*, **51**, 147–157.

[20] Stephens, M.A. (1974). EDF statistics for goodness of fit and some comparisons, *Journal of American Statistical Association*, **69**, 730–737.

[21] Toussaint, G.T. (1980). The relative neighborhood graph of a finite planar set, *Pattern Recognition*, **12**(4), 261–268.

REVSTAT – Statistical Journal
Volume 14, Number 4, October 2016, 395–413

ON THE IDENTIFIABILITY CONDITIONS IN SOME NONLINEAR TIME SERIES MODELS

Authors: Jungsik Noh
– Department of Bioinformatics,
University of Texas Southwestern Medical Center, Dallas, USA
jungsik.noh@utsouthwestern.edu

Sangyeol Lee
– Department of Statistics, Seoul National University,
Seoul, Korea
sylee@stats.snu.ac.kr

Received: March 2014 Revised: February 2015 Accepted: March 2015

Abstract:

- In this study, we consider the identifiability problem for nonlinear time series models. Special attention is paid to smooth transition GARCH, nonlinear Poisson autoregressive, and multiple regime smooth transition autoregressive models. Some sufficient conditions are obtained to establish the identifiability of these models.

Key-Words:

- *identifiability, nonlinear time series models; GARCH-type models; smooth transition GARCH models; Poisson autoregressive models; smooth transition autoregressive models.*

AMS Subject Classification:

- 62F10, 62M10.

1. INTRODUCTION

Verifying the identifiability conditions for time series models is a fundamental task in constructing the consistent estimators of model parameters and ensuring the positive definiteness of their asymptotic covariance matrices. Although time series models are assumed to be identifiable in many situations, its verification is often nontrivial and even troublesome, especially in handling nonlinear generalized autoregressive conditional heteroscedasticity (GARCH) models. This issue has a long history and there exist a vast amount of relevant studies in the literature. For instance, Rothenberg [28] introduced the global and local identification concept and verified that local identifiability is equivalent to the nonsingularity of the information matrix. Phillips [26] derived asymptotic theories in partially identified models. Hansen [12] and Francq *et al.* [7] proposed a test for the hypothesis wherein nuisance parameters are unidentifiable. Komunjer [16] provided the primitive conditions for global identification in moment restriction models. In most cases, the identifiability condition is inherent to given statistical models; for example, the multiple linear regression model is unidentifiable when exact multicollinearity exists. Thus, in nature, the verification of identifiability is more complicated in nonlinear time series models with volatilities, such as threshold autoregressive and smooth transition GARCH models (see, for instance, Chan [3] and Meitz and Saikkonen [24]). Thus, there is a need to develop a more refined approach than the existing ones to cope with the problem more adequately.

In this study, we deal with the identifiability problem within a framework similar to that of the M-estimation. To elucidate, let us consider the nonlinear least squares (NLS) estimation from a strictly stationary ergodic process $\{(Y_t, Z_t)\}$, with $E(Y_t|Z_t) = f(Z_t, \beta^\circ)$ for some known function f. Then, the limit of the random objective functions for parameter estimation is uniquely minimized at β° when the following identifiability condition holds:

$$f(Z_1, \beta) = f(Z_1, \beta^\circ) \quad \text{a.s. implies} \quad \beta = \beta^\circ\,. \tag{1.1}$$

In most M-estimation procedures, the identifiability conditions are given in the form of (1.1), where f can be a conditional mean, variance, or quantile function (see Hayashi [13, p. 463], Berkes *et al.* [2], and Lee and Noh [19]). Moreover, as seen in Wu [33], to ensure the positive definiteness of asymptotic covariance matrices of the NLS estimator, one needs to verify that $\lambda^T \partial f(Z_1, \beta^\circ)/\partial\beta = 0$ a.s. implies $\lambda = 0$. The method described in this study is also useful to verify the positive definiteness of asymptotic covariance matrices of parameter estimators.

As a representative study on the issue with nonlinear time series, we can refer to Chan and Tong [4], who studied the asymptotic theory of NLS estimators for the smooth transition AR (STAR) models and verified the positive

definiteness of asymptotic variance matrices. Later, many authors handled this problem using various GARCH-type models because it is crucial when verifying the asymptotic properties of quasi-maximum likelihood estimators (QMLEs). For example, Straumann–Mikosch [29], Medeiros and Veiga [22], Kristensen and Rahbek [17], Meitz and Saikkonen [24], and Lee and Lee [18] consider the identifiability problem in exponential and asymmetric GARCH(p, q) models, flexible coefficient GARCH$(1, 1)$ models nesting a smooth transition GARCH$(1, 1)$ (STGARCH) model, nonlinear ARCH models, nonlinear AR(p) models with nonlinear GARCH$(1, 1)$ errors, STAR(p)–STGARCH$(1, 1)$ models, and Box–Cox transformed threshold GARCH(p, q) models. To ensure (1.1), these authors developed their own methods that reflect the nonlinear structure of underlying models.

In this study, we develop a method that refines existing ones to deduce the identifiability conditions for various nonlinear time series models, tribute to STGARCH(p, q), Poisson autoregressive, and multiple regime STAR(p) models. The remainder of this paper is organized as follows. In Section 2, we describe our method using some examples. In Section 3, we investigate the identifiability conditions in the aforementioned models. The proofs are provided in Section 4.

2. EXAMPLES AND MOTIVATION

In this section, we explore some existing methods that verify the identifiability of STAR models and asymmetric GARCH (AGARCH) models. In what follows, $\{X_t\}$ and $\mathcal{F}_t$ denote the data-generating process and the σ-field generated by $\{X_s : s \le t\}$.

First, we consider the STAR model with two regimes as follows:

$$\begin{aligned} X_t &= m(X_{t-1}, ..., X_{t-p}; \theta^\circ) + \varepsilon_t \,, \\ m(X_{t-1}, ..., X_{t-p}; \theta^\circ) &= \beta_0^{\circ T} \mathbf{X}_{t-1} + \beta_1^{\circ T} \mathbf{X}_{t-1} F\left(\frac{X_{t-d} - c^\circ}{z^\circ}\right), \end{aligned}$$

where $\{\varepsilon_t\}$ are iid random variables, $\theta^{\circ T} = (\beta_0^{\circ T}, \beta_1^{\circ T}, c^\circ, r^\circ)$ and $\mathbf{X}_{t-1} = (1, X_{t-1}, ..., X_{t-p})^T$, and $F(\cdot)$ is a smooth distribution function. Chan and Tong [4] verified the positive definiteness of $E[\dot{m}_t(\theta^\circ)\dot{m}_t(\theta^\circ)^T]$, where $\dot{m}_t(\theta^\circ) = \dot{m}(X_{t-1}, ..., X_{t-p}; \theta^\circ)$ denotes the gradient of $m(x; \theta)$ at θ°, by showing that for a given $\lambda \neq 0$, there exists $S \subset \mathbb{R}^p$, such that $\{\lambda^T \dot{m}(x; \theta^\circ)\}^2$ is positive for any $x \in S$ and $P(\{(X_{t-1}, ..., X_{t-p}) \in S\}) > 0$. On the other hand, Meitz and Saikkonen [24] also considered the above model and verified that $m(X_{t-1}, ..., X_{t-p}; \theta) = m(X_{t-1}, ..., X_{t-p}; \theta^\circ)$ a.s. implies $\theta = \theta^\circ$. In both cases, the main step is commonly to show that the function $x \mapsto g(x; \theta, \theta^\circ)$, which equals $(\theta - \theta^\circ)^T \dot{m}(x; \theta^\circ)$ in Chan and Tong [4] and $m(x; \theta) - m(x; \theta^\circ)$ in Meitz and Saikkonen [24], satisfies $g(x; \theta, \theta^\circ) = 0$ for

all $x \in \mathrm{supp}(X_{t-1}, ..., X_{t-p})$, where $\mathrm{supp}(Y)$ denotes the distribution support of the random vector Y. With this equation, they could deduce certain conditions to guarantee $\theta = \theta^\circ$. Motivated by these studies, we take a similar approach to deduce the identifiability conditions for nonlinear time series models. In fact, our method is handier than those in the existing studies, such as Kristensen–Rahbek [17], Meitz and Saikkonen [24], and Lee and Lee [18]. For example, our method no longer requires the condition that either the observations or their conditional volatilities should take all values of an open interval with a positive probability.

Next, we consider the case of an AGARCH(1,1) model with power 2:

$$X_t = \sigma_t \eta_t \ , \qquad \sigma_t^2 \ = \ \omega^\circ + \alpha^\circ \left(|X_{t-1}| - \gamma^\circ X_{t-1}\right)^2 + \beta^\circ \sigma_{t-1}^2 \ , \tag{2.1}$$

where $\{\eta_t\}$ is a sequence of iid random variables with $E\eta_t = 0$ and $E\eta_t^2 = 1$. Kristensen and Rahbek [17] and Straumann and Mikosch [29] derived identifiability conditions for asymmetric power ARCH and AGARCH models. We denote $\theta^\circ = (\omega^\circ, \alpha^\circ, \beta^\circ, \gamma^\circ)^T$ and $\Theta = (0, \infty) \times [0, \infty) \times [0, 1) \times [-1, 1]$, where $\alpha^\circ > 0$. Assuming that Model (2.1) has a strictly stationary solution $\{X_t\}$, for $\theta \in \Theta$, we define a strictly stationary process $\{\sigma_t^2(\theta)\}$ as the solution of

$$\sigma_t^2(\theta) \ = \ \omega + \alpha \left(|X_{t-1}| - \gamma X_{t-1}\right)^2 + \beta \sigma_{t-1}^2(\theta) \ , \qquad \forall t \in \mathbb{Z} \ , \tag{2.2}$$

where $\sigma_t^2(\theta^\circ)$ is equal to σ_t^2.

In this case, the identifiability condition is that $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$ implies $\theta = \theta^\circ$, which is crucial to verify the strong consistency of QMLE. Below, we demonstrate the approach of Straumann and Mikosch [29]. Note that $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for all t because $\{\sigma_t^2 - \sigma_t^2(\theta)\}$ is stationary. Then, one can obtain

$$\omega^\circ - \omega + \sigma_{t-1}^2 Y_{t-1} = 0 \quad \text{a.s.} \, , \tag{2.3}$$

where $Y_{t-1} = \alpha^\circ \left(|\eta_{t-1}| - \gamma^\circ \eta_{t-1}\right)^2 - \alpha \left(|\eta_{t-1}| - \gamma \eta_{t-1}\right)^2 + \beta^\circ - \beta$. As shown in Lemma 5.3 of Straumann and Mikosch [29], Y_{t-1} is $\mathcal{F}_{t-2}$-measurable due to (2.3), but at the same time, it is independent of $\mathcal{F}_{t-2}$. Then, $\theta = \theta^\circ$ can be easily deduced from the degeneracy of Y_{t-1} and certain mild conditions on the distribution of η_{t-1}. This approach, however, cannot be extended straightforwardly to more complicated models. Thus, in our study, we take a different approach.

Our idea is to interpret the left-hand side of equation (2.3) as a function of η_{t-1}. Considering that σ_{t-1} is given, for example, as constant σ, we introduce the continuous function:

$$g(x, \sigma) \ = \ \omega^\circ - \omega + \sigma^2 \left\{ \alpha^\circ \left(|x| - \gamma^\circ x\right)^2 - \alpha \left(|x| - \gamma x\right)^2 + \beta^\circ - \beta \right\} .$$

Since (2.3) implies $g(\eta_{t-1}, \sigma_{t-1}) = 0$ a.s., it follows that $g(x, \sigma) = 0$ for all $(x, \sigma) \in \mathrm{supp}(\eta_{t-1}, \sigma_{t-1})$. Further, owing to the independence of η_{t-1} and σ_{t-1}, we have

$g(x, \sigma) = 0$ for all $(x, \sigma) \in \text{supp}(\eta_{t-1}) \times \text{supp}(\sigma_{t-1})$. This, in turn, implies

$$P\Big\{g(x, \sigma_{t-1}) = 0 \ \text{ for all } \ x \in \text{supp}(\eta_{t-1})\Big\} \ = \ 1 \, . \tag{2.4}$$

Assume that $\text{supp}(\eta_{t-1}) = \mathbb{R}$; in fact, it is sufficient to assume that $\text{supp}(\eta_{t-1})$ comprises three distinct (one positive and one negative) real numbers. Then, $g(x, \sigma_{t-1}) = 0$ a.s. for all $x \in \mathbb{R}$ and, particularly $g(0, \sigma_{t-1}) = \omega^\circ - \omega + \sigma_{t-1}^2 \, (\beta^\circ - \beta) = 0$ a.s., which leads to $\beta = \beta^\circ$ and $\omega = \omega^\circ$ owing to the nondegeneracy of σ_{t-1}^2. Henceforth, the equation $g(x, \sigma_{t-1}) = 0$ a.s. $\forall x \in \mathbb{R}$ is now reduced to

$$\alpha^\circ \, (|x| - \gamma^\circ x)^2 - \alpha \, (|x| - \gamma x)^2 \ = \ 0 \, , \qquad \forall x \in \mathbb{R} \, , \tag{2.5}$$

and thus, $\theta = \theta^\circ$ is derived. This AGARCH(1, 1) example demonstrates that equation (2.4) plays a crucial role in obtaining the conditions to guarantee the identifiability of a time series model. Later, to obtain the desired results for general nonlinear time series models, such as STGARCH, nonlinear Poisson autoregressive, and multiple regime STAR models, we will often apply the equations analogous to (2.4) and results such as $P\left\{\lim_{x \to \infty} g(x, \sigma_{t-1}) = 0\right\} = 1$ or $P\left\{\lim_{x \to -\infty} x^{-2} g(x, \sigma_{t-1}) = 0\right\} = 1$, as seen in the proof of Theorem 3.1.

3. IDENTIFIABILITY IN NONLINEAR TIME SERIES

3.1. Smooth transition GARCH models

González-Rivera [11] introduced the STGARCH(p, q, d) model:

$$\begin{aligned} X_t \ &= \ \sigma_t \eta_t \, , \\ \sigma_t^2 \ &= \ \omega^\circ + \sum_{i=1}^{q} \alpha_{1i}^\circ X_{t-i}^2 + \left(\sum_{i=1}^{q} \alpha_{2i}^\circ X_{t-i}^2 \right) F(X_{t-d}, \gamma^\circ) + \sum_{j=1}^{p} \beta_j^\circ \sigma_{t-j}^2 \, , \end{aligned} \tag{3.1}$$

where $\{\eta_t\}$ is the same as that in Model (2.1),

$$F(X_{t-d}, \gamma^\circ) \ = \ \frac{1}{1 + e^{\gamma^\circ X_{t-d}}} - \frac{1}{2} \, ,$$

$d \in \{1, ..., q\}$ is pre-specified, and $\gamma^\circ > 0$ is the smoothness parameter that determines the speed of transition. It is noteworthy that when $\gamma^\circ \to \infty$, the STGARCH(1, 1, 1) model becomes a GJR-GARCH(1, 1) model proposed by Glosten *et al.* [10], which is identical to Model (2.1).

We denote the true parameter vector by $\theta^\circ = (\gamma^\circ, \omega^\circ, \alpha_{11}^\circ, ..., \alpha_{1q}^\circ, \alpha_{21}^\circ, ..., \alpha_{2q}^\circ, \beta_1^\circ, ..., \beta_p^\circ)^T$. Let $\Theta = [0,\infty) \times (0,\infty) \times A \times B$ be the parameter space, where

$$A = \left\{(\alpha_{11}, ..., \alpha_{1q}, \alpha_{21}, ..., \alpha_{2q}) \in \mathbb{R}^{2q} : \alpha_{1i} \geq 0,\ |\alpha_{2i}| \leq 2\alpha_{1i}, \forall i\right\},$$

$$B = \left\{(\beta_1, ..., \beta_p) \in [0,1)^p : \sum_{j=1}^{p} \beta_j < 1\right\}, \tag{3.2}$$

and assume that $\theta^\circ \in \Theta$ for the conditional variance to be positive.

Sufficient conditions to ensure the existence of a stationary solution for Model (3.1) are not specified in the literature. For instance, Straumann–Mikosch [29] and Meitz and Saikkonen [23] derived such conditions only for general GARCH-type models. However, for example, it can be seen that the STGARCH$(1,1,1)$ model is stationary when $E\left[\log\left\{\beta_1^\circ + \left(\alpha_{11}^\circ + \frac{1}{2}|\alpha_{21}^\circ|\right)\eta_{t-1}^2\right\}\right] < 0$ (cf. Example 4 and Table 1 of Meitz and Saikkonen [23]).

Given the stationary solution $\{X_t\}$ and a parameter vector $\theta \in \Theta$, we define

$$c_t(\alpha) = \omega + \sum_{i=1}^{q} \alpha_{1i} X_{t-i}^2 + \left(\sum_{i=1}^{q} \alpha_{2i} X_{t-i}^2\right) F(X_{t-d}, \gamma)\,,$$

where $\alpha = (\gamma, \omega, \alpha_{11}, ..., \alpha_{1q}, \alpha_{21}, ..., \alpha_{2q})$. Note that the polynomial $\beta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ has all its zeros outside the unit disc because of (3.2). Define $\sigma_t^2(\theta) = \beta(B)^{-1} c_t(\alpha)$, where B is the backshift operator. Then, we have the following.

Theorem 3.1. *Let $\{X_t\}$ be a stationary process satisfying (3.1) and suppose that*

(**a**) *$\alpha_{2i}^\circ \neq 0$ for some $1 \leq i \leq q$ and $\gamma^\circ > 0$.*

(**b**) *The support of the distribution of η_1 is $\mathbb{R}$.*

Then, if $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, we have $\theta = \theta^\circ$.

Remark 3.1. It is remarkable that the identifiability in the STGARCH models needs no restriction concerning orders p and q. The above theorem shows that the STGARCH(p,q,d) models can be consistently estimated by fitting any STGARCH(p^*,q^*,d) models with $p^* \geq p$ and $q^* \geq q$. However, this is not true for GARCH and AGARCH models, wherein conditions such as (c) in Theorem 3.2 below are necessary. See Francq and Zakoïan [8] and Straumann and Mikosch [29].

Remark 3.2. As pointed out by a referee, the common root condition for the STGARCH models is not required owing to the reasons described below. Consider a STGARCH$(0,1,d)$ model and let σ_t^2 be the conditional variance. Multiplying $(1-\beta B)$ to both sides of the volatility equation, we get $(1-\beta B)\sigma_t^2 = (1-\beta)\omega + \alpha_{11} X_{t-1}^2 - \beta\alpha_{11} X_{t-2}^2 + \alpha_{21} X_{t-1}^2 F(X_{t-d}, \gamma) - \beta\alpha_{21} X_{t-2}^2 F(X_{t-d-1}, \gamma)$.

This, however, is not expressible as a form of STGARCH$(1, 2, d)$ models, unlike we see in GARCH and AGARCH models.

Remark 3.3. As in the case of the AGARCH model in Section 2, the support needs not be $\mathbb{R}$. For example, $\text{supp}(\eta_1) = \mathbb{Z}$ is sufficient.

Condition (a) in Theorem 3.1 suggests that there exists a smooth transition mechanism, that is, conditional variances asymmetrically respond to positive and negative news. When it fails, the STGARCH model becomes a standard GARCH model. The following theorem demonstrates that model parameters in (3.1) are only partially identified when no such transition mechanism exists.

Theorem 3.2. *Let $\{X_t\}$ be a stationary process satisfying (3.1) with $\gamma^\circ = 0$ or $\alpha_{2i}^\circ = 0$, $i = 1, ..., q$. Suppose that condition (b) in Theorem 3.1 and the following condition hold:*

(**c**) *$\alpha_{1i}^\circ > 0$ for some $1 \le i \le q$, $(\alpha_{1q}^\circ, \beta_p^\circ) \neq (0, 0)$, and the polynomials $\alpha_1^\circ(z) = \sum_{i=1}^q \alpha_{1i}^\circ z^i$ and $\beta^\circ(z) = 1 - \sum_{j=1}^p \beta_j^\circ z^j$ have no common zeros.*

If $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, then $\omega = \omega^\circ$, $\alpha_{1i} = \alpha_{1i}^\circ$, $\beta_j = \beta_j^\circ$ for $1 \le i \le q$, $1 \le j \le p$, and either $\gamma = 0$ or $\alpha_{2i} = 0$, $1 \le i \le q$ holds.

Remark 3.4. The hypothesis testing of whether the smoothness mechanism exists has been studied by González-Rivera [11]. This is a type of testing problem wherein nuisance parameters are unidentifiable under the null hypothesis. In addition, inference in a similar situation has been studied by Hansen [12] and Francq *et al.* [7].

3.2. Threshold Poisson autoregressive models

Poisson autoregressive models (or integer-valued GARCH models) are used to model time series of counts with over-dispersion and have been widely applied in fields ranging from finance to epidemiology to estimate, for example, the number of transactions per minute of certain stocks and the daily epileptic seizure counts of patients. See Fokianos *et al.* [5], Kang and Lee [15], and the references therein.

Let $\{X_t : t \ge 0\}$ be a time series of counts and $\{\lambda_t : t \ge 0\}$ its intensity process. Let $\mathcal{F}_{0,t}$ denote the σ-field generated $\{\lambda_0, X_0, ..., X_t\}$. An integer-valued threshold GARCH (INTGARCH) model is then defined by

$$\begin{aligned} &X_t|\mathcal{F}_{0,t-1} \sim \text{Poisson}(\lambda_t)\ , \\ &\lambda_t = \omega^\circ + \alpha_1^\circ X_{t-1} + (\alpha_2^\circ - \alpha_1^\circ)(X_{t-1} - l^\circ)^+ + \beta^\circ \lambda_{t-1}\ , \end{aligned} \tag{3.3}$$

for $t \geq 1$, where a^+ denotes $\max\{0, a\}$. We assume that the true parameter vector $\theta^\circ = (\omega^\circ, \alpha_1^\circ, \alpha_2^\circ, \beta^\circ, l^\circ)$ belongs to a parameter space $\Theta = (0, \infty) \times [0, 1)^3 \times \mathbb{N}$. Theorem 2.1 of Neumann [25] indicates that if $\beta^\circ + \max\{\alpha_1^\circ, \alpha_2^\circ\} < 1$, there exists a unique stationary bivariate process $\{(X_t, \lambda_t) : t \geq 0\}$ satisfying (3.3). Then, the time domain can be extended from $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ to $\mathbb{Z}$. Franke *et al.* [9] considered the conditional LS estimation in these models.

Given the stationary process $\{X_t : t \in \mathbb{Z}\}$ and a parameter vector $\theta \in \Theta$, we define a stationary process $\{\lambda_t(\theta)\}$ as the solution of

$$\lambda_t(\theta) \,=\, \omega + \alpha_1 X_{t-1} + (\alpha_2 - \alpha_1)(X_{t-1} - l)^+ + \beta \lambda_{t-1}(\theta) \;, \qquad t \in \mathbb{Z} \;.$$

Then, we have the following.

Theorem 3.3. *Suppose that $\{X_t : t \in \mathbb{Z}\}$ is a stationary process satisfying (3.3) and $\alpha_1^\circ \neq \alpha_2^\circ$. Then, if $\lambda_t = \lambda_t(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, we have $\theta = \theta^\circ$.*

Remark 3.5. When $\alpha_1^\circ = \alpha_2^\circ > 0$, Model (3.3) becomes an integer-valued GARCH(1, 1) model. In this case, it can be seen that parameters, except the threshold parameter l, are identifiable.

3.3. General Poisson autoregressive models

Neumann [25] considered a class of nonlinear Poisson autoregressive models $\{X_t : t \in \mathbb{Z}\}$ of counts with intensity process $\{\lambda_t : t \in \mathbb{Z}\}$ such as

$$X_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \;, \qquad \lambda_t = f(\lambda_{t-1}, X_{t-1}, \theta^\circ) \;, \tag{3.4}$$

for some known function $f : [0, \infty) \times \mathbb{N}_0 \times \Theta \to [0, \infty)$. According to Theorems 2.1 and 3.1 of Neumann [25], when $f(\cdot, \theta^\circ)$ satisfies the following contractive condition:

$$|f(\lambda, y, \theta^\circ) - f(\lambda', y', \theta^\circ)| \,\leq\, \kappa_1 |\lambda - \lambda'| + \kappa_2 |y - y'| \;, \qquad \forall \lambda, \lambda' \geq 0 \;, \;\; \forall y, y' \in \mathbb{N}_0 \;,$$

where $\kappa_1, \kappa_2 \geq 0$ and $\kappa_1 + \kappa_2 < 1$, there exists a stationary process $\{(X_t, \lambda_t)\}$ with $\lambda_t \in \mathcal{F}_{t-1}$ satisfying (3.4). Further, in view of Theorem 3.1 in Neumann [25], one can define a stationary process $\{\lambda_t(\theta)\}$ satisfying

$$\lambda_t(\theta) = f(\lambda_{t-1}(\theta), X_{t-1}, \theta) \;, \qquad \forall t \in \mathbb{Z} \;,$$

for the stationary process $\{X_t\}$ and parameter vector $\theta \in \Theta$. Fokianos and Tjøstheim [6] studied ML estimation in these models.

The following theorem presents the mild requirements of f for their identifiability assumptions. Its proof is straightforward in view of the proof of Theorem 3.3.

Theorem 3.4. *Let $\{(X_t, \lambda_t)\}$ be a stationary process satisfying (3.4) and suppose that*

(**a**) *For each $\theta \in \Theta$, $f(\cdot, \theta)$ is continuous on* $\mathrm{supp}(\lambda_1) \times \mathbb{N}_0$.

(**b**) $f(\lambda, y, \theta) = f(\lambda, y, \theta^\circ),\ \forall \lambda \in \mathrm{supp}(\lambda_1),\ \forall y \in \mathbb{N}_0$ *implies* $\theta = \theta^\circ$.

Then, if $\lambda_t = \lambda_t(\theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, $\theta = \theta^\circ$.

3.4. Multiple regime smooth transition autoregressive models

Regime switching models for financial data have received considerable attention. For example, Teräsvirta [32] studied inference for two-regime STAR models and McAleer and Medeiros [21] and Li and Ling [20] considered multiple-regime smooth transition and threshold AR models. In this subsection, we consider the nonlinear LS estimation in a multiple-regime STAR model with heteroscedastic errors proposed by McAleer and Medeiros [21].

Suppose that $\{X_t\}$ follows a multiple-regime STAR model of order p with $M+1$ (limiting) regimes, that is,

$$X_t = \beta_0^{\circ T}\mathbf{X}_{t-1} + \sum_{i=1}^{M} \beta_i^{\circ T}\mathbf{X}_{t-1} G(X_{t-d^\circ}; \gamma_i^\circ, c_i^\circ) + \varepsilon_t\ , \tag{3.5}$$

where $\{\varepsilon_t\}$ is white noise, $\beta_i^\circ = (\phi_{i0}^\circ, \phi_{i1}^\circ, ..., \phi_{ip}^\circ)^T$ for $0 \le i \le M$, $\mathbf{X}_{t-1} = (1, X_{t-1}, ..., X_{t-p})^T$, and $G(X_{t-d^\circ}; \gamma_i^\circ, c_i^\circ)$ is a logistic transition function given by

$$G(X_{t-d^\circ}; \gamma_i^\circ, c_i^\circ) = \frac{1}{1 + e^{-\gamma_i^\circ (X_{t-d^\circ} - c_i^\circ)}}\ , \tag{3.6}$$

wherein the regime switches according to the value of transition variable X_{t-d°: $d^\circ \in \{1, ..., p\}$ is a delay parameter, $-\infty < c_1^\circ < \cdots < c_M^\circ < \infty$ are threshold parameters, and $\gamma_i^\circ > 0$, $i = 1, ..., M$, are smoothing parameters. When γ_i° is quite large, Model (3.5) is barely distinguishable from the threshold model studied by Li and Ling [20].

In the literature, one can find sufficient conditions under which Model (3.5) is stationary when the error terms are iid. For example, Theorem 2 of McAleer and Medeiros [21] ensures the stationarity of Model (3.5) of order 1. Using the same reasoning and Lemma 2.1 of Berkes *et al.* [2], we can see that Model (3.5)

has a stationary solution if

$$\sum_{j=1}^{p} \sup_{x\in\mathbb{R}} \left| \phi_{0j}^{\circ} + \sum_{i=1}^{M} \phi_{ij}^{\circ} G(x; \gamma_i^{\circ}, c_i^{\circ}) \right| < 1 \, .$$

It is also true if $\max_{0\le i\le M} \sum_{j=1}^{p} \left| \sum_{k=0}^{i} \phi_{kj}^{\circ} \right| < 1$, which can be deduced from Theorem 3.2 and Example 3.6 in An and Huang [1].

We denote by $\theta = (\beta_0^T, \beta_1^T, ..., \beta_M^T, \gamma_1, ..., \gamma_M, c_1, ..., c_M, d)^T$ a parameter vector belonging to a parameter space $\Theta \subset \mathbb{R}^{(M+1)(p+1)+2M} \times \{1, ..., p\}$ and set

$$m(X_{t-1}, ..., X_{t-p}, \theta) \; = \; {\beta_0}^T \mathbf{X}_{t-1} + \sum_{i=1}^{M} {\beta_i}^T \mathbf{X}_{t-1} G(X_{t-d}; \gamma_i, c_i) \, .$$

Then, we have the following.

Theorem 3.5. *Let $\{X_t\}$ be a stationary process satisfying (3.5). Assume that*

- (**a**) *For each $i = 1, ..., M$, $\beta_i^{\circ} \neq (0, ..., 0)^T \in \mathbb{R}^{p+1}$.*
- (**b**) *The support of the stationary distribution of $(X_p, ..., X_1)$ is $\mathbb{R}^p$.*
- (**c**) *The parameter space Θ satisfies that $\gamma_i > 0$, $i = 1, ..., M$, and $-\infty < c_1 < \cdots < c_M < \infty$.*

Then, if $m(X_{t-1}, ..., X_{t-p}, \theta^{\circ}) = m(X_{t-1}, ..., X_{t-p}, \theta)$ a.s. for some $t \in \mathbb{Z}$ and $\theta \in \Theta$, we have $\theta = \theta^{\circ}$.

Remark 3.6. Theorem 3.5 is closely related to the identifiability of the finite mixture of logistic distributions (see Lemma 4.1 in Section 4). Although the restriction on threshold parameters has a natural interpretation, it is not necessarily required. In fact, if we only assume that $(\gamma_i^{\circ}, c_i^{\circ}), i = 1, ..., M$, are distinct, instead of the condition $c_i^{\circ} < c_{i+1}^{\circ}$, then Model (3.5) is weakly identifiable in the sense of Redner and Walker [27].

4. PROOFS

Proof of Theorem 3.1: We only prove the theorem when $d = 1$ since the other cases can be handled similarly. Owing to the stationarity, we have $\sigma_t^2 = \sigma_t^2(\theta)$ a.s. for any $t \in \mathbb{Z}$. Since $\beta^{\circ}(z) \neq 0$ for $|z| \le 1$ and $\sigma_t^2 = \beta^{\circ}(B)^{-1} c_t(\alpha^{\circ})$, we can express

$$c_t(\alpha) \; = \; \beta(B)\sigma_t^2(\theta) \; = \; \beta(B)\beta^{\circ}(B)^{-1} c_t(\alpha^{\circ}) \; = \; c_t(\alpha^{\circ}) + \sum_{j=1}^{\infty} b_j c_{t-j}(\alpha^{\circ}) \, , \tag{4.1}$$

where $1+\sum_{j=1}^{\infty} b_j z^j = \beta(z)/\beta^\circ(z)$ for $|z| \le 1$. As discussed in Section 2, we can express (4.1) as a function of η_{t-1} and $\mathcal{F}_{t-2}$-measurable random variables:

$$\begin{aligned} & g_1(\eta_{t-1}, \sigma_{t-1}, A_{t,2}, B_{t,2}, A^\circ_{t,2}, B^\circ_{t,2}, D_{t,2}) \ := \\ & \quad := \ (\alpha_{11} - \alpha^\circ_{11})\sigma^2_{t-1}\eta^2_{t-1} + A_{t,2} - A^\circ_{t,2} + \big(\alpha_{21}\sigma^2_{t-1}\eta^2_{t-1} + B_{t,2}\big) F(\sigma_{t-1}\eta_{t-1}, \gamma) \\ & \qquad - \big(\alpha^\circ_{21}\sigma^2_{t-1}\eta^2_{t-1} + B^\circ_{t,2}\big) F(\sigma_{t-1}\eta_{t-1}, \gamma^\circ) - D_{t,2} \\ & \quad = \ 0 \quad \text{a.}s., \end{aligned}$$

where for $2 \le i^* \le q$ and $2 \le k$,

$$A_{t,i^*} = \omega + \sum_{i=i^*}^{q} \alpha_{1i} X^2_{t-i}\,, \qquad B_{t,i^*} = \sum_{i=i^*}^{q} \alpha_{2i} X^2_{t-i}\,, \quad D_{t,k} = \sum_{j=k-1}^{\infty} b_j c_{t-j}(\alpha^\circ)\,,$$

$$A^\circ_{t,i^*} = \omega^\circ + \sum_{i=i^*}^{q} \alpha^\circ_{1i} X^2_{t-i}\,, \quad B^\circ_{t,i^*} = \sum_{i=i^*}^{q} \alpha^\circ_{2i} X^2_{t-i}\,.$$

Using the arguments that obtain (2.4) and condition (b), we can see that with probability 1, $g_1(x, \sigma_{t-1}, A_{t,2}, B_{t,2}, A^\circ_{t,2}, B^\circ_{t,2}, D_{t,2}) = 0$ for all $x \in \mathbb{R}$. Particularly, this implies

$$g_1(0, \sigma_{t-1}, A_{t,2}, B_{t,2}, A^\circ_{t,2}, B^\circ_{t,2}, D_{t,2}) \ = \ A_{t,2} - A^\circ_{t,2} - D_{t,2} \ = \ 0 \quad \text{a.s..} \tag{4.2}$$

Then, viewing (4.2) as a function of η_{t-2} and $\mathcal{F}_{t-3}$-measurable random variables, we can express

$$\begin{aligned} & g_2(\eta_{t-2}, \sigma_{t-2}, A_{t,3}, A^\circ_{t,3}, A^\circ_{t-1,2}, B^\circ_{t-1,2}, D_{t,3}) \ := \\ & \quad := \ (\alpha_{12} - \alpha^\circ_{12})\sigma^2_{t-2}\eta^2_{t-2} + A_{t,3} - A^\circ_{t,3} - b_1 c_{t-1}(\alpha^\circ) - D_{t,3} \\ & \quad = \ (\alpha_{12} - \alpha^\circ_{12})\sigma^2_{t-2}\eta^2_{t-2} + A_{t,3} - A^\circ_{t,3} - D_{t,3} \\ & \qquad - b_1\big\{\alpha^\circ_{11}\sigma^2_{t-2}\eta^2_{t-2} + A^\circ_{t-1,2} + \big(\alpha^\circ_{21}\sigma^2_{t-2}\eta^2_{t-2} + B^\circ_{t-1,2}\big) F(\sigma_{t-2}\eta_{t-2}, \gamma^\circ)\big\} \\ & \quad = \ 0 \quad \text{a.s.}\,, \end{aligned} \tag{4.3}$$

which entails

$$P\Big(g_2(x, \sigma_{t-2}, A_{t,3}, A^\circ_{t,3}, A^\circ_{t-1,2}, B^\circ_{t-1,2}, D_{t,3}) = 0,\ \forall x \in \mathbb{R}\Big) \ = \ 1\,. \tag{4.4}$$

Note that if

$$f(x) \ := \ ax^2 + b + (cx^2 + d)F(\sigma x, \gamma^\circ) \ = \ 0 \tag{4.5}$$

for all $x \in \mathbb{R}$, where a, b, c, d, $\sigma > 0$, $\gamma^\circ > 0$ are real numbers, because $\lim_{x\to\pm\infty} x^{-2} f(x) = 0$ and $\lim_{x\to\pm\infty} f(x) = 0$, it must hold that $a = c = 0$ and $b = d = 0$. Then, combining this and (4.4), we get $b_1\alpha^\circ_{21} = 0$ and $b_1 B^\circ_{t-1,2} = 0$ a.s.. Further, $B^\circ_{t-1,2} = 0$ a.s. if and only if $\alpha^\circ_{22} = \cdots = \alpha^\circ_{2q} = 0$. Due to condition (a) and (4.3), we have $b_1 = 0$ and $A_{t,3} - A^\circ_{t,3} - D_{t,3} = 0$ a.s., and similarly, it can be seen that $b_k = 0$, $k \ge 2$, $A_{t,k+2} - A^\circ_{t,k+2} - D_{t,k+2} = 0$ a.s., $2 \le k \le q-2$, and $\omega - \omega^\circ - D_{t,k+2} = 0$ a.s., $k \ge q-1$. This implies $\beta(\cdot) = \beta^\circ(\cdot)$, $\omega = \omega^\circ$, and

$A_{t,2} = A^\circ_{t,2}, ..., A_{t,q} = A^\circ_{t,q}$ a.s., and subsequently, $\alpha_{1q} = \alpha^\circ_{1q}, ..., \alpha_{12} = \alpha^\circ_{12}$. From this and (4.1), we can obtain

$$
\begin{aligned}
& h_1(\eta_{t-1}, \sigma_{t-1}, B_{t,2}, B^\circ_{t,2}) \ := \\
& \qquad := \ (\alpha_{11} - \alpha^\circ_{11})\sigma^2_{t-1}\eta^2_{t-1} + \big(\alpha_{21}\sigma^2_{t-1}\eta^2_{t-1} + B_{t,2}\big)\, F(\sigma_{t-1}\eta_{t-1}, \gamma) \\
& \qquad\quad - \ \big(\alpha^\circ_{21}\sigma^2_{t-1}\eta^2_{t-1} + B^\circ_{t,2}\big)\, F(\sigma_{t-1}\eta_{t-1}, \gamma^\circ) \\
& \qquad = \ 0 \quad \text{a.}s.
\end{aligned} \tag{4.6}
$$

Suppose that $\gamma = 0$. Then, $F(X_{t-1}, \gamma) \equiv 0$, and using (4.5) and (4.6), we get $\alpha^\circ_{21} = 0$ and $B^\circ_{t,2} = 0$ a.s. Since this is a contradiction to condition (a), γ must be positive. Thus, from (4.6), we have

$$\lim_{x\to\infty} x^{-2} h_1(x, \sigma_{t-1}, B_{t,2}, B^\circ_{t,2}) \ = \ \sigma^2_{t-1}\left\{\alpha_{11} - \alpha^\circ_{11} - 2^{-1}(\alpha_{21} - \alpha^\circ_{21})\right\} \ = \ 0 \quad \text{a.s..}$$

Further, taking the limit $x \to -\infty$, we obtain $\alpha_{11} = \alpha^\circ_{11}$ and $\alpha_{21} = \alpha^\circ_{21}$, so that

$$\lim_{x\to\infty} h_1(x, \sigma_{t-1}, B_{t,2}, B^\circ_{t,2}) \ = \ -2^{-1}B_{t,2} + 2^{-1}B^\circ_{t,2} = 0 \quad \text{a.s.}\,,$$

which results in $\alpha_{2i} = \alpha^\circ_{2i},\ 2 \le i \le q$. Then, in view of (4.6), we obtain

$$
\begin{aligned}
h_2(\eta_{t-1}, \sigma_{t-1}, B^\circ_{t,2}) \ &:= \ \big(\alpha^\circ_{21}\sigma^2_{t-1}\eta^2_{t-1} + B^\circ_{t,2}\big)\left(\frac{1}{1 + e^{\gamma\sigma_{t-1}\eta_{t-1}}} - \frac{1}{1 + e^{\gamma^\circ\sigma_{t-1}\eta_{t-1}}}\right) \\
&= \ 0 \quad \text{a.s.}\,.
\end{aligned}
$$

If $\gamma < \gamma^\circ$ and additionally if $\alpha^\circ_{21} \neq 0$, we should have

$$\lim_{x\to\infty} x^{-2} e^{\gamma\sigma_{t-1}x} h_2(x, \sigma_{t-1}, B^\circ_{t,2}) \ = \ \alpha^\circ_{21}\sigma^2_{t-1} \ = \ 0 \quad \text{a.s.}\,,$$

which leads to a contradiction. However, if $\alpha^\circ_{21} = 0$, we have $\lim_{x\to\infty} e^{\gamma\sigma_{t-1}x} h_2(x, \sigma_{t-1}, B^\circ_{t,2}) = B^\circ_{t,2} = 0$ a.s., which also leads to a contradiction to condition (a). Hence, we must have $\gamma \ge \gamma^\circ$. Since $\gamma > \gamma^\circ$ is also impossible, we conclude that $\gamma = \gamma^\circ$, which completes the proof. □

Proof of Theorem 3.2: As in handling (4.3), we follow the same lines in the proof of Theorem 3.1 to obtain

$$
\begin{aligned}
& g'_2(\eta_{t-2}, \sigma_{t-2}, A_{t,3}, A^\circ_{t,3}, A^\circ_{t-1,2}, D_{t,3}) \ := \\
& \qquad := \ (\alpha_{12} - \alpha^\circ_{12} - b_1\alpha^\circ_{11})\sigma^2_{t-2}\eta^2_{t-2} + A_{t,3} - A^\circ_{t,3} - b_1 A^\circ_{t-1,2} - D_{t,3} \\
& \qquad = \ 0 \quad \text{a.s.}\,.
\end{aligned}
$$

Then, as in handling (4.2), we get $A_{t,3} - A^\circ_{t,3} - b_1 A^\circ_{t-1,2} - D_{t,3} = 0$ a.s. Similarly, it can be seen that $\omega - \omega^\circ - b_1 A^\circ_{t-1,q} - b_2 A^\circ_{t-2,q-1} - \cdots - b_{q-1} A^\circ_{t-q+1,2} - D_{t,q+1} = 0$ a.s. Then, with probability 1, for all $x \in \mathbb{R}$,

$$
\begin{aligned}
g(x) \ := \ & (\omega - \omega^\circ) - b_1\big(\alpha^\circ_{1q}\sigma^2_{t-q-1}x^2 + \omega^\circ\big) - b_2\big(\alpha^\circ_{1,q-1}\sigma^2_{t-q-1}x^2 + A^\circ_{t-2,q}\big) - \cdots \\
& - b_{q-1}\big(\alpha^\circ_{12}\sigma^2_{t-q-1}x^2 + A^\circ_{t-q+1,3}\big) - b_q\big(\alpha^\circ_{11}\sigma^2_{t-q-1}x^2 + A^\circ_{t-q,2}\big) - D_{t,q+2} \\
= \ & 0\,,
\end{aligned} \tag{4.7}
$$

which, in turn, implies

$$P\left(\lim_{x\to\infty}\frac{-g(x)}{\sigma_{t-q-1}^2x^2}=b_1\alpha_{1q}^\circ+\cdots+b_q\alpha_{11}^\circ=0\right)=1\,.$$

In fact, we can obtain an analogous relationship between η_{t-q-k} and $\mathcal{F}_{t-q-k-1}$-measurable random variables, $k\geq 2$, and as such, $b_k\alpha_{1q}^\circ+\cdots+b_{k+q-1}\alpha_{11}^\circ=0$ for all $k\geq 1$, which implies that $\beta(z)\beta^\circ(z)^{-1}\alpha_1^\circ(z)$ is a polynomial of at most q orders. Then, using condition (c) and the arguments similar to those in Straumann and Mikosch (2006), p. 2481, we can see that $\beta(\cdot)=\beta^\circ(\cdot)$, and thus, $b_j=0$ for $j\geq 1$. Combining this, (4.2) and (4.7), we get $A_{t,2}=A_{t,2}^\circ,...,A_{t,q}=A_{t,q}^\circ$ a.s. and $\omega=\omega^\circ$, which, in turn, implies $\alpha_{1q}=\alpha_{1q}^\circ,...,\alpha_{12}=\alpha_{12}^\circ$. Hence, (4.1) can be reexpressed as

$$\begin{aligned}h_1'(\eta_{t-1},\sigma_{t-1},B_{t,2}) &:= (\alpha_{11}-\alpha_{11}^\circ)\sigma_{t-1}^2\eta_{t-1}^2+\left(\alpha_{21}\sigma_{t-1}^2\eta_{t-1}^2+B_{t,2}\right)F(\sigma_{t-1}\eta_{t-1},\gamma)\\ &= 0 \quad \text{a.}s.\,.\end{aligned}$$

From this, we can easily obtain $\alpha_{11}=\alpha_{11}^\circ$ and the same equation as in (4.5), which finally leads to $\alpha_{21}=\cdots=\alpha_{2q}=0$. This completes the proof. □

Proof of Theorem 3.3: First, we conjecture that the support of the stationary distribution of (X_1,λ_1) is a Cartesian product of $\mathbb{N}_0$ and $\text{supp}(\lambda_1)$. If it is not true, there exists $(m',\lambda')\in\mathbb{N}_0\times\text{supp}(\lambda_1)$ such that $(m',\lambda')\notin\text{supp}(X_1,\lambda_1)$, and for some positive real number r,

$$0\;=\;P\left(X_1=m',\lambda_1\in(\lambda'-r,\lambda'+r)\right)\;=\;\int_{\lambda'-r}^{\lambda'+r}(m'!)^{-1}e^{-u}u^{m'}\mathrm{d}F_{\lambda_1}(u)\,,$$

where F_{λ_1} is the distribution function of λ_1. Since the integrand is positive, it must hold that $P(\lambda_1\in(\lambda'-r,\lambda'+r))=0$, which, however, contradicts to the fact that $\lambda'\in\text{supp}(\lambda_1)$. Thus, our conjecture is validated.

Note that owing to the stationarity, for all $t\in\mathbb{Z}$,

$$\begin{aligned}g(X_{t-1},\lambda_{t-1}) &:= (\omega-\omega^\circ)+(\alpha_1-\alpha_1^\circ)X_{t-1}+(\alpha_2-\alpha_1)(X_{t-1}-l)^+\\ &\quad-(\alpha_2^\circ-\alpha_1^\circ)(X_{t-1}-l^\circ)^+-(\beta-\beta^\circ)\lambda_{t-1}\\ &= 0 \quad \text{a.s.}\,,\end{aligned}$$

and therefore,

$$g(m,\lambda)=0 \qquad \text{for all} \quad m\in\mathbb{N}_0 \;\text{ and }\; \lambda\in\text{supp}(\lambda_1)\,, \tag{4.8}$$

since $g(\cdot)$ is continuous and $\text{supp}(X_1,\lambda_1)=\mathbb{N}_0\times\text{supp}(\lambda_1)$. In particular, $g(0,\lambda)=(\omega-\omega^\circ)+(\beta-\beta^\circ)\lambda=0$ for any $\lambda\in\text{supp}(\lambda_1)$. Note that λ_t is not degenerate when $\alpha_1^\circ\neq\alpha_2^\circ$, since otherwise, X_{t-1} should be degenerate. Thus, we have $\omega=\omega^\circ$ and $\beta=\beta^\circ$, so that $g(1,\lambda)=\alpha_1-\alpha_1^\circ=0$. Further, it follows from (4.8) that $\lim_{m\to\infty}m^{-1}g(m,\lambda)=\alpha_2-\alpha_2^\circ=0$. Then, using the fact that $g(l,\lambda)=g(l^\circ,\lambda)=0$ and $\alpha_1^\circ\neq\alpha_2^\circ$, we obtain $l=l^\circ$, which completes the proof. □

Proof of Theorem 3.5: For simplicity, we assume that $d^\circ = 1$: the other cases can be handled similarly. From condition (b) and the continuity of $m(\cdot, \theta)$, we can see that

$$m(x_1, ..., x_p, \theta^\circ) = m(x_1, ..., x_p, \theta) \ , \qquad \forall x_j \in \mathbb{R} \, , \quad 1 \le j \le p \ . \tag{4.9}$$

Suppose that $d \neq 1$. From (4.9), we can express

$$\begin{aligned} m(x_1, ..., x_p, \theta^\circ) - m(x_1, ..., x_p, \theta) \ &= \\ &= \ \left\{ f_0^\circ(\mathbf{x}_2) - f_0(\mathbf{x}_2) - \sum_{i=1}^{M} f_i(\mathbf{x}_2) G(x_d; \gamma_i, c_i) \right\} \\ &\quad + \left\{ \phi_{01}^\circ - \phi_{01} - \sum_{i=1}^{M} \phi_{i1} G(x_d; \gamma_i, c_i) \right\} x_1 \\ &\quad + \sum_{i=1}^{M} \left(f_i^\circ(\mathbf{x}_2) + \phi_{i1}^\circ x_1 \right) G(x_1; \gamma_i^\circ, c_i^\circ) \\ &= \ 0 \, , \end{aligned} \tag{4.10}$$

where $G(\cdot)$ is the one in (3.6), $\mathbf{x}_2 = (x_2, ..., x_p)^T$, and

$$f_i^\circ(\mathbf{x}_2) = \phi_{i0}^\circ + \sum_{2 \le j \le p} \phi_{ij}^\circ x_j \, , \quad f_i(\mathbf{x}_2) = \phi_{i0} + \sum_{2 \le j \le p} \phi_{ij} x_j \, , \qquad \text{for} \quad i = 0, 1, ..., M \ .$$

Then, applying Lemma 4.1 below to (4.10), we have $\phi_{11}^\circ = 0$ and $f_1^\circ(\mathbf{x}_2) = 0$ for each $\mathbf{x}_2 \in \mathbb{R}^{p-1}$, which, however, contradicts to condition (a). Thus, it must hold that $d = d^\circ = 1$. Owing to the above, we can reexpress (4.9) as

$$\begin{aligned} \left(f_0^\circ(\mathbf{x}_2) + \phi_{01}^\circ x_1 \right) + \sum_{i=1}^{M} \left(f_i^\circ(\mathbf{x}_2) + \phi_{i1}^\circ x_1 \right) G(x_1; \gamma_i^\circ, c_i^\circ) \ &= \\ = \ \left(f_0(\mathbf{x}_2) + \phi_{01} x_1 \right) + \sum_{i=1}^{M} \left(f_i(\mathbf{x}_2) + \phi_{i1} x_1 \right) G(x_1; \gamma_i, c_i) \, , & \\ \forall x_j \in \mathbb{R} \, , \quad 1 \le j \le p \ . & \end{aligned} \tag{4.11}$$

Lemma 4.1 ensures that a family of real-valued functions $\mathcal{G} = \{1, i(\cdot)\} \cup \{G(\cdot; \gamma, c) : \gamma > 0, c \in \mathbb{R}\} \cup \{i(\cdot) G(\cdot; \gamma, c) : \gamma > 0, c \in \mathbb{R}\}$, where $i(\cdot)$ is an identity function, i.e., $i(y) = y$, are linearly independent. Thus, any element of the linear span of $\mathcal{G}$ is uniquely represented as a linear combination of the elements of $\mathcal{G}$: see Yakowitz and Spragins [34]. Further, there exists a vector $\mathbf{x}_2' \in \mathbb{R}^{p-1}$ such that $(f_i^\circ(\mathbf{x}_2'), \phi_{i1}^\circ) \neq (0, 0)$ for all $i = 1, ..., M$; unless otherwise, $\phi_{i0}^\circ = \cdots = \phi_{ip}^\circ = 0$ for some i, which contradicts condition (a). Then, viewing (4.11) with $\mathbf{x}_2$ substituted by $\mathbf{x}_2'$ as a function of x_1 and using condition (c), we obtain $\phi_{01}^\circ = \phi_{01}$ and $\phi_{i1}^\circ = \phi_{i1}$, $\gamma_i^\circ = \gamma_i$, $c_i^\circ = c_i$ for $i = 1, ..., M$. Subsequently, owing to (4.11), for all $x_1 \in \mathbb{R}$ and $\mathbf{x}_2 \in \mathbb{R}^{p-1}$, we get

$$\left(f_0^\circ(\mathbf{x}_2) - f_0(\mathbf{x}_2) \right) + \sum_{i=1}^{M} \left(f_i^\circ(\mathbf{x}_2) - f_i(\mathbf{x}_2) \right) G(x_1; \gamma_i^\circ, c_i^\circ) \ = \ 0 \, .$$

Then, applying Lemma 4.1 again, we conclude that $\phi^{\circ}_{i0} = \phi_{i0}$ and $\phi^{\circ}_{ij} = \phi_{ij}$, $j = 2, ..., p$, $i = 0, 1, ..., M$. This completes the proof. □

Lemma 4.1. *Let $(\gamma_1, c_1), ..., (\gamma_k, c_k)$ be distinct real vectors with $\gamma_i > 0$, $i = 1, ..., k$. Suppose that for all $y \in \mathbb{R}$,*

$$d_{00} + d_{01}y + \sum_{i=1}^{k}(d_{i0} + d_{i1}y)\frac{1}{1 + e^{-\gamma_i(y-c_i)}} \;=\; 0\,. \tag{4.12}$$

Then, $d_{i0} = d_{i1} = 0$ for $i = 0, 1, ..., k$.

Proof: Denote by $g(y)$ the left-hand side of (4.12). Then, $\lim_{y\to-\infty} y^{-1}g(y) = d_{01} = 0$, and thus, $\lim_{y\to-\infty} g(y) = d_{00} = 0$. In what follows, for function $f : \mathbb{R} \to \mathbb{R}$, we denote by $\mathcal{L}\{f\}$ its two-sided Laplace transform, that is, $\mathcal{L}\{f(\cdot)\}(s) = \int_{-\infty}^{\infty} e^{-sy} f(y)\mathrm{d}y$. Note that the transform of the logistic distribution function is as follows:

$$F_0(s;\gamma,c) \;:=\; \mathcal{L}\{G(\cdot;\gamma,c)\}(s) \;=\; \frac{\pi\gamma^{-1}e^{-cs}}{\sin\pi\gamma^{-1}s}\,, \qquad 0 < s < \gamma\,.$$

Further,

$$\begin{aligned} F_1(s;\gamma,c) \;&:=\; \mathcal{L}\{i(\cdot)G(\cdot;\gamma,c)\}(s) \\ &=\; \frac{\pi\gamma^{-1}ce^{-cs}}{\sin\pi\gamma^{-1}s} + \frac{\pi^2\gamma^{-2}e^{-cs}\cos\pi\gamma^{-1}s}{\sin^2\pi\gamma^{-1}s}\,, \qquad 0 < s < \gamma\,. \end{aligned}$$

Without loss of generality, assume that (γ_i, c_i), $i = 1, ..., k$, satisfy a lexicographical ordering, that is, $\gamma_i \le \gamma_{i+1}$ and $c_i < c_{i+1}$ when $\gamma_i = \gamma_{i+1}$. Suppose that $\gamma_1 = \cdots = \gamma_l < \gamma_{l+1} \le \cdots \le \gamma_k$ and $c_1 < \cdots < c_l$. Then, applying the two-sided Laplace transformation to (4.12), we have that for all $0 < s < \gamma_1$,

$$\sum_{i=1}^{k} d_{i0}F_0(s;\gamma_i,c_i) + \sum_{i=1}^{k} d_{i1}F_1(s;\gamma_i,c_i) \;=\; 0\,. \tag{4.13}$$

Since the numerator of the left-hand side of (4.13) is an analytic function on $\mathbb{R}$, (4.13) is still valid for all $s \in \mathbb{R}\backslash D$, where $D = \{s : s = \gamma_i m, 1 \le i \le k, m \in \mathbb{Z}\}$. Multiplying $\sin^2\pi\gamma_1^{-1}s$ to both the sides of (4.13), we attain

$$\begin{aligned} &\sin\pi\gamma_1^{-1}s\sum_{i=1}^{l}\left\{d_{i0}\pi\gamma_1^{-1}e^{-c_is} + d_{i1}\pi\gamma_1^{-1}c_ie^{-c_is}\right\} \;+ \\ &+\; \sin^2\pi\gamma_1^{-1}s\sum_{i=l+1}^{k}\left\{d_{i0}\frac{\pi\gamma_i^{-1}e^{-c_is}}{\sin\pi\gamma_i^{-1}s} + d_{i1}\frac{\pi\gamma_i^{-1}c_ie^{-c_is}}{\sin\pi\gamma_i^{-1}s}\right\} \\ &+\; \cos\pi\gamma_1^{-1}s\sum_{i=1}^{l}d_{i1}\pi^2\gamma_1^{-2}e^{-c_is} + \sin^2\pi\gamma_1^{-1}s\sum_{i=l+1}^{k}d_{i1}\frac{\pi^2\gamma_i^{-2}e^{-c_is}\cos\pi\gamma_i^{-1}s}{\sin^2\pi\gamma_i^{-1}s} \;=\; 0\,. \end{aligned}$$

Then, if we set $\mathbb{N}_1 = \{n \in \mathbb{N} : \gamma_1 n \neq \gamma_i m \text{ for all } l < i \leq k, m \in \mathbb{N}\}$, for any fixed $n \in \mathbb{N}_1$, letting $s \to \gamma_1 n$ through the values in $\mathbb{R}\backslash D$, we can have

$$\sum_{i=1}^{l} d_{i1} e^{-c_i \gamma_1 n} = 0 \,. \tag{4.14}$$

Since (4.14) holds for all $n \in \mathbb{N}_1$, multiplying $e^{c_1 \gamma_1 n}$ to both the sides of (4.14) and letting $n \to \infty$ through the values in $\mathbb{N}_1$, we get $d_{11} = 0$. Similarly, it can be seen that $d_{21} = \cdots = d_{l1} = 0$. Meanwhile, multiplying $\sin \pi \gamma_1^{-1} s$ to both the sides to (4.13) and letting $s \to \gamma_1 n$, we can have $\sum_{i=1}^{l} d_{i0} e^{-c_i \gamma_1 n} = 0$ for any $n \in \mathbb{N}_1$, and henceforth, $d_{10} = \cdots = d_{l0} = 0$. Continuing the above process, one can finally establish the lemma. □

Remark 4.1. Lemma 4.1 actually entails the identifiability of logistic mixture distributions (cf. Yakowitz and Spragins [34] and Sussmann [30]). Hwang and Ding [14] also proved the linear independence of logistic distributions and their density functions to deal with the identifiability problem in artificial neural networks. However, their results do not directly imply Lemma 4.1. Our proof is simpler and is based on Theorem 2 of Teicher [31].

ACKNOWLEDGMENTS

The authors thank the referees for their careful reading and valuable comments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A2A2A01046092).

REFERENCES

[1] An, H.Z. and Huang, F.C. (1996). The geometrical ergodicity of nonlinear autoregressive models, *Statist. Sinica*, **6**, 943–956.

[2] Berkes, I.; Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation, *Bernoulli*, **9**, 201–227.

[3] Chan, K.S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *Ann. Statist.*, **21**, 520–533.

[4] Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models, *J. Time Ser. Anal.*, **7**, 179–190.

[5] Fokianos, K.; Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression, *J. Amer. Statist. Assoc.*, **104**, 1430–1439. With electronic supplementary materials available online.

[6] Fokianos, K. and Tjøstheim, D. (2012). Nonlinear Poisson autoregression, *Ann. Inst. Statist. Math.*, **64**, 1205–1225.

[7] Francq, C.; Horvath, L. and Zakoïan, J.-M. (2010). Sup-tests for linearity in a general nonlinear AR(1) model, *Economet. Theor.*, **26**, 965–993.

[8] Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes, *Bernoulli*, **10**, 605–637.

[9] Franke, J.; Kirch, C. and Kamgaing, J. (2012). Changepoints in times series of counts, *J. Time Series Anal.*, **33**, 757–770.

[10] Glosten, L.; Jagannathan, R. and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *J. Finance*, **48**, 1779–1801.

[11] González-Rivera, G. (1998). Smooth-transition GARCH models, *Stud. Nonlinear Dyn. Econom.*, **3**, 61–78.

[12] Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413–430.

[13] Hayashi, F. (2000). *Econometrics*, Princeton University Press, Princeton, NJ.

[14] Hwang, J.T.G. and Ding, A.A. (1997). Prediction intervals for artificial neural networks, *J. Amer. Statist. Assoc.*, **92**, 748–757.

[15] Kang, J. and Lee, S. (2014). Parameter change test for Poisson autoregressive models, *Scand. J. Statist.* (to appear).

[16] Komunjer, I. (2012). Global identification in nonlinear models with moment restrictions, *Economet. Theor.*, **28**, 719–729.

[17] Kristensen, D. and Rahbek, A. (2009). Asymptotics of the QMLE for nonlinear ARCH models, *J. Time Ser. Econom.*, **1**, 38.

[18] Lee, S. and Lee, T. (2012). Inference for Box–Cox transformed threshold GARCH models with nuisance parameters, *Scand. J. Statist.*, **39**, 568–589.

[19] Lee, S. and Noh, J. (2013). Quantile regression estimator for GARCH models, *Scand. J. Stat.*, **40**, 2–20.

[20] Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models, *J. Econometrics*, **167**, 240–253.

[21] McAleer, M. and Medeiros, M.C. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries, *J. Econometrics*, **147**, 104–119.

[22] Medeiros, M.C. and Veiga, A. (2009). Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model, *Economet. Theor.*, **25**, 117–161.

[23] Meitz, M. and Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, *Economet. Theor.*, **24**, 1291–1320.

[24] Meitz, M. and Saikkonen, P. (2011). Parameter estimation in nonlinear AR–GARCH models, *Economet. Theor.*, **27**, 1236–1278.

[25] Neumann, M.H. (2011). Absolute regularity and ergodicity of Poisson count processes, *Bernoulli*, **17**, 1268–1284.

[26] Phillips, P.C.B. (1989). Partially identified econometric models, *Economet. Theor.*, **5**, 181–240.

[27] Redner, R.A. and Walker, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm, *SIAM Rev.*, **26**, 195–239.

[28] Rothenberg, T.J. (1971). Identification in parametric models, *Econometrica*, **39**, 577–591.

[29] Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach, *Ann. Statist.*, **34**, 2449–2495.

[30] Sussmann, H. (1992). Uniqueness of the weights for minimal feedforward nets with a given input-output map, *Neural Networks*, **5**, 589–593.

[31] Teicher, H. (1963). Identifiability of finite mixtures, *Ann. Math. Statist.*, **34**, 1265–1269.

[32] Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *J. Amer. Statist. Assoc.*, **89**, 208–218.

[33] Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation, *Ann. Statist.*, **9**, 501–513.

[34] Yakowitz, S.J. and Spragins, J.D. (1968). On the identifiability of finite mixtures, *Ann. Math. Statist.*, **39**, 209–214.

REVSTAT – Statistical Journal
Volume 14, Number 4, October 2016, 415–431

ESTIMATING THE SHAPE PARAMETER OF TOPP–LEONE DISTRIBUTION BASED ON PROGRESSIVE TYPE II CENSORED SAMPLES

Author: Husam Awni Bayoud
– College of Sciences and Humanities, Fahad Bin Sultan University, Tabuk-Saudi Arabia
hbayoud@fbsu.edu.sa

Received: March 2014 Revised: November 2014 Accepted: March 2015

Abstract:

- In this paper, classical and Bayesian point estimations of the Topp–Leone distribution shape parameter are studied when the sample is progressive Type II censored. The maximum likelihood estimator (MLE) of the unknown parameter is proposed. Since the MLE does not have an explicit form, an approximate MLE has been derived. The Bayes estimate and the associated credible interval are also studied. Lindley's method is proposed to approximate the Bayes estimate. The importance sampling technique is also proposed to approximate the Bayes estimate and to construct the associated credible interval. Monte Carlo simulations are performed to compare the performances of the proposed methods, and two data sets have been analyzed for illustrative purposes.

Key-Words:

- *Bayes estimate; credible interval; Lindley's approximation; maximum likelihood estimate; Markov chain Monte Carlo.*

AMS Subject Classification:

- 62N02, 62F15.

1. INTRODUCTION

Topp and Leone [19] introduced a family of distributions with finite support whose cumulative distribution function (cdf) is given by

$$F(x|\theta,\beta) = \begin{cases} 0\,, & x<0 \\ \left(\frac{x}{\beta}\left(2-\frac{x}{\beta}\right)\right)^{\theta}\,, & 0\le x<\beta \\ 1\,, & x\ge\beta \end{cases}\,, \qquad \theta>0\,, \tag{1.1}$$

and the probability density function (pdf) is given by

$$f(x|\theta,\beta) = \frac{2\theta}{\beta}\left(1-\frac{x}{\beta}\right)\left(\frac{x}{\beta}\left(2-\frac{x}{\beta}\right)\right)^{\theta-1}, \qquad 0<x<\beta\,,\ \ \theta>0\,. \tag{1.2}$$

For simplicity, we denote this distribution by $TL\,(\theta,\beta)$. Topp–Leone (T-L) distribution is a continuous unimodal distribution with bounded support; this makes it appropriate for modeling lifetime of distributions with finite support. Topp and Leone [19] did not provide any motivation for this family of distributions except to saying that it could be used to model failure data. Nadarajah and Kotz [15] showed that this distribution exhibit bathtub failure rate functions with widespread applications in reliability. Moreover, Ghitany *et al.* [10] showed that T-L distribution possesses some attractive reliability properties such as the bathtub-shape hazard rate, decreasing reversed hazard rate, upside-down mean residual life, and increasing expected inactivity time. Moments for T-L distribution were derived by Nadarajah and Kotz [15]. Zghoul [21] provided expressions for moments of ordered statistics from T-L distribution. Recently, Bayoud [6] derived admissible minimax estimates for the shape parameter of the T-L distribution under squared and linear-exponential loss functions. A reflected version of the Generalized T-L distribution was used by Van Drop and Kotz [20] to fit the U.S. income data for the year 2001 for Caucasian, Hispanic and Afro American populations.

Classical and Bayesian inferences of the parameters of T-L distribution have not yet been studied in the presence of censored samples. In this paper, we study classical and Bayesian estimations for the shape parameter of the T-L distribution when the sample is progressive Type II censored. A Type II progressive censoring scheme can be expressed as follows: suppose that n units are placed on a life test at time zero and the experimenter decides beforehand the quantity m, the number of failures to be observed. When the first failure time $X_{1:m:n}$ is observed, R_1 of the remaining $n-1$ surviving units are randomly selected and removed. At the second observed failure time $X_{2:m:n}$, R_2 of the remaining $n-R_1-2$ surviving units are randomly selected and removed. This experiment terminates at the time $X_{m:m:n}$ when the m^{th} failure

is observed, and the remaining $R_m = n - R_1 - R_2 - \cdots - R_{m-1} - m$ surviving units are all removed. The sample $\{X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}\}$ is called progressively Type II censored sample of size m from a sample of size n with censoring scheme $\{R_1, R_2, ..., R_m\}$. The values $\{m; R_1, R_2, ..., R_m\}$ are determined prior to the study. Note that, if $R_1 = R_2 = \cdots = R_m = 0$, so that $n = m$, then the progressively Type II censoring scheme reduces to the case of complete sample. Also note that if $R_1 = R_2 = \cdots = R_{m-1} = 0$, so that $R_m = n - m$, then the censoring scheme reduces to a conventional Type II censoring scheme. Readers may refer to [2] for more details about the progressive censoring.

The rest of this paper is organized as follows. In Section 2, we provide the model assumptions based on the progressive Type II censoring. The MLE is studied in Section 3. We propose an approximate MLE (AMLE) in Section 4. The Bayes estimate and the construction of the credible interval are discussed in Section 5. In Section 6, data analysis and some simulation studies are carried out to investigate the performance of the proposed estimation methods. Finally, we conclude the paper in Section 7.

2. MODEL ASSUMPTIONS

Let $X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}$ be a progressively Type II censored sample from T-L lifetime distribution (1.2), with $\{m; R_1, R_2, ..., R_m\}$ being the progressive censoring scheme. The likelihood function based on the observed progressive Type II censored sample $D = \{x_{1:m:n}, x_{2:m:n}, ..., x_{m:m:n}\}$ is given by:

$$L(D|\theta,\beta) = c\left(\frac{2\theta}{\beta}\right)^m \prod_{i=1}^{m}\left(1-\frac{x_{i:m:n}}{\beta}\right) u^{(\theta-1)}(x_{i:m:n})\left[1-u^{\theta}(x_{i:m:n})\right]^{R_i}, \tag{2.1}$$

where

$$c = n(n-1-R_1)(n-2-R_1-R_2)\cdots\left(n-\sum_{i=1}^{m-1}(R_i+1)\right), \qquad 0 < x_{i:m:n} < \beta,$$

and

$$u(x_{i:m:n}) = \frac{x_{i:m:n}}{\beta}\left(2-\frac{x_{i:m:n}}{\beta}\right) \in (0,1) \qquad \forall i = 1, 2, ..., m.$$

The log-likelihood function, $l(D|\theta,\beta) = \ln L(D|\theta,\beta)$, may be written from (2.1) as:

$$l(D|\theta,\beta) \propto m\ln(\theta) + \sum_{i=1}^{m}\theta\ln u(x_{i:m:n}) + \sum_{i=1}^{m} R_i \ln\left[1-u^{\theta}(x_{i:m:n})\right]. \tag{2.2}$$

3. MAXIMUM LIKELIHOOD ESTIMATE

Equating the partial derivative of the log-likelihood function $l(D|\theta,\beta)$ in (2.2) to zero, we have that:

$$\frac{\partial l(D|\theta,\beta)}{\partial\theta} = \frac{m}{\theta} + \sum_{i=1}^{m} \ln u\left(x_{i:m:n}\right) - \sum_{i=1}^{m} \frac{u^{\theta}\left(x_{i:m:n}\right)}{1-u^{\theta}\left(x_{i:m:n}\right)} \ln u\left(x_{i:m:n}\right) R_i = 0\,. \tag{3.1}$$

The MLE of θ is the solution of the likelihood equation (3.1). Since (3.1) is a non-linear equation, a numerical technique is needed. Newton-Raphson method is proposed to obtain the MLE iteratively. A suitable initial guess for the iterative method will be proposed in the next section. However, numerical results, presented in Section 6, show that the numerical MLE converges to the true parameter quite accurately without showing any problem with convergence.

4. APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATE

The likelihood equation (3.1), as mentioned in the previous section, does not admit explicit solution for the shape parameter. Therefore, we expand the function $g_i(\theta) = \frac{u^{\theta}(x_{i:m:n})}{1-u^{\theta}(x_{i:m:n})}$ in a first-order Taylor series around $v_i = \frac{\ln p_i}{\ln u(x_{i:m:n})}$, where $p_i = 1 - \prod_{j=m-i+1}^{m} \frac{j+\sum_{i=m-j+1}^{m} R_i}{1+j+\sum_{i=m-j+1}^{m} R_i}$ for $i = 1, 2, ..., m$. We may then consider the following approximation:

$$g_i(\theta) \approx \frac{u^{v_i}\left(x_{i:m:n}\right)}{1-u^{v_i}\left(x_{i:m:n}\right)} + (\theta - v_i)\frac{u^{v_i}\left(x_{i:m:n}\right)}{\left[1-u^{v_i}\left(x_{i:m:n}\right)\right]^2} \ln u\left(x_{i:m:n}\right). \tag{4.1}$$

Using the approximation in (4.1), (3.1) is roughly:

$$\begin{aligned} &\frac{m}{\theta} + \sum_{i=1}^{m} \ln u\left(x_{i:m:n}\right) - \\ &- \sum_{i=1}^{m} \frac{u^{v_i}\left(x_{i:m:n}\right)}{1-u^{v_i}\left(x_{i:m:n}\right)} \left[1 + \frac{(\theta - v_i)}{1-u^{v_i}\left(x_{i:m:n}\right)} \ln u\left(x_{i:m:n}\right)\right] \ln u\left(x_{i:m:n}\right) R_i = 0\,. \end{aligned} \tag{4.2}$$

From (4.2), we obtain the AMLE of θ as a solution of the quadratic equation:

$$A\theta^2 + B\theta + m = 0\,,$$

where

$$A = -\sum_{i=1}^{m} \frac{u^{v_i}\left(x_{i:m:n}\right)}{\left[1-u^{v_i}\left(x_{i:m:n}\right)\right]^2} \left[\ln u\left(x_{i:m:n}\right)\right]^2 R_i$$

and

$$B \;=\; \sum_{i=1}^{m} \ln u\left(x_{i:m:n}\right) \left[1 - R_i \frac{u^{v_i}\left(x_{i:m:n}\right)}{1-u^{v_i}\left(x_{i:m:n}\right)} \left[1 - v_i \frac{\ln u\left(x_{i:m:n}\right)}{1-u^{v_i}\left(x_{i:m:n}\right)}\right]\right].$$

Therefore, the AMLE, say $\hat{\theta}_{AMLE}$, is obtained as

$$\hat{\theta}_{AMLE} \;=\; \frac{-B - \sqrt{B^2 - 4Am}}{2A}\,, \tag{4.3}$$

which is the only positive root. This procedure has been used, for example, by Balakrishnan and Aggarwala [2], Balakrishnan and Varadan [3], Balosooriya and Balakrishnan [4] and Kim and Han [12].

It is worth mentioning that the proposed AMLE (4.3) may provide a convenient starting value for the iterative solution for the MLE in (3.1).

5. BAYESIAN INFERENCE

In this section, we discuss the Bayes estimate and the associated credible interval for the shape parameter. The squared error loss function (SELF) is considered, which is defined as

$$L(\hat{\theta}) \;=\; \left(\theta - \hat{\theta}\right)^2,$$

where $\hat{\theta}$ is the estimator of θ.

5.1. Prior and posterior analysis

The shape parameter θ is positive. So, it is assumed that θ has an Exponential prior with pdf:

$$g\left(\theta\right) = ae^{-a\theta}\,, \qquad \theta > 0 \quad \text{and} \quad a > 0\,.$$

This prior is conjugate when the complete sample is considered; see [6].

It follows, from (2.1) and the prior pdf, that the posterior density function of θ can be written as:

$$\begin{aligned} \pi\left(\theta|D,\beta\right) \;&=\; \frac{L(D|\theta,\beta)g\left(\theta\right)}{\int_0^\infty L(D|\theta,\beta)g\left(\theta\right)d\theta} \\ &=\; \frac{\theta^m e^{-a\theta}\prod_{i=1}^{m} u^{\theta}\left(x_{i:m:n}\right)\left[1-u^{\theta}\left(x_{i:m:n}\right)\right]^{R_i}}{K}\,, \end{aligned} \tag{5.1}$$

where $K = \int_0^\infty \theta^m e^{-a\theta} \prod_{i=1}^m u^\theta (x_{i:m:n}) \left[1 - u^\theta (x_{i:m:n})\right]^{R_i} d\theta$, the normalizing constant.

Under the SELF, the Bayes estimate of θ, say $\hat{\theta}_B(a)$, is the posterior mean, which is given by:

$$\hat{\theta}_B(a) = E_\pi (\theta|D, \beta) = \frac{1}{K} \int_0^\infty \theta^{m+1} e^{-a\theta} \prod_{i=1}^m u^\theta (x_{i:m:n}) \left[1 - u^\theta (x_{i:m:n})\right]^{R_i}. \tag{5.2}$$

It is obvious that (5.2) cannot be evaluated explicitly. Therefore, we propose two approaches to approximate (5.2): Lindley's procedure and the MCMC using the importance sampling technique.

5.2. Lindley's approximation

Lindley [14] proposed an approximation procedure to evaluate the ratio of two integrals, such that the Bayes estimate in (5.2) takes a form containing no integrals. This procedure has been used by several authors in the literature to obtain the Bayes estimates for various distributions; see, for instance, Press [18]. Consider:

$$I(D, a) = \frac{\int y(\theta) e^{l(\theta)+\tau(\theta)} d\theta}{\int e^{l(\theta)+\tau(\theta)} d\theta} \tag{5.3}$$

where l is the log-likelihood function of the observed sample, $y(\theta)$ is a continuous function in θ, and $\tau(\theta) = \ln g(\theta)$ where $g(\theta)$ is the prior pdf of θ.

Based on Lindley's procedure, the ratio (5.3) is approximated by:

$$I(D, a) \approx y\left(\hat{\theta}\right) + \frac{1}{2} \left(\hat{y}_{\theta\theta} + 2\hat{y}_\theta \hat{\tau}_\theta\right) \hat{\sigma}_{\theta\theta} + \frac{1}{2} \left(\hat{y}_\theta \hat{\sigma}_{\theta\theta}^2 \hat{l}_{\theta\theta\theta}\right) \tag{5.4}$$

where $y_{\theta\theta}$ denotes the second derivative of the function $y(\theta)$ with respect to θ, $\hat{y}_{\theta\theta}$ represents the same expression evaluated at $\theta = \hat{\theta}_{MLE}$, $\hat{\tau}_\theta = \frac{\partial}{\partial\theta}\tau(\theta)|_{\theta=\hat{\theta}_{MLE}}$, $\hat{l}_{\theta\theta} = \frac{\partial^2 l}{\partial\theta^2}|_{\theta=\hat{\theta}_{MLE}}$, $\hat{l}_{\theta\theta\theta} = \frac{\partial^3 l}{\partial\theta^3}|_{\theta=\hat{\theta}_{MLE}}$ and $\hat{\sigma}_{\theta\theta} = -\frac{1}{\hat{l}_{\theta\theta}}$.

Hence, the approximate Bayes estimate can be obtained using Lindley's procedure, by substituting $y(\theta) = \theta$, l=log-likelihood function (2.2) and $g(\theta) = ae^{-a\theta}$ in Lindley's approximation (5.4), as:

$$\hat{\theta}_{B,L}(a) \approx \hat{\theta}_{MLE} + \frac{a}{\hat{l}_{\theta\theta}} + \frac{1}{2} \frac{\hat{l}_{\theta\theta\theta}}{\hat{l}_{\theta\theta}^2} \tag{5.5}$$

where $\hat{\theta}_{MLE}$ is the MLE of θ,

$$\hat{l}_{\theta\theta} = \frac{\partial^2 l}{\partial\theta^2}|_{\theta=\hat{\theta}_{MLE}} = -\frac{m}{\hat{\theta}^2_{MLE}} - \sum_{i=1}^{m} \left[\ln u\left(x_{i:m:n}\right)\right]^2 \frac{u^{\hat{\theta}_{MLE}}\left(x_{i:m:n}\right)}{\left[1-u^{\hat{\theta}_{MLE}}\left(x_{i:m:n}\right)\right]^2} R_i \,,$$

and

$$\begin{aligned}\hat{l}_{\theta\theta\theta} &= \frac{\partial^3 l}{\partial\theta^3}|_{\theta=\hat{\theta}_{MLE}} \\ &= \frac{2m}{\hat{\theta}^3_{MLE}} - \sum_{i=1}^{m} \left[\ln u\left(x_{i:m:n}\right)\right]^3 \frac{u^{\hat{\theta}_{MLE}}\left(x_{i:m:n}\right)\left[1+u^{\hat{\theta}_{MLE}}\left(x_{i:m:n}\right)\right]}{\left[1-u^{\hat{\theta}_{MLE}}\left(x_{i:m:n}\right)\right]^3} R_i \,.\end{aligned}$$

5.3. MCMC method

Unfortunately, Lindley's procedure fails to construct credible intervals for the unknown parameter. Hence, we propose to use the importance sampling technique to approximate the Bayes estimate and to construct the associated credible interval. Similar procedure was used, for example, by Chen *et al.* [7], Kundu and Pradhan [13], Pradhan and Kundu [16] and [17]. To implement the importance sampling technique, we rewrite the posterior pdf (5.1) as follows:

$$\pi(\theta|D,\beta) \propto f_1(\theta|D)\, f_2(\theta)$$

where

$$f_1(\theta|D) = \frac{\left[a-\sum_{i=1}^{m}\ln u(x_{i:m:n})\right]^m}{\Gamma(m+1)}\ \theta^m e^{-\theta\left[a-\sum_{i=1}^{m}\ln u(x_{i:m:n})\right]}$$

which is clearly a gamma density function with the shape parameter $(m+1)$ and scale parameter $[a-\sum_{i=1}^{m}\ln u(x_{i:m:n})]^{-1}$; and

$$f_2(\theta) = \prod_{i=1}^{m}\left[1-u^{\theta}(x_{i:m:n})\right]^{R_i}.$$

Therefore, (5.2) can be written as:

$$\hat{\theta}_B\left(a\right) = \frac{\int_0^\infty \theta f_1(\theta|D) f_2(\theta) d\theta}{\int_0^\infty f_1(\theta|D) f_2(\theta) d\theta}\,. \tag{5.6}$$

Now, we propose the following algorithms, along the line of Kundu and Pradhan [13], to compute the approximate Bayes estimate and to construct the associated credible interval for the parameter θ.

5.3.1. Algorithm 1 (BE)

Step 1) Generate a random sample of size M from $f_1(\theta|D)$, gamma density function with the shape parameter $(m+1)$ and scale parameter $[a - \sum_{i=1}^{m} \ln u(x_{i:m:n})]^{-1}$, say $\theta_1, \theta_2, ..., \theta_M$;

Step 2) Compute $f_2(\theta_j) = \prod_{i=1}^{m} [1 - u^{\theta_j}(x_{i:m:n})]^{R_i}$, for $j = 1, 2, ..., M$;

Step 3) Under the assumption of SELF, a simulation consistent estimate of θ can be obtained using the importance sampling technique as:

$$\hat{\theta}_{B,IS}(a) = \frac{\sum_{j=1}^{M} \theta_j f_2(\theta_j)}{\sum_{j=1}^{M} f_2(\theta_j)} .$$

Using this algorithm, it is possible to construct the Bayes estimate of any function of θ, say $H(\theta)$ as:

$$\hat{H}(\theta) = \frac{\sum_{j=1}^{M} H(\theta_j) f_2(\theta_j)}{\sum_{j=1}^{M} f_2(\theta_j)}, \quad \text{provided that } \hat{H}(\theta) \text{ is defined at all } j = 1, 2, ..., m .$$

Now, to compute the credible interval of θ. Let, for $0 < p < 1$, θ_p be such that $P(\theta \leq \theta_p|D, \beta) = \int_0^{\theta_p} \pi(\theta|D, \beta) d\theta = p$, where $\pi(\theta|D, \beta)$ is the posterior pdf defined in (5.1).

5.3.2. Algorithm 2 (credible interval)

Here, we use the sample $\theta_1, \theta_2, ..., \theta_M$ that is obtained from Algorithm 1.

Step 1) Compute $w_j = \frac{f_2(\theta_j)}{\sum_{j=1}^{M} f_2(\theta_j)}$ for $j = 1, 2, ..., M$;

Step 2) Arrange the set $\{(\theta_1, w_1), (\theta_2, w_2), ..., (\theta_M, w_M)\}$ as

$$\left\{(\theta_{(1)}, w_{[1]}), (\theta_{(2)}, w_{[2]}), ..., (\theta_{(M)}, w_{[M]})\right\} ,$$

where $\theta_{(1)} \leq \theta_{(2)}, ..., \leq \theta_{(M)}$;

Step 3) The $100(1-\alpha)\%$ credible interval for θ is given by:

$$\left(\hat{\theta}_{\frac{\alpha}{2}}, \hat{\theta}_{1-\frac{\alpha}{2}}\right)$$

where $\hat{\theta}_p$ is a simulation consistent Bayes estimate for θ_p, which is given by $\theta_{(M_p)}$ such that M_p is the integer satisfying:

$$\sum_{j=1}^{M_p} w_{[j]} \leq p < \sum_{j=1}^{M_p+1} w_{[j]} .$$

Proposition 5.1. *The posterior pdf $\pi(\theta|D,\beta)$ in (5.1) is log-concave.*

Proof: Since $u(x_i) = \frac{x_i}{\beta}(2 - \frac{x_i}{\beta}) > 0$, then it is easy to see that:

$$\frac{\partial^2 \ln \pi(\theta|D,\beta)}{\partial \theta^2} = -\left[\frac{m}{\theta^2} + \sum_{i=1}^{m} [\ln u(x_{i:m:n})]^2 \frac{u^\theta(x_{i:m:n})}{[1-u^\theta(x_{i:m:n})]^2} R_i\right] < 0$$

for any θ, this proves the result. □

Since the posterior distribution (5.1) is log-concave, then one can apply Devroye's algorithm introduced in Devroye [8] to generate a sample from the posterior distribution, say $\theta_1, \theta_2, ..., \theta_M$. Based on this sample and under the SELF, the approximate Bayes estimate of θ is given by:

$$\hat{\theta}_{MCMC} = \hat{E}(\theta|D) = \frac{1}{M}\sum_{j=1}^{M} \theta_j \, .$$

The $100(1-\alpha)\%$ credible interval of θ can be computed by ordering $\theta_1, \theta_2, ..., \theta_M$ as $\theta_{(1)} \leq \theta_{(2)} \leq ... \leq (\theta_M)$ and taking the interval as:

$$\left(\theta_{(M(\frac{\alpha}{2}))}, \theta_{(M(1-\frac{\alpha}{2}))}\right).$$

6. SIMULATION STUDY AND DATA ANALYSIS

6.1. Simulations

In this section, we present some simulation studies to observe the behavior of the proposed estimation methods for different sample sizes, different priors and for different censoring schemes. We have considered three sample sizes, $n = 15$, 25 and 50; and three progressive Type II censoring schemes with $m = 5$, namely, $(n-m, 0, 0, 0, 0)$, $(0, 0, 0, 0, n-m)$ and $(R_1, R_2, R_3, R_4, R_5)$ where $R_i = \frac{n-m}{m}$ for $i = 1, 2, ..., 5$.

In all cases, the parameter β is assumed without loss of generality to equal 1. Simulations are performed for three values of the shape parameter, namely, $\theta = 0.5$, $\theta = 1$ and $\theta = 10$. For a given n, m and $(R_1, R_2, ...R_m)$, we have generated a sample for the given censoring scheme. The AMLE is computed for the shape parameter based on the method proposed in Section 3. We use this AMLE as a starting value to obtain the MLE iteratively by using Newton-Raphson method as discussed in Section 2. The approximate Bayes estimate is computed for

the shape parameter using Lindley's procedure and the importance sampling technique based on 1000 importance sampling. For Bayesian estimation, the following priors are considered: Prior 0: assuming $a = 0.0001$, a very small value, and Prior 1: informative prior with $a \approx 1/\theta$ and $a \approx 2/\theta$, separately, since $E(\theta) = 1/a$. The expected value and the corresponding mean squared error (MSE) of the proposed estimates are computed over 1000 replications. The results are reported in Tables 1, 2 and 3 when $\theta = 0.5, \theta = 1$ and $\theta = 10$, respectively.

From Tables 1, 2 and 3, it is clear that as the sample size increases, the MSE decreases for all estimation methods. This verifies the consistency of the proposed methods. It is also obvious that the AMLE and the approximate Bayes estimates under Prior 1 perform, in terms of MSE, better than the iterative MLE and the approximate Bayes estimates under Prior 0. For fixed sample size n, fixed θ and for any censoring scheme, the approximate Bayes estimates under Prior 1 with $a \approx 2/\theta$ outperform the other estimates in terms of the MSE. It is noticeable that the AMLE performs better than the MLE in all cases. The approximate Bayes estimates under Prior 0 do not perform as efficiently as the other estimates.

Table 1: Expected value of the proposed estimators and the corresponding MSE when $\theta = 0.5$.

n	**Scheme**	$\hat{\theta}_{MLE}$	$\hat{\theta}_{AMLE}$	$\hat{\theta}_{B,L}(a)$			$\hat{\theta}_{B,IS}(a)$		
				$a=2$	$a=4$	$a=10^{-4}$	$a=2$	$a=4$	$a=10^{-4}$
15	(0,0,0,0,10)	0.532 0.0204	0.521 0.0193	0.530 0.0170	0.483 0.0112	0.587 0.0353	0.528 0.0175	0.488 0.0132	0.586 0.0356
	(2,2,2,2,2)	0.534 0.0235	0.477 0.0167	0.534 0.0192	0.481 0.0120	0.593 0.0394	0.533 0.0200	0.488 0.0141	0.593 0.0392
	(10,0,0,0,0)	0.533 0.0308	0.517 0.0286	0.543 0.0252	0.472 0.0125	0.623 0.0604	0.540 0.0254	0.483 0.0166	0.624 0.0607
25	(0,0,0,0,20)	0.522 0.0145	0.516 0.0141	0.524 0.0131	0.497 0.0105	0.555 0.0190	0.515 0.0137	0.490 0.0123	0.548 0.0188
	(4,4,4,4,4)	0.523 0.0167	0.474 0.0128	0.527 0.0150	0.493 0.0111	0.562 0.0226	0.524 0.0155	0.493 0.0122	0.558 0.0225
	(20,0,0,0,0)	0.524 0.0252	0.509 0.0232	0.541 0.0230	0.494 0.0131	0.587 0.0360	0.537 0.0225	0.497 0.0153	0.588 0.0364
50	(0,0,0,0,45)	0.509 0.0078	0.507 0.0077	0.514 0.0075	0.500 0.0063	0.532 0.0102	0.476 0.0093	0.452 0.0097	0.505 0.0117
	(9,9,9,9,9)	0.508 0.0087	0.470 0.0076	0.515 0.0084	0.498 0.0075	0.534 0.0111	0.494 0.0099	0.474 0.0094	0.520 0.0114
	(45,0,0,0,0)	0.503 0.0135	0.491 0.0133	0.527 0.0138	0.497 0.0114	0.561 0.0215	0.522 0.0133	0.495 0.0120	0.563 0.0221

Table 2: Expected value of the proposed estimators and the corresponding MSE when $\theta = 1$.

n	**Scheme**	$\hat{\theta}_{MLE}$	$\hat{\theta}_{AMLE}$	$\hat{\theta}_{B,L}(a)$			$\hat{\theta}_{B,IS}(a)$		
				$a=1$	$a=2$	$a=10^{-4}$	$a=1$	$a=2$	$a=10^{-4}$
15	(0,0,0,0,10)	1.08 0.1010	1.06 0.0946	1.08 0.0816	0.978 0.0524	1.13 0.1231	1.07 0.0842	0.989 0.0632	1.13 0.1228
	(2,2,2,2,2)	1.08 0.1060	0.969 0.0760	1.08 0.0853	0.974 0.0513	1.15 0.1427	1.08 0.0884	0.990 0.0628	1.15 0.1434
	(10,0,0,0,0)	1.09 0.1500	1.06 0.1370	1.11 0.1160	0.956 0.0509	1.19 0.2014	1.10 0.1190	0.981 0.0684	1.20 0.2032
25	(0,0,0,0,20)	1.03 0.0611	1.02 0.0594	1.04 0.0549	0.994 0.0443	1.09 0.0709	1.02 0.0559	0.973 0.0504	1.08 0.0751
	(4,4,4,4,4)	1.03 0.0694	0.937 0.0548	1.04 0.0619	0.985 0.0465	1.10 0.0792	1.04 0.0648	0.981 0.0507	1.10 0.0784
	(20,0,0,0,0)	1.04 0.1020	1.01 0.0960	1.08 0.0925	0.986 0.0527	1.16 0.1368	1.07 0.0917	0.993 0.0612	1.16 0.1388
50	(0,0,0,0,45)	1.02 0.0352	1.05 0.0347	1.03 0.0340	0.996 0.0245	1.06 0.0356	0.953 0.0384	0.903 0.0372	0.990 0.0390
	(9,9,9,9,9)	1.02 0.0387	0.947 0.0326	1.04 0.0374	0.991 0.0260	1.07 0.0466	1.00 0.0431	0.947 0.0352	1.04 0.0470
	(45,0,0,0,0)	1.01 0.0628	0.992 0.0600	1.06 0.0639	0.983 0.0408	1.12 0.0920	1.05 0.0618	0.979 0.0427	1.13 0.0930

Table 3: Expected value of the proposed estimators and the corresponding MSE when $\theta = 10$.

n	**Scheme**	$\hat{\theta}_{MLE}$	$\hat{\theta}_{AMLE}$	$\hat{\theta}_{B,L}(a)$			$\hat{\theta}_{B,IS}(a)$		
				$a=0.1$	$a=0.2$	$a=10^{-4}$	$a=0.1$	$a=0.2$	$a=10^{-4}$
15	(0,0,0,0,10)	10.7 11.1	10.5 10.6	10.5 8.9	9.6 4.9	11.5 13.0	10.6 9.4	9.8 5.9	11.4 12.8
	(2,2,2,2,2)	10.8 12.2	9.6 8.9	10.7 9.7	9.6 5.1	11.7 14.7	10.7 9.9	9.7 6.1	11.7 14.7
	(10,0,0,0,0)	10.8 15.0	10.5 13.4	11.0 11.5	9.4 5.2	12.2 21.3	10.9 11.9	9.6 6.9	12.3 21.5
25	(0,0,0,0,20)	10.5 5.8	10.3 5.6	10.5 5.2	10.0 4.1	10.9 7.4	10.3 5.5	9.8 4.8	10.8 7.6
	(4,4,4,4,4)	10.5 6.5	9.5 4.9	10.6 5.9	10.0 4.5	11.1 8.6	10.5 6.1	10.0 5.0	11.0 8.7
	(20,0,0,0,0)	10.4 9.4	10.1 8.9	10.7 8.6	9.8 5.3	11.8 14.9	10.7 8.4	9.9 6.1	11.8 15.1
50	(0,0,0,0,45)	10.1 3.1	10.1 3.1	10.2 3.0	10.1 2.7	10.6 3.9	9.5 3.8	9.2 3.9	9.9 4.2
	(9,9,9,9,9)	10.2 3.8	9.4 3.2	10.3 3.7	10.1 3.0	10.6 4.5	9.9 4.3	9.6 3.8	10.3 4.8
	(45,0,0,0,0)	10.1 6.3	9.9 6.0	10.6 6.5	10.0 4.4	11.2 10.2	10.5 6.2	10.0 4.7	11.3 10.4

6.2. Data analysis

In this section, we analyze real and simulated data sets using the proposed estimation methods for illustrative purposes.

6.2.1. Real data

We analyze the failure time (in mileage) of eighteen military carriers presented by Grubbs [11] as follows:

$$162,\ 200,\ 271,\ 302,\ 393,\ 508,\ 539,\ 629,\ 706,\ 777,$$
$$884,\ 1101,\ 1182,\ 1463,\ 1603,\ 1984,\ 2355,\ 2880\,.$$

First, it was checked whether the T-L distribution can be used or cannot to analyze this data set. The MLE of β is 2880, the maximum order statistic, and the MLE of θ is 1.133 . The Bayes estimate of θ, under the SELF, is 1.125 when $a = 1$, see [6]. It is obvious that the MLE and the Bayes estimate are almost the same. The Kolmogorov-Smirnov (KS) distance between the empirical distribution function and the fitted distribution function, using the MLEs, has been used to check the goodness of fit. The KS statistic value is 0.135, and the KS critical value is 0.309 at at $n = 18$ and $\alpha = 0.05$. Accordingly, one cannot reject the hypothesis that the data are coming from T-L distribution. We consider the following censoring schemes, assuming $m = 6$:

Scheme 1) $(R_1 = R_2 = \cdots = R_5 = 0, R_6 = 12)$.

Scheme 2) $(R_1 = R_2 = \cdots = R_5 = R_6 = 2)$.

Scheme 3) $(R_1 = 12, R_2 = \cdots = R_5 = R_6 = 0)$.

Based on Schemes 1, 2 and 3, we have generated the following progressive Type II censored samples:

$$D = (162, 200, 271, 302, 393, 508)\ ,$$
$$D = (162, 271, 393, 508, 539, 884)$$

and

$$D = (162, 302, 508, 777, 884, 1463)\ ,$$

respectively. The proposed estimates and the credible interval for the shape parameter are computed and reported in Table 4. It is observed from Table 4 that all estimates are in quite similar agreement and close to the estimates obtained using the complete sample. The approximate Bayes estimates dominate the other estimates when the hyper-parameter a is assumed to equal 1. The associated credible intervals for the shape parameter are satisfactory in all the cases.

Table 4: Real Data Analysis.

Estimate		**Censoring Scheme**		
		Scheme 1	**Scheme 2**	**Scheme 3**
$\hat{\theta}_{MLE}$		1.169	1.303	1.241
$\hat{\theta}_{AMLE}$		1.153	1.289	1.236
$\hat{\theta}_{B,L}(a)$	$a = 0.5$	1.205	1.346	1.307
	$a = 1$	1.163	1.290	1.246
	$a = 3.5$	0.996	1.066	1.002
$\hat{\theta}_{B,IS}(a)$	$a = 0.5$	1.151	1.307	1.285
	$a = 1$	1.128	1.287	1.246
	$a = 3.5$	1.085	1.141	1.053
90% Credible Interval	$a = 0.5$	(0.76, 1.51)	(0.69, 1.49)	(0.79, 1.85)
	$a = 1$	(0.74, 1.47)	(0.66, 1.47)	(0.77, 1.84)
	$a = 3.5$	(0.66, 1.38)	(0.58, 1.27)	(0.65, 1.48)

6.2.2. Simulated data

We analyze the following simulated data set presented by Genc [9] assuming $\theta = 0.3$ and $\beta = 1$:

$$0.1425, 0.2707, 0.2783, 0.0718, 0.4537, 0.0615, 0.0047, 0.3454, 0.4428, 0.1909,$$
$$0.1028, 0.0013, 0.0592, 0.5413, 0.2442, 0.0001, 0.0002, 0.0178, 0.0114, 0.5388$$

We consider the following censoring schemes, assuming $m = 4$:

Scheme 1) $(R_1 = R_2 = R_3 = 0, R_4 = 16)$.

Scheme 2) $(R_1 = R_2 = R_3 = R_4 = 4)$.

Scheme 3) $(R_1 = 16, R_2 = R_3 = R_4 = 0)$.

Based on Schemes 1, 2 and 3, we have generated the following progressive Type II censored samples:

$$D = (0.0001, 0.0002, 0.0013, 0.0047),$$
$$D = (0.0001, 0.0047, 0.01114, 0.0178)$$

and

$$D = (0.0001, 0.0013, 0.0718, 0.2707)\ , \qquad \text{respectively.}$$

The proposed estimates and the credible interval are computed and reported in Table 5. It is clear from Table 5 that all estimates are quite similar, and the approximate Bayes estimates dominate the other when the hyper-parameter $a = 5$. It is also observed that the credible intervals are satisfactory under all the cases.

Table 5: Simulated Data Analysis.

Estimate		**Censoring Scheme**		
		Scheme 1	**Scheme 2**	**Scheme 3**
$\hat{\theta}_{MLE}$		0.3699	0.4381	0.3664
$\hat{\theta}_{AMLE}$		0.3694	0.4266	0.3662
$\hat{\theta}_{B,L}(a)$	$a = 0.75$	0.3901	0.4651	0.4041
	$a = 3.5$	0.3674	0.4294	0.3720
	$a = 5$	0.3550	0.4100	0.3546
$\hat{\theta}_{B,IS}(a)$	$a = 0.75$	0.3541	0.4731	0.3972
	$a = 3.5$	0.3792	0.4120	0.3675
	$a = 5$	0.3530	0.4054	0.3555
90% Credible Interval	$a = 0.75$	$(0.25, 0.46)$	$(0.23, 0.46)$	$(0.24, 0.58)$
	$a = 3.5$	$(0.24, 0.47)$	$(0.21, 0.44)$	$(0.22, 0.56)$
	$a = 5$	$(0.23, 0.45)$	$(0.21, 0.42)$	$(0.22, 0.51)$

7. CONCLUSIONS

In this article, classical and Bayesian point estimations were proposed for the shape parameter of the Topp–Leone distribution when the sample is progressive Type II censored. It was observed that the MLE cannot be derived in explicit form. Hence, an approximate MLE was proposed. Bayes estimate of the shape parameter cannot be obtained in explicit form. Lindley's procedure and the importance sampling technique were proposed to obtain the approximate Bayes estimate and to construct the credible interval for the shape parameter. The performance of the different estimation methods was compared by Monte Carlo simulations. It was observed that the approximate Bayes estimates, based on the informative prior with $a \approx 2/\theta$, outperform the other estimates in terms of the MSE. It was also noticed that the AMLE performs well and dominates the MLE in terms of the MSE in all cases.

ACKNOWLEDGMENTS

The author would like to express his appreciation to the Editor-In-Chief Professor M. Ivette Gomes, anonymous Associate Editor and referee for their valuable comments and suggestions that have improved the article.

REFERENCES

[1] Balakrishnan, N. (2007). Progressive censoring methodology: an appraisal, *TEST*, **16**, 211–296.

[2] Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring Theory, Methods and Applications*, Birkhauser, Boston.

[3] Balakrishnan, N. and Varadan, J. (1991). Approximate MLEs for the location and scale parameters of the extreme value distribution with censoring, *IEEE Transaction on Reliability*, **40**, 146–151.

[4] Balasooriya, U. and Balakrishnan, N. (2000). Reliability sampling plans for log-normal distribution based on progressively-censored samples, *IEEE Transaction on Reliability*, **49**, 199–203.

[5] Banerjee, A. and Kundu, D. (2008). Inference based on type-II hybrid censored data from a Weibull distribution, *IEEE Transaction on Reliability*, **57**, 369–378.

[6] Bayoud, H. (2015). Admissible minimax estimators for the shape parameter of Topp–Leone distribution, *Communications in Statistics-Theory and Methods*, doi:10.1080/03610926.2013.818700.

[7] Chen, M.H.; Shao, Q.M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York.

[8] Devroye, L. (1984). A simple algorithm for generating random variables with log-concave density, *Computing*, **33**, 247–257.

[9] Genc, A.I. (2011). Estimation of $P(X > Y)$ with Topp–Leone distribution, *Journal of Statistical Computation And Simulation*, doi:10.1080/00949655.2011.607821.

[10] Ghitany, M.E.; Kotz, S. and Xie, M. (2005). On some reliability measures and their stochastic ordering for the Topp–Leone distribution, *Journal of Applied Statistics*, **32**, 715–722.

[11] Grubbs, F.E. (1971). Approximate fiducial bounds on reliability for the two parameter negative exponential distribution, *Technometrics*, **13**, 873–876.

[12] Kim, C. and Han, K. (2010). Estimation of the scale parameter of the half-logistic distribution under progressively type II censored sample, *Stat. Papers*, **51**, 375–387.

[13] Kundu, D. and Pradhan, B. (2009). Bayesian inference and life testing plans for generalized exponential distribution, *Science in China, Series A: Mathematics (Special volume dedicated to Professor Z.D. Bai)*, **52**, 1373–1388.

[14] Lindley, D.V. (1980). Approximate Bayesian methods, *Trabajos de Estadistica Y de Investigacion Operativa*, **31**(1), 223–245.

[15] Nadarajah, S. and Kotz, S. (2003). Moments of some J-shaped distributions, *Journal of Applied Statistics*, **30**, 311–317.

[16] Pradhan, S. and Kundu, S. (2009). On progressively censored generalized exponential distribution, *Test*, **18**, 497–515.

[17] Pradhan, S. and Kundu, S. (2011). Bayes estimation and prediction of the two-parameter gamma distribution, *Journal of Statistical Computation and Simulation*, **81**, 1187–1198.

[18] Press, S.J. (2001). *The Subjectivity of Scientists and the Bayesian Approach*, Wiley, New York.

[19] Topp, C.W. and Leone, F.C. (1955). A family of J-shaped frequency functions, *Journal of the American Statistical Association*, **50**, 209–219.

[20] Van Dorp, J.R. and Kotz, S. (2006). *Modeling Income Distributions Using Elevated Distributions. Distribution Models Theory*, World Scientific Press, Singapore 1–25.

[21] Zghoul, A. (2010). Order statistics from a family of J-Shaped distributions, *Metron*, **LXVIII**(2), 127–136.

REVSTAT – Statistical Journal
Volume 14, Number 4, October 2016, 433–454

OBJECTIVE BAYESIAN ESTIMATORS FOR THE RIGHT CENSORED RAYLEIGH DISTRIBUTION: EVALUATING THE AL-BAYYATI LOSS FUNCTION

Author: J.T. Ferreira
– Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa
johan.ferreira@up.ac.za

A. Bekker
– Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa

M. Arashi
– Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa
and
Department of Statistics, School of Mathematical Sciences, Shahrood Unviversity of Technology, Shahrood, Iran

Received: December 2013 Revised: March 2015 Accepted: March 2015

Abstract:

- The Rayleigh distribution, serving as a special case of the Weibull distribution, is known to have wide applications in survival analysis, reliability theory and communication engineering. In this paper, Bayesian estimators (including shrinkage estimators) of the unknown parameter of the censored Rayleigh distribution are derived using the Al-Bayyati loss function, whilst simultaneously considering different objective prior distributions. Comparisons are made between the proposed estimators by calculating the risk functions using simulation studies and an illustrative example.

Key-Words:

- *Al-Bayyati loss; Rayleigh distribution; risk efficiency; shrinkage estimation; squared error loss.*

AMS Subject Classification:

- 62501, 62N10.

1. INTRODUCTION

The Rayleigh distribution is a continuous probability distribution serving as a special case of the well-known Weibull distribution. This distribution has long been considered to have significant applications in fields such as survival analysis, reliability theory and especially communication engineering.

When considering the complete Rayleigh model, the probability density function is given by

$$f(x;\theta) = 2\theta x e^{-\theta x^2}\,, \qquad x, \theta > 0\,, \tag{1.1}$$

using the parametrization of the distribution as proposed by Bhattacharya and Tyagi (1990), and is denoted by $X \sim Rayleigh(\theta)$. The parameter θ is a scale parameter, and characterizes the lifetime of the object under consideration in application.

Mostert (1999) did extensive work concerning the censored model, and showed that the censored Rayleigh model is relatively easy to use compared to other more complex models (such as the Weibull- and compound Rayleigh models). In certain types of applications, it is not uncommon that some observations may cease to be observed due to machine failure, budgetary constraints, and the likes. To compensate for such events, right censored analyses utilizes information only obtained from the first d observations. Thus, the right censored sample consists of n observations, where only d lifetimes (d an integer), $x_1 < x_2 < ... < x_d$ are measured fully, while the remainder $n - d$ are censored. These $n - d$ censored observations are ordered separately and are denoted by $x_{d+1} < x_{d+2} < ... < x_n$. In the context of reliability analysis (for example), a lifetime would be the time until a unit / machine fails to operate successfully.

In the paper of Soliman (2000), a family of non-informative priors were introduced:

$$g(\theta) = \frac{1}{\theta^m}\,, \qquad m, \theta > 0\,, \tag{1.2}$$

and was termed a "quasi-density" prior family. This paper explores the application of this prior family with regards to the right censored Rayleigh model. Different known prior densities are contained within (1.2), namely the Jeffreys prior ($m = 1$), Hartigan's prior ($m = 3$), and a third prior illustrating the diminishing effect of the prior density family — this is termed a "vanishing" prior (some large value of m, chosen arbitrarily such that $m = 10$). The choice of m would be up to the practitioner to determine the extent of the objectivity required. It is worth noting that Hartigan's prior ($m = 3$) is known as an asymptotically invariant prior as well. Liang (2008) provides valuable contributions when considering relevant choices of hyperparameters.

Mostert (1999) showed that the likelihood of the censored Rayleigh model is given by

$$\mathcal{L}(\theta) \propto (2\theta)^d u e^{-\theta T}$$

where $T = \sum_{i=1}^{n} x_i^2 \sim gamma(\frac{n}{2}, \theta)$. The quantity u is defined as $u = \prod_{i=1}^{d} x_i$, see Mostert (1999) for further details. It can be shown that the posterior distribution results in

$$g(\theta|T) = \frac{T^{d-m+1}}{\Gamma(d-m+1)} \theta^{(d-m+1)-1} e^{-\theta T} \tag{1.3}$$

which characterizes a $gamma(d-m+1, T)$ distribution, where $\Gamma(\cdot)$ denotes the gamma function. Note that, since the posterior distribution is always a proper distribution, it ensures the need of restrictions on the parameter space. In order for (1.3) to be well-defined, it is thus assumed throughout that $m < d+1$.

Together with Soliman (2000), Mostert (1999) compared the Bayesian estimators under the linear exponential (LINEX) loss function and squared-error loss (SEL) function, and Dey and Dey (2011) did similar work for the complete model by applying Jeffreys prior and a loss function as proposed by Al-Bayyati (2002). This paper extends concepts in the literature for the *censored* Rayleigh model by considering this new loss function, namely the Al-Bayyati loss (ABL), and comparing it to other known results.

Gruber (2004) proposed a method where a balanced loss function is used in a Bayesian context. A balanced loss function is where a weighted *loss* value is constructed by substituting each estimate into its corresponding loss function and determining some weighted value thereof. In this paper an extension of this methodology is considered, by obtaining a new *estimator* as a weighted value of the Bayesian estimator under either SEL or ABL, and some other estimate of the unknown parameter (in this case, θ). This is also known as a shrinkage based estimation approach.

The focus of this paper is the evaluation of the ABL estimator in terms of its performance by considering its risk efficiency in comparison to the SEL estimator, and also the effect of the parameter m, the prior density family degree. In this respect the following proposal is adopted:

1. Obtain the Bayes estimator under SEL, and evaluate under ABL;
2. Obtain the Bayes estimator under ABL, and evaluate under SEL; and
3. Obtain shrinkage estimators of both SEL and ABL estimators by combining the Bayesian estimators with some prespecified point estimate of the parameter, and evaluate under SEL.

In Section 2 the respective Bayesian estimators are determined and the risk (expected loss) are studied comparatively. The effect of risk efficiency is

also investigated, and a shrinkage approach is also then considered. In section 3 an illustrative example involving a simulation study and a real data analysis presented, and section 4 contains a discussion and some final conclusions.

2. SQUARED-ERROR LOSS (SEL) & AL-BAYYATI LOSS (ABL)

2.1. Parameter estimation under SEL & ABL

This section explores the Bayesian estimators under the loss functions for the model discussed in the introduction. The SEL is defined by

$$L_{SEL}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 \tag{2.1}$$

and the loss function proposed by Al-Bayyati (2002):

$$L_{ABL}(\hat{\theta}, \theta) = \theta^c(\hat{\theta} - \theta)^2 \ , \qquad c \in \mathbb{R} \ . \tag{2.2}$$

SEL is a widely used loss function due to its attractive feature of symmetry — where the function focuses on the size of the loss rather than the direction (over- or underestimation) of the loss. The ABL introduces the additional parameter c, which assists in determining a flatter loss function (albeit still symmetric) or the alternative, and it specifically generalizes the SEL (2.1). c can also be considered the order of weighting of the quadratic component. Under SEL, the (posterior) risk function has the following form:

$$\begin{aligned} R_{SEL}(\hat{\theta}, \theta) &= \int_0^\infty L_{SEL}(\hat{\theta}, \theta) g(\theta|T) d\theta \\ &= \hat{\theta}^2_{SEL} - 2\hat{\theta}_{SEL} \frac{\Gamma(d-m+2)}{\Gamma(d-m+1)T} + \frac{\Gamma(d-m+3)}{\Gamma(d-m+1)T^2} \ . \end{aligned}$$

From (1.3) the Bayesian estimator $\hat{\theta}_{SEL}$ is given by the posterior mean of θ:

$$\hat{\theta}_{SEL} = \frac{d-m+1}{T} \ . \tag{2.3}$$

Since (1.1) indicates that the parameter θ must be positive, a restriction implied by (2.3) is that $m < d+1$ (corresponding to the restriction discussed in the Introduction regarding the posterior distribution). Under ABL, the (posterior) risk

function has the following form:

$$\begin{aligned} R_{ABL}(\hat{\theta},\theta) &= \int_0^\infty L_{ABL}(\hat{\theta},\theta)g(\theta|T)d\theta \\ &= \hat{\theta}^2_{ABL}\frac{\Gamma(d-m+c+1)}{\Gamma(d-m+1)T^c} - 2\hat{\theta}_{ABL}\frac{\Gamma(d-m+c+2)}{\Gamma(d-m+1)T^{c+1}} \\ &\quad + \frac{\Gamma(d-m+c+3)}{\Gamma(d-m+1)T^{c+2}}\,. \end{aligned}$$

The Bayesian estimator $\hat{\theta}_{ABL}$ is

$$\hat{\theta}_{ABL} = \frac{d-m+c+1}{T}\,. \tag{2.4}$$

Similar to the case of the SEL estimator, $m < d+c+1$ for positive c, and $m+c < d+1$ for negative c in order for the gamma function to be well-defined.

2.2. Comparing the risk of SEL and ABL

The three different prior degrees are of interest here, namely the Jeffreys prior ($m = 1$), Hartigan's prior ($m = 3$), and the vanishing prior ($m = 10$). The posterior risk of the two loss functions was compared against each other for certain parameter values — notably for increasing values of θ and for the three different values of m.

The risk was determined empirically by simulating 5000 samples of sizes $n = 30, 40$ and 50 each, using the inverse-transform method and $uniform\,(0,1)$ random variates. From each of these obtained samples, the parameter was estimated under SEL and ABL (with $c = 0.5$), and the average loss of all 5000 samples was determined. The value of d was set at $d = 0.2n$, which implies that 20% of lifetimes have been observed. There are practical examples were a censoring of between 70% and 90% have been observed (see Stablein, Carter, and Novak (1981)), which is why, as an illustration, a censoring of 80% is used.

In Figures 1 to 3 it is seen that the shape of the functions do not change for different values of m, but it is observed that the risk is increasing for larger m values. Also, as the sample size n increases, the magnitude of the risk is decreasing. From the simulation it is evident that for positive c, SEL has least risk and would thus be preferable. An effective way of comparing the risk of different loss functions is by determining the risk efficiency — which is explored in the next section.

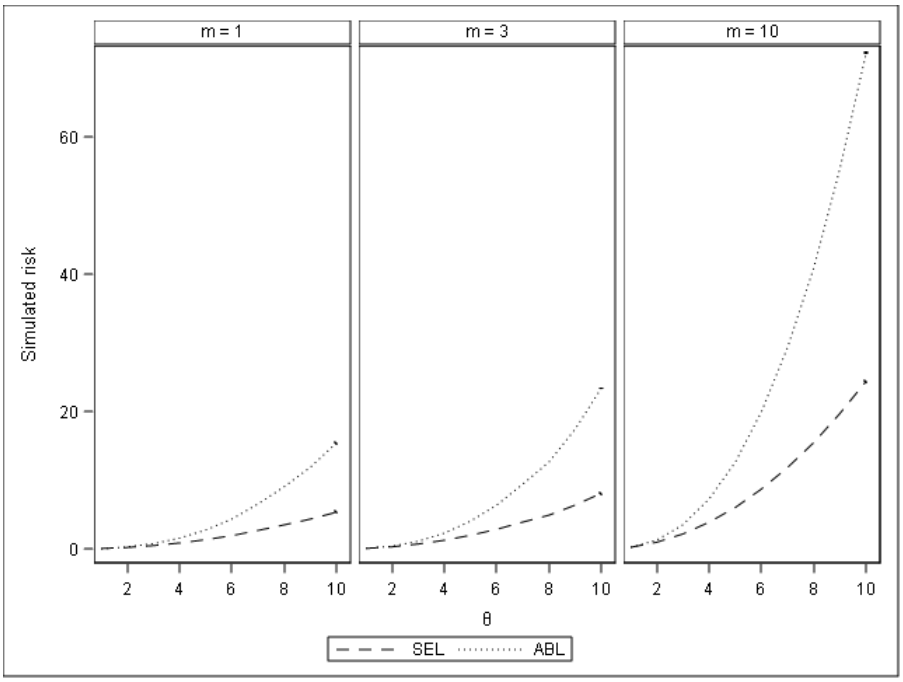

Figure 1: Simulated risk for SEL and ABL ($n = 30$).

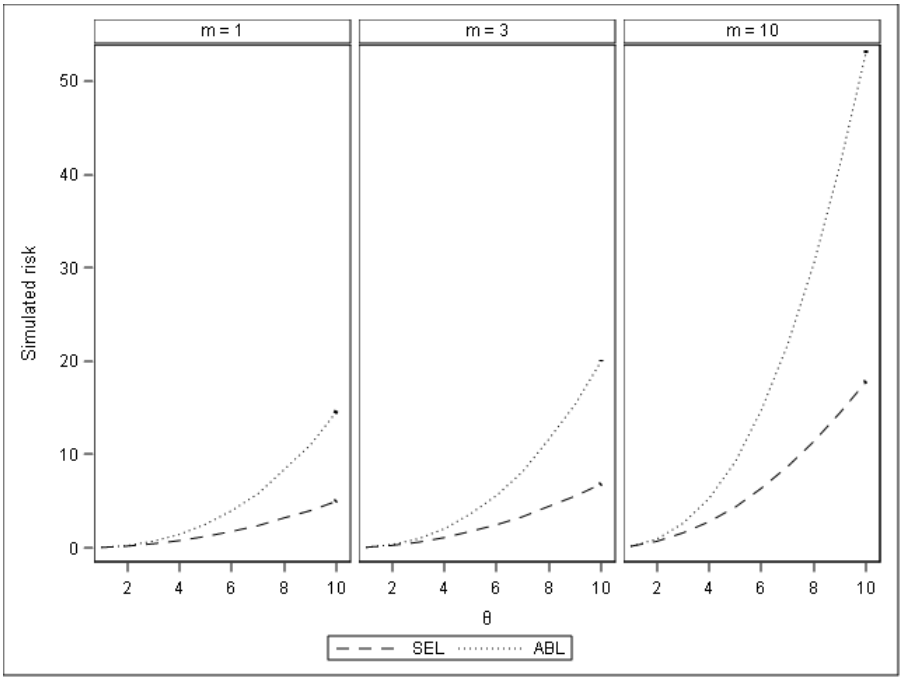

Figure 2: Simulated risk for SEL and ABL ($n = 40$).

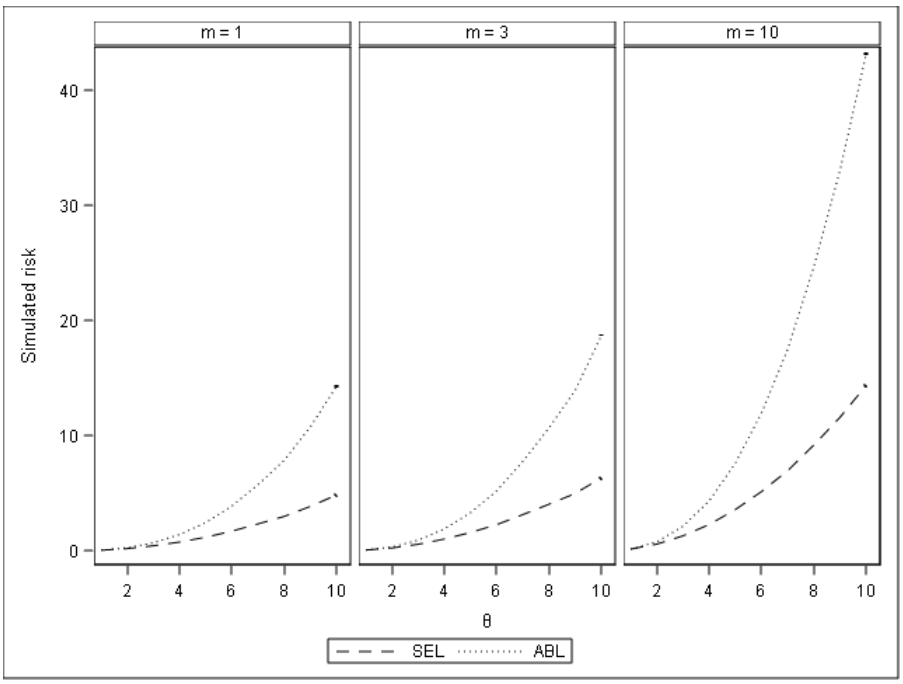

Figure 3: Simulated risk for SEL and ABL ($n = 50$).

2.3. Risk efficiency between SEL and ABL

Risk efficiency is a method that provides an intuitive way of determining which estimator — under a certain loss function — performs better than the other. The form of the risk function considered is

$$R_L^*(\hat{\theta}_{est}, \theta) \;=\; E_T(L(\hat{\theta}_{est}, \theta)) \;=\; \int_0^\infty L(\hat{\theta}_{est}, \theta) f(T) dT$$

using the distribution of T. Here, L denotes the loss function under which the risk efficiency is calculated, and $\hat{\theta}_{est}$ denotes its estimator of θ. The risk efficiency is then given by:

$$RE_L(\hat{\theta}_L, \hat{\theta}_y) \;\equiv\; \frac{R_L^*(\hat{\theta}_y, \theta)}{R_L^*(\hat{\theta}_L, \theta)}$$

translating to, the risk efficiency of $\hat{\theta}_L$ with respect to $\hat{\theta}_y$ under L loss ($\hat{\theta}_y$ denotes an estimator under any other loss function than L). This is similar as the approach by Dey (2011). Now, $\hat{\theta}_L$ denotes the estimator for the parameter that needs to be estimated under loss L, and $\hat{\theta}_y$ denotes the estimator for the parameter under the loss y. The interpretation of this expression is that when $RE_L(\hat{\theta}_L, \hat{\theta}_y) > 1$, the estimator $\hat{\theta}_L$ is preferable under L loss than that of $\hat{\theta}_y$.

2.3.1. SEL vs. ABL under SEL

The risk efficiency for the estimators derived in section 2.1 under SEL are given by:

$$RE_{SEL}(\hat{\theta}_{SEL}, \hat{\theta}_{ABL}) = \frac{R^*_{SEL}(\hat{\theta}_{ABL}, \theta)}{R^*_{SEL}(\hat{\theta}_{SEL}, \theta)}.$$

The expressions required by above equation are obtained as:

$$\begin{aligned} R^*_{SEL}(\hat{\theta}_{ABL}, \theta) &= \int_0^\infty L_{SEL}(\hat{\theta}_{ABL}, \theta) f(T) dT \\ &= \theta^2 \left(\frac{(d-m+1+c)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1+c}{(\frac{n}{2}-1)} + 1 \right) \end{aligned}$$

and

$$\begin{aligned} R^*_{SEL}(\hat{\theta}_{SEL}, \theta) &= \int_0^\infty L_{SEL}(\hat{\theta}_{SEL}, \theta) f(T) dT \\ &= \theta^2 \left(\frac{(d-m+1)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1}{(\frac{n}{2}-1)} + 1 \right). \end{aligned}$$

The risk efficiency of $\hat{\theta}_{sel}$ with respect to $\hat{\theta}_{abl}$ under SEL is then

$$\begin{aligned} RE_{SEL}(\hat{\theta}_{SEL}, \hat{\theta}_{ABL}) &= \frac{R^*_{SEL}(\hat{\theta}_{ABL}, \theta)}{R^*_{SEL}(\hat{\theta}_{SEL}, \theta)} \\ &= \frac{\left(\frac{(d-m+1+c)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1+c}{(\frac{n}{2}-1)} + 1 \right)}{\left(\frac{(d-m+1)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1}{(\frac{n}{2}-1)} + 1 \right)}. \end{aligned} \tag{2.5}$$

An interesting characteristic of this equation (2.5) is that it is independent from the sample information i.e. independent of x_i. It is only dependent on n, d, c, and m.

Figure 4 illustrates the risk efficiency (2.5) for arbitrary parameter values. Since the function is not dependent on sample information, no simulation from (1.1) is required. A sample size of $n = 30$ was specified along with $d = 0.2n$ and for different values of c. The risk efficiency values is plotted against values of m, the prior family degree. It is of special interest that for negative values of c, the ABL estimator performs better than that of the SEL counterpart for small values of m. The converse holds when this "threshold" value of m is reached, where the more efficient estimator becomes the SEL estimator.

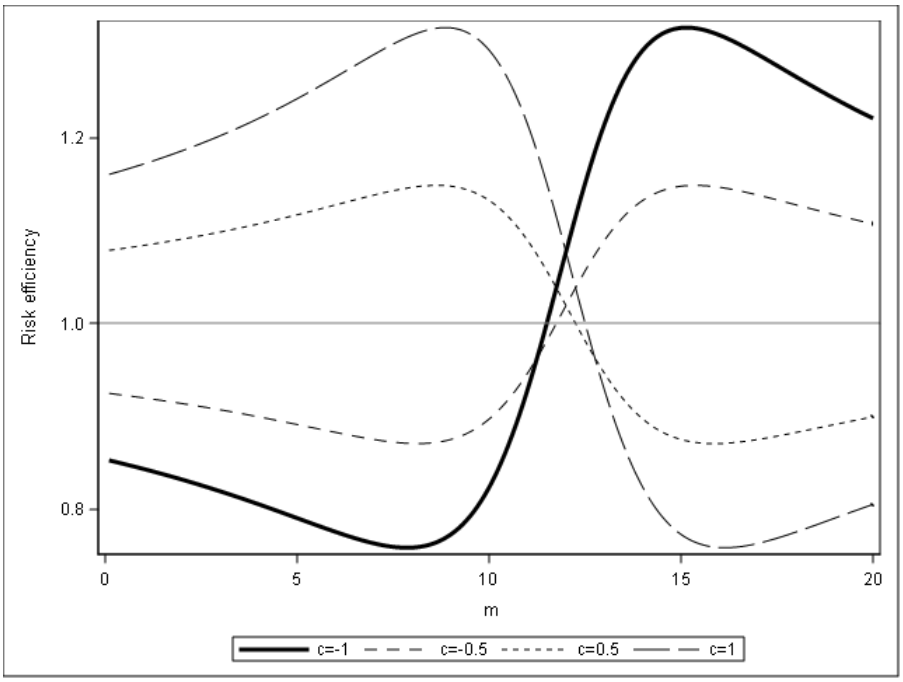

Figure 4: Risk efficiency of SEL- and ABL estimator under SEL.

2.3.2. ABL vs. SEL under ABL

The risk efficiency for SEL and ABL under ABL is given by:

$$RE_{ABL}(\hat{\theta}_{ABL}, \hat{\theta}_{SEL}) = \frac{R^*_{ABL}(\hat{\theta}_{SEL}, \theta)}{R^*_{ABL}(\hat{\theta}_{ABL}, \theta)}.$$

The expressions required by above equation are obtained as:

$$\begin{aligned} R^*_{ABL}(\hat{\theta}_{SEL}, \theta) &= \int_0^\infty L_{ABL}(\hat{\theta}_{SEL}, \theta) f(T) dT \\ &= \theta^{c+2} \left(\frac{(d-m+1)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1}{(\frac{n}{2}-1)} + 1 \right) \end{aligned}$$

and

$$\begin{aligned} R^*_{ABL}(\hat{\theta}_{ABL}, \theta) &= \int_0^\infty L_{ABL}(\hat{\theta}_{ABL}, \theta) f(T) dT \\ &= \theta^{c+2} \left(\frac{(d-m+1+c)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1+c}{(\frac{n}{2}-1)} + 1 \right) \end{aligned}$$

again using the relations derived in section (2.3.1). The risk efficiency of $\hat{\theta}_{abl}$ versus $\hat{\theta}_{sel}$ under ABL is:

$$
\begin{aligned}
RE_{ABL}(\hat{\theta}_{ABL}, \hat{\theta}_{SEL}) &= \frac{R^*_{ABL}(\hat{\theta}_{SEL}, \theta)}{R^*_{ABL}(\hat{\theta}_{ABL}, \theta)} \\
&= \frac{\left(\frac{(d-m+1)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1}{(\frac{n}{2}-1)} + 1\right)}{\left(\frac{(d-m+1+c)^2}{(\frac{n}{2}-1)(\frac{n}{2}-2)} - 2\frac{d-m+1+c}{(\frac{n}{2}-1)} + 1\right)} .
\end{aligned} \tag{2.6}
$$

It is observed that this last result is the reciprocal of the (2.5). Figure 5 illustrates this result; where the converse of the discussion of (2.5) holds.

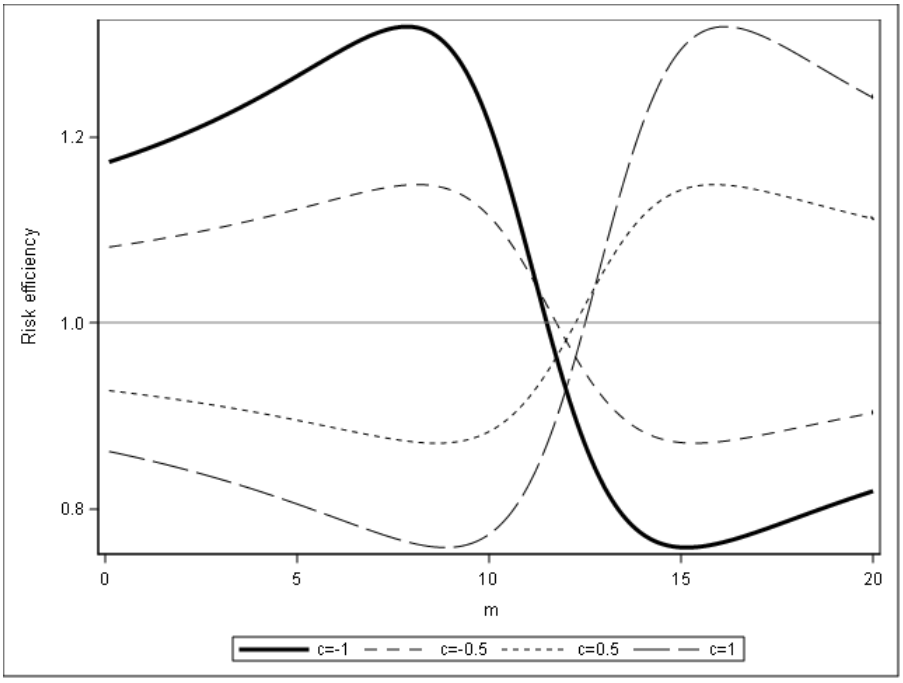

Figure 5: Risk efficiency of SEL- and ABL estimator under ABL.

2.4. Shrinkage estimation approach

Gruber (2004) proposed a method where a balanced loss function is used for Bayesian analysis. A balanced loss function is where a weighted loss value is constructed by substituting each estimate into its corresponding loss function and determining some weighted value thereof. As a slight twist on this approach, consider obtaining a new estimator as a weighted value of the Bayesian estimator

under either SEL or ABL, and some other estimate of the unknown parameter (in this case, θ). This is also known as a shrinkage based estimation approach. Define the SEL-based Bayesian shrinkage estimator by

$$\hat{\theta}_{S_1} = \lambda\hat{\theta}_{SEL} + (1-\lambda)\theta_o \ , \qquad 0 \leq \lambda \leq 1 \ , \tag{2.7}$$

and the ABL-based Bayesian shrinkage estimator by

$$\hat{\theta}_{S_2} = \lambda\hat{\theta}_{ABL} + (1-\lambda)\theta_o \ , \qquad 0 \leq \lambda \leq 1 \ . \tag{2.8}$$

where θ_o is a pre-specified point realization of θ. Similar as in the case of the SEL- and ABL estimators, these two newly proposed estimators ((2.7) and (2.8)) is compared in terms of their risk functions. The analysis here is only considered under the SEL. For the SEL-based shrinkage Bayesian estimator we have from (2.1) and (2.7)

$$\begin{aligned} R^*_{SEL}(\hat{\theta}_{S_1},\theta) &= E_T\left(\lambda\hat{\theta}_{SEL} - \lambda\theta + \lambda\theta + (1-\lambda)\theta_0 - \theta\right)^2 \\ &= \lambda^2\left(\theta^2\left(\frac{(d-m+1)^2}{\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)} - \frac{2(d-m+1)}{\left(\frac{n}{2}-1\right)} + 1\right)\right) \\ &\quad + (1-\lambda^2)(\theta_0-\theta)^2 \\ &\quad + 2\lambda(1-\lambda)\left(\theta_0 E_T(\hat{\theta}_{SEL}) - \theta E_T(\hat{\theta}_{SEL}) - \theta\theta_0 + \theta^2)\right) \end{aligned}$$

where $E_T(\hat{\theta}_{SEL}) = (d-m+1)\frac{\theta}{\left(\frac{n}{2}-1\right)}$, using the expected value of the gamma distribution of T. The ABL-based shrinkage Bayesian estimator is, from (2.2) and (2.8), given by

$$\begin{aligned} R^*_{SEL}(\hat{\theta}_{S_2},\theta) &= E_T\left(\lambda\hat{\theta}_{ABL} - \lambda\theta + \lambda\theta + (1-\lambda)\theta_0 - \theta\right)^2 \\ &= \lambda^2\left(\theta^2\left(\frac{(d-m+1+c)^2}{\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)} - \frac{2(d-m+1+c)}{\left(\frac{n}{2}-1\right)} + 1\right)\right) \\ &\quad + (1-\lambda^2)(\theta_0-\theta)^2 \\ &\quad + 2\lambda(1-\lambda)\left(\theta_0 E_T(\hat{\theta}_{ABL}) - \theta E_T(\hat{\theta}_{ABL}) - \theta\theta_0 + \theta^2)\right) \end{aligned}$$

where $E_T(\hat{\theta}_{ABL}) = (d-m+1+c)\frac{\theta}{\left(\frac{n}{2}-1\right)}$. When this method is repeated with ABL as the underlying loss functions, similar expressions are obtained but in a scaled form (stemming from the scaling value θ^c from the ABL), and is omitted here.

2.4.1. Risk comparison under SEL and ABL for shrinkage estimators

A similar approach was followed as in Dey (2011) and as discussed in section 2.2, but in this instance the shrinkage estimators were considered with the true risk. Again because of the inferential nature of the ABL, it is only discussed here for the SEL. Two viewpoints were considered: the first of which was for different prior point estimates and for varying λ, and the second was for fixed prior point estimate, different values of m, and for varying λ. This was all considered in the same simulated data setting as in section 2.2, with the addition that the "true" value of θ was 10. An underestimated value, an overestimated value, together with the MLE of θ was considered; i.e. $\theta_0 = 7$, 7.7625, and 15 (here, $\hat{\theta}_{MLE} = \frac{d}{T}$). Figure 6 illustrates the effect of these different prior point estimates and $m = 1$, whilst Figure 7 illustrates for different values of m and the prior point estimate equal to the MLE of the censored Rayleigh distribution. The two figures illustrate these effects.

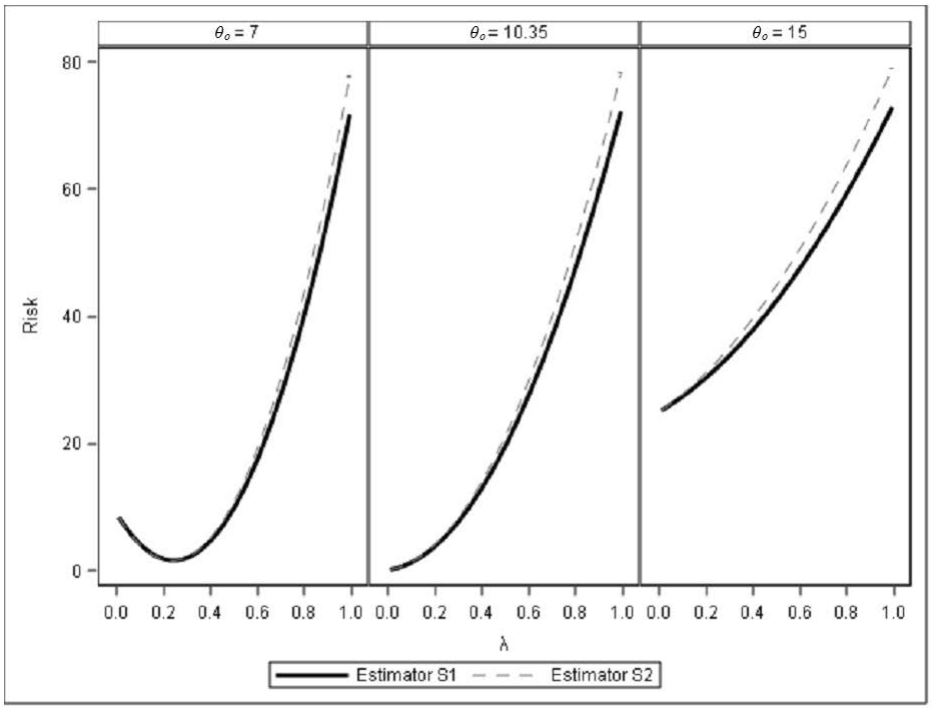

Figure 6: Risk under SEL for shrinkage estimators $\hat{\theta}_{S_1}$ and $\hat{\theta}_{S_2}$, different θ_0, and varying λ ($m = 1$ (fixed)).

As can be seen in both cases, least risk can be obtained for some nonzero and nonunity value of λ, except for the case depicted in Figure 7 when $m = 10$.

This however makes little practical sense if not viewed in comparison with that of the "original" risk for only the Bayesian estimators. In the next section, this comparison is explored with reference to the risk efficiency.

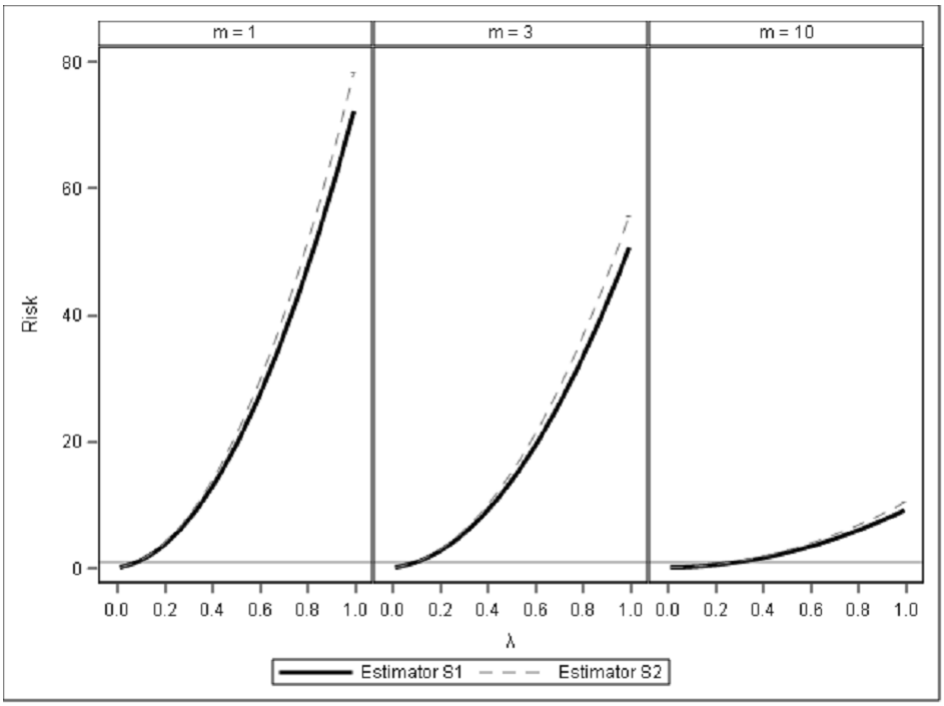

Figure 7: Risk under SEL for shrinkage estimators $\hat{\theta}_{S_1}$ and $\hat{\theta}_{S_2}$, different m, and varying λ ($\theta_0 = MLE$ (fixed)).

2.4.2. Risk efficiency under SEL and ABL for shrinkage estimators

Now, the risk efficiency for the shrinkage estimators was determined under these two loss functions. The following comparisons are considered:

$$RE_{SEL}(\hat{\theta}_{SEL}, \hat{\theta}_{S_1}) = \frac{R^*_{SEL}(\hat{\theta}_{S_1}, \theta)}{R^*_{SEL}(\hat{\theta}_{sel}, \theta)} \tag{2.9}$$

and

$$RE_{ABL}(\hat{\theta}_{ABL}, \hat{\theta}_{S_2}) = \frac{R^*_{ABL}(\hat{\theta}_{S_2}, \theta)}{R^*_{ABL}(\hat{\theta}_{ABL}, \theta)} \, . \tag{2.10}$$

The same parameter choices as used previously was employed here, and different values of θ_0 were chosen arbitrarily, to assist with the comparison.

The prior density degree was $m = 1$, the Jeffreys prior, and the true value of θ from which the observations were simulated from, is 10. Three values were considered: a value that underestimates the true value of θ, the MLE, and a value that overestimated the true value of θ. Two considerations were examined and is illustrated by the respective figures below. Figure 7 illustrates the risk efficiency under SEL for varying λ, and these different prior point estimates. Figure 8 illustrates the same, but for the case where the underlying loss function is ABL. For these illustrative purposes, the ABL constant c was set to 0.5.

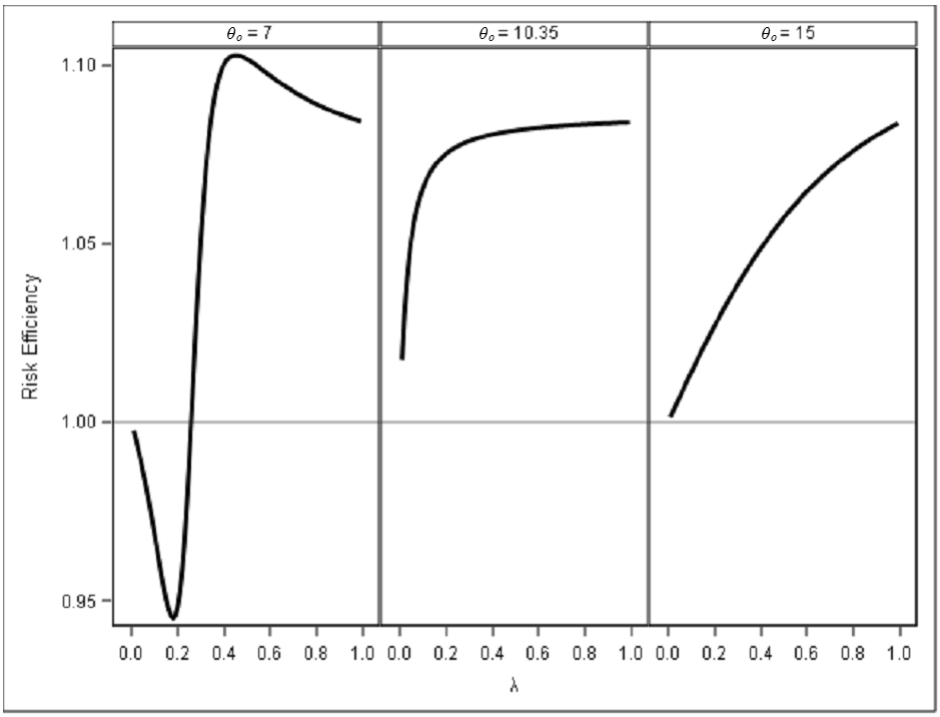

Figure 8: Risk efficiency under SEL for shrinkage estimators $\hat{\theta}_{S_1}$ and $\hat{\theta}_{S_2}$, different θ_0, and varying λ.

Figure 8 clearly shows that there is indeed some shrinkage estimator value (i.e. $0 \leq \lambda \leq 0.25$) that is more appropriate to use than the the true corresponding Bayesian estimator (for a risk efficiency value of < 1). This seems only true for the case of underestimation ($\theta_0 = 7$). For the case of the MLE and overestimation ($\theta_0 = 15$), only the Bayesian estimate seems appropriate. Figure 9 shows the reciprocal results, where the shrinkage estimator seems more appropriate to use in overestimation.

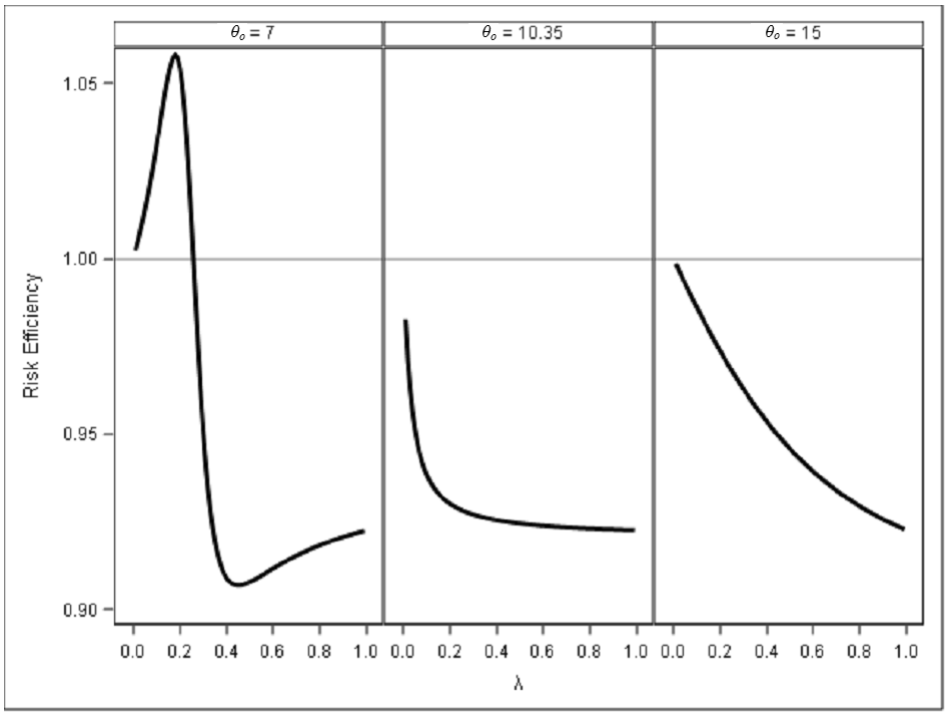

Figure 9: Risk efficiency under ABL for shrinkage estimators $\hat{\theta}_{S_1}$ and $\hat{\theta}_{S_2}$, different θ_0, and varying λ.

3. ILLUSTRATIVE EXAMPLES

3.1. Simulation study

In this section, the $RMSE$ (root mean square error) comparison of the SEL estimator (2.3), the ABL estimator (2.4), and the shrinkage counterparts (2.7) and (2.8) is calculated via simulation. It is known that an estimator with least $RMSE$ is considered preferable. As the parameter θ in (1.1) indicates a lifetime, it is important to use an estimator which estimates the true value of the population parameter as closely as possible, otherwise the chosen estimator may overestimate or underestimate the value too severely, resulting in catastrophic events in real life. For example, when estimating the lifetime of an airplane engine, underestimating the lifetime is much less serious than overestimating the lifetime of the engine. By using the $RMSE$ the estimator which exhibits the smallest error in estimation can be determined.

The $RMSE$ is given by $RMSE = \sqrt{\frac{\sum_{i=1}^{p}\left(\hat{\theta}_{est}-\theta\right)^2}{p}}$, where p denotes the number of observations of θ. $\hat{\theta}_{est}$ denotes the estimated value of θ under a specific loss function. The following steps outline the method followed in this simulation.

1. Simulate $p = 5000$ random samples from (1.1) for a given value of θ. From each simulated sample, determine $\hat{\theta}_{est}$ under SEL, ABL, and both considered shrinkage estimators (for the shrinkage estimators, the value of $\theta_0 = MLE$). Then, calculate the value of the $RMSE$.
2. Repeat Step 1 for a successive range of θ values, in this case, $\theta = 1...40$.
3. Plot the $RMSE$ for all four estimators upon the same set of axis. The estimator with lowest $RMSE$ is considered the preferable estimator.

Figure 10 and 11 shows the results for different choices of λ.

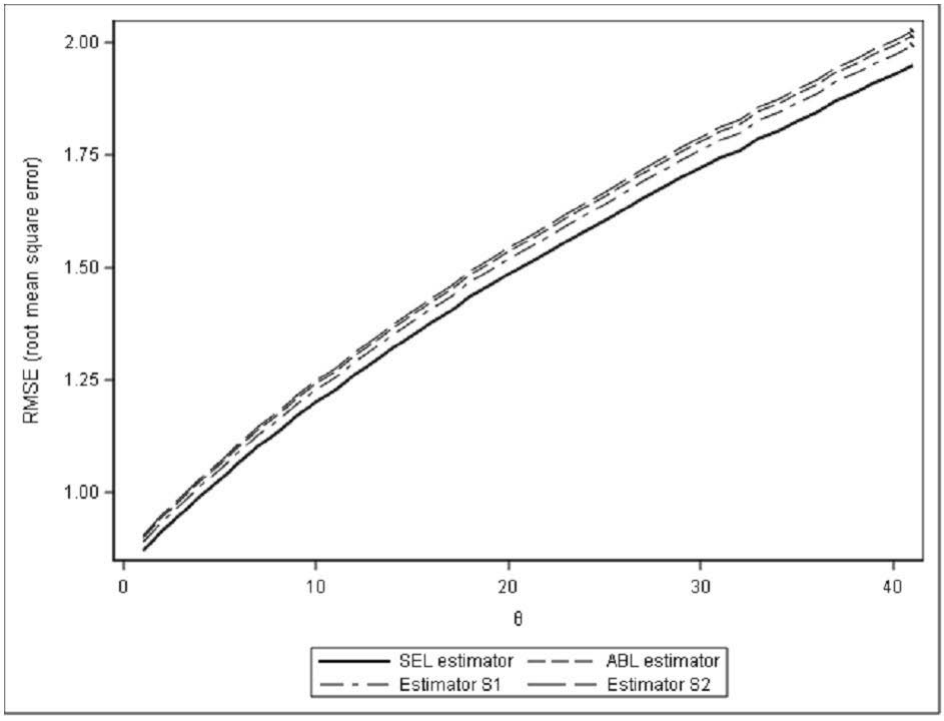

Figure 10: Root mean square error for $\hat{\theta}_{est}$ under SEL, ABL, S_1 and S_2 where $\theta_0 = MLE$, and varying θ ($m = 1$ (fixed), $c = 0.5$, $\lambda = 0.5$).

It is observed that the SEL estimator is preferable for the considered Rayleigh model against that of the ABL estimator, and both considered shrinkage

estimators. The SEL estimator is also preferable to its corresponding shrinkage estimator, and the ABL estimator is also preferable to its corresponding shrinkage estimator. These are for the cases when the MLE and the Bayesian estimate carries equal weight in the shrinked estimator.

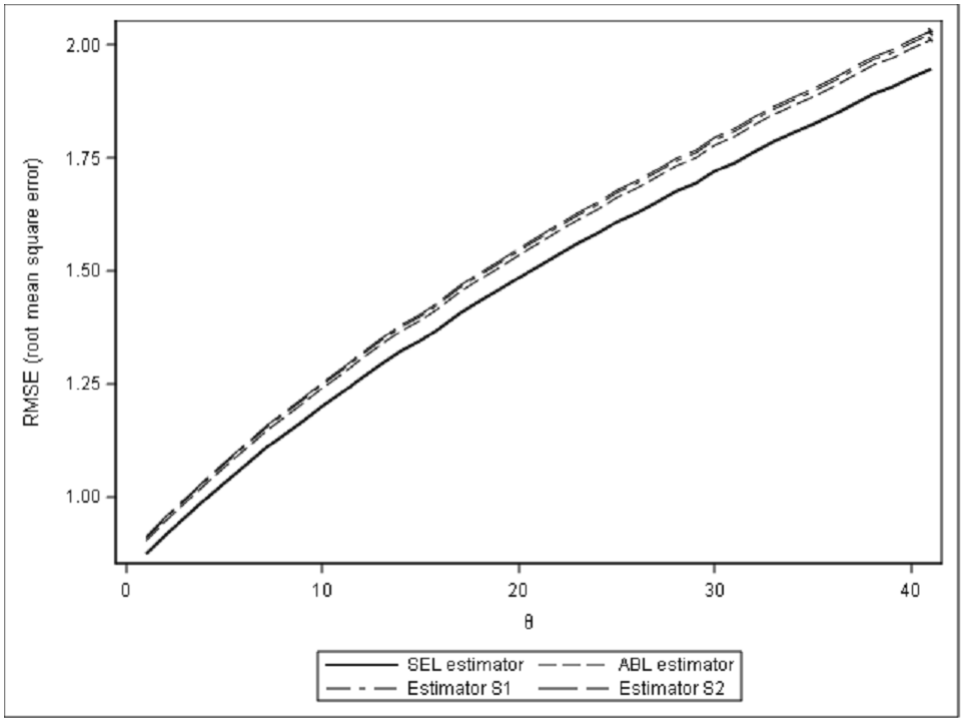

Figure 11: Root mean square error for $\hat{\theta}_{est}$ under SEL, ABL, S_1 and S_2 where $\theta_0 = MLE$, and varying θ ($m = 1$ (fixed), $c = 0.5$, $\lambda = 0.1$).

Figure 11 shows the case when the weight of the shrinkage estimators are skewed toward the MLE. Even in this case, both Bayesian estimates are preferred compared to their respective shrinkage estimators.

3.2. Practical application: gastrointestinal tumor group

The results are illustrated using gastrointestinal tumor study group data, obtained from Stablein, Carter, and Novak (1981) from a clinical trial in the treatment of locally advanced nonresectable gastric carcinoma. Mostert (1999) showed that the Rayleigh model is suitable for this data — it is also of censored nature which applies here. The sample size is $n = 45$, and the number of fully

observed lifetimes is $d = 37$, where $T = 133.643$. The MLE of θ was used as the estimate θ_0. Table 1 below gives the parameter estimates under different loss function ((2.3) and (2.4)) for different parameter combinations.

Table 1: Parameter estimates under SEL and ABL for the real data set, for different values of m and c.

Value of m	Estimate value	$c = -1$	$c = -0.5$	$c = 0.5$	$c = 1$
$m = 1$	$\hat{\theta}_{SEL} = 0.27685$				
	$\hat{\theta}_{MLE} = 0.27685$				
	$\hat{\theta}_{ABL}$	0.26937	0.27311	0.28059	0.28433
$m = 3$	$\hat{\theta}_{SEL} = 0.26189$				
	$\hat{\theta}_{MLE} = 0.27685$				
	$\hat{\theta}_{ABL}$	0.25440	0.25815	0.26563	0.26937
$m = 10$	$\hat{\theta}_{SEL} = 0.20951$				
	$\hat{\theta}_{MLE} = 0.27685$				
	$\hat{\theta}_{ABL}$	0.20203	0.20577	0.21325	0.21699

This example aims to emphasize the effect of the shrinkage effect of the respective shrinkage estimators ((2.7) and (2.8)) and was achieved via a bootstrapping approach. By using the bootstrap method, a sampling distribution of the mentioned estimators can be constructed, and determined whether the estimator has a convergent nature — also, to have small standard error. The convergent nature of the bootstrap in parameter estimation is expected to illustrate the shrinkage effect to determine which estimator seems more appropriate for the given data set.

As mentioned, the performance of the estimator was studied via bootstrapping from the sample $k = 1000$ times. Thus, 1000 samples were drawn from the original sample with replacement, and for each of the drawn samples, the estimator under SEL was computed, and the risk value. The risk value was computed via

$$R^*\left(\hat{\theta}_{SEL}, \theta\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\hat{\theta}_{S_1,i} - \theta\right)^2$$

where $\hat{\theta}_{S_1,i}$ is the shrinkage estimator (2.3) for the i^{th} bootstrapped sample, and θ the fixed sample parameter (determined via reparametrization of the mean of the distribution, equal to $\mu = \frac{1}{2}\sqrt{\frac{\pi}{\theta}}$, thus $\theta = \frac{\pi}{(2\mu)^2}$). This risk value was determined for increasing λ and graphed correspondingly, and is presented in Figure 12. It can be concluded that the estimator is indeed accurate and stable; in addition, from visual inspection it is observed that the estimator indeed has a small standard error. However, because of its near-convergent nature as $\lambda \to 1$, in this example, $\hat{\theta}_{SEL}$ is preferred to that of the MLE. This is in accordance with the $RMSE$

study in the preceeding section. This could be attributed to the shrinkage effect present in the shrinkage expression (2.7).

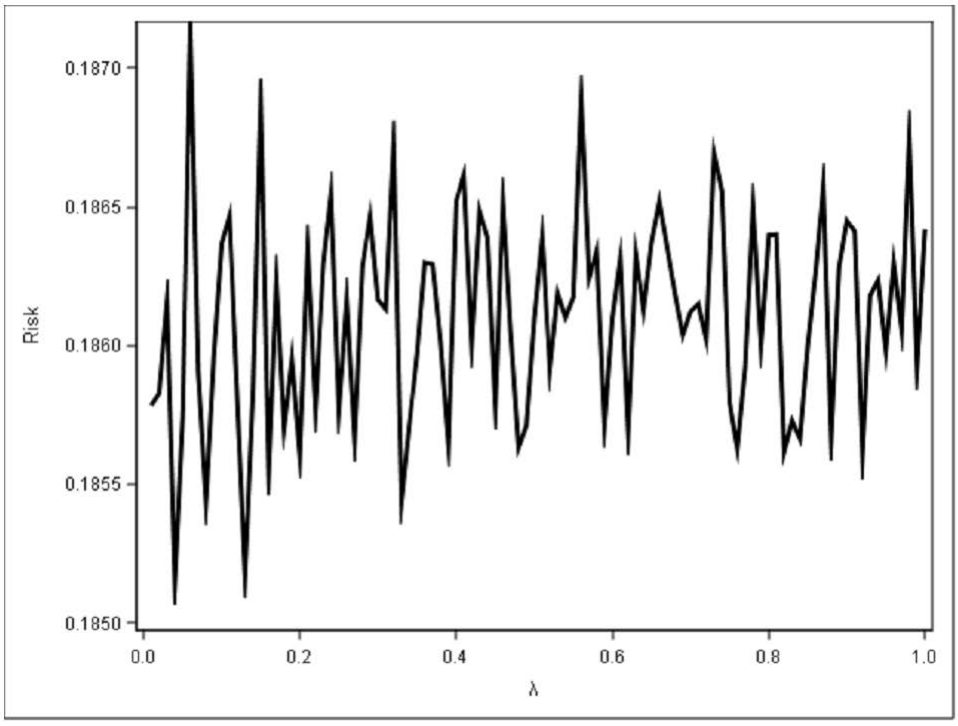

Figure 12: Bootstrap estimated values of $\hat{\theta}_{S_1,i}$, for $m = 1$ and increasing λ.

4. CONCLUSION

This paper explored the behaviour of the loss function proposed by Al-Bayyati (2002) by comparing it to the well-known squared error loss function. Bayes- and shrinkage estimators were derived. Their performance was studied under each of the mentioned loss functions in terms of their respective risk. It was observed that for positive values of c, the Al-Bayyati loss parameter, the risk of SEL was lower than that of ABL. Another focus of this paper was the effect of the prior family degree m. It was observed that the risk of both SEL and ABL became larger as m increased. In a risk efficiency perspective, it was seen that negative values of c results in the ABL estimator being more efficient under SEL since the risk is then smaller. The reciprocal result holds when the underlying loss function is the ABL. When the underlying loss is ABL, then for positive values of c the SEL estimator performs better in terms of risk.

After proposing shrinkage estimators (where the derived Bayesian estimators are combined in linear fashion with some pre-specified point estimate of the parameter) their risk and risk efficiency was also studied. It was observed that for underestimation of the parameter, the shrinkage estimator yielded lower risk than that of only the Bayesian estimator itself. For overestimation, only the Bayesian estimator performed better than the shrinkage estimator. In the risk efficiency setting it was observed that there does exist some values of λ which results in the shrinkage estimator under ABL performing better than the SEL estimator when the underlying loss function is SEL.

As a simulation study the RMSE was determined for each of the proposed estimators and subsequently compared. It was seen that the estimator under SEL remains preferable when considering the RMSE criterion. A numerical example also followed showing the applicational use of the estimators to a real data set.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Office of the Dean of the Faculty of Natural and Agricultural Sciences, University of Pretoria, for their financial assistance toward this study. In addition, the support from STATOMET, Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria is also humbly acknowledged. Finally the anonymous reviewer is thanked for his/her constructive comments and suggestions for greatly improving the quality of this paper.

REFERENCES

[1] Al-Bayyati, H.N. (2002). Comparing methods of estimating Weibull failure models using simulation, *Unpublished PhD thesis*, College of Administration and Economics, Baghdad University, Iraq.

[2] Bhattacharya, S.K. and Tyagi, R.K. (1990). Bayesian survival analysis based on the Raleigh model, *Trabajos de Estadistica*, **5**(1), 81–92.

[3] Delaportas, P. and Wright, D.E. (1991). Numerical prediction for the two-parameter Weibull distribution, *The Statistician*, **40**, 365–372.

[4] Dey, S. (2011). Comparison of relative risk functions of the Rayleigh distribution under Type II censored samples: Bayesian approach, *Jordan Journal of Mathematics and Statistics*, **4**(1), 61–68.

[5] Dey, S. and Dey, T. (2011). Rayleigh distribution revisited via an extension of Jeffreys prior information and a new loss function, *Revstat*, **9**(3), 213–226.

[6] Gruber, M.H.J. (2004). The efficiency of shrinkage estimators with respect to Zellner's balanced loss function, *Communications in Statistics — Theory and Methods*, **33**(2), 235–249.

[7] Liang, F.; Paulo, R.; German, G.; Clyde, M.A. and Berger, J.O. (2008). Mixtures of g priors for Bayesian variable selection, *Journal of the American Statistical Association*, **103**, 401–414.

[8] Mostert, P.J. (1999). A Bayesian method to analyse cancer lifetimes using Rayleigh models, *Unpublished PhD thesis*, University of South Africa.

[9] Soliman, A.A. (2000). Comparison of LINEX and quadratic Bayes estimators for the Rayleigh distribution, *Communications in Statistics — Theory and Methods*, **29**(1), 95–107.

[10] Stablein, D.M.; Carter, W.H. and Novak, J.W. (1981). Analysis of survival data with nonproportional hazard functions, *Controlled Clinical Trials*, **2**, 149–159.

REVSTAT – Statistical Journal
Volume 14, Number 4, October 2016, 455–479

ON HITTING TIMES FOR MARKOV TIME SERIES OF COUNTS WITH APPLICATIONS TO QUALITY CONTROL

Authors: Manuel Cabral Morais
– CEMAT and Department of Mathematics, Instituto Superior Técnico, Universidade de Lisboa, Portugal
maj@math.ist.utl.pt

António Pacheco
– CEMAT and Department of Mathematics, Instituto Superior Técnico, Universidade de Lisboa, Portugal
apacheco@math.ist.utl.pt

Received: July 2014 Revised: March 2015 Accepted: March 2015

Abstract:

- Examples of time series of counts arise in several areas, for instance in epidemiology, industry, insurance and network analysis. Several time series models for these counts have been proposed and some are based on the binomial thinning operation, namely the integer-valued autoregressive (INAR) model, which mimics the structure and the autocorrelation function of the autoregressive (AR) model.

  The detection of shifts in the mean of an INAR process is a recent research subject and it can be done by using quality control charts. Underlying the performance analysis of these charts, there is an indisputable popular measure: the run length (RL), the number of samples until a signal is triggered by the chart. Since a signal is given as soon as the control statistic falls outside the control limits, the RL is nothing but a hitting time.

  In this paper, we use stochastic ordering to assess:

  – the ageing properties of the RL of charts for the process mean of Poisson INAR(1) output;
  – the impact of shifts in model parameters on this RL.

  We also explore the implications of all these properties, thus casting interesting light on this hitting time for a Markov time series of counts.

Key-Words:

- *statistical process control; run length; phase-type distributions; stochastic ordering.*

AMS Subject Classification:

- 60E15, 62P30.

1. THE INAR(1) PROCESS

Time series of counts become apparent in areas such as:

- epidemiology — the number of new cases of some infectious and notifiable diseases is monitored monthly to assess and surveil the incidence of acute viral infections such as poliomyelitis, as reported by Zeger (1988) and Silva (2005, pp. 145–147);
- industry — the monthly number of accidents in a manufacturing plant (Silva *et al.*, 2009), the number of defects per sample (Weiss, 2009a) and the number non-conforming units within a sample of finite size counts (Weiss, 2009b,c) have to be controlled;
- insurance — modelling the number of claim counts is an extremely important part of insurance pricing (Boucher *et al.*, 2008);
- network analysis — the number of intrusions on computers and network systems (Weiss, 2009d, p. 11) also requires surveillance.

In some cases, the integer values of the time series are large and continuous-valued models can be (and are frequently) used. However, when the time series consists only of small integer numbers, ARMA processes are of limited use for modelling purposes, namely because the multiplication of an integer-valued random variable (r.v.) by a real constant may lead to a non-integer r.v. (Silva, 2005, p. 22).

A possible way out is to replace the scalar multiplication by a random operation, such as the binomial thinning operation. This operation can be thought as the scalar multiplication counterpart in the integer-valued setting which preserves the integer structure of the process, it is due to Steutel and Van Harn (1979) and may be stated as follows.

Definition 1.1. Let X be a discrete r.v. with range $\mathbb{N}_0 = \{0, 1, ...\}$ and α a scalar in $[0, 1]$. Then the binomial thinning operation on X results in the following r.v.:

$$\alpha \circ X = \sum_{i=1}^{X} Y_i \,, \tag{1.1}$$

where $\{Y_i : i \in \mathbb{N}\}$ is a sequence of i.i.d. Bernoulli(α) r.v. independent of X.

In this case $\alpha \circ X$ emerges from X by *binomial thinning*, and $\circ$ represents the *binomial thinning operator*. Furthermore, according to Steutel and Van Harn (1979), $\alpha \circ X$ also takes values in $\mathbb{N}_0$ and: $1 \circ X = X$; $0 \circ X = 0$; $\alpha \circ X \mid X =$

$x \sim \text{Bin}(x, \alpha)$; $E(\alpha \circ X) = \alpha \times E(X)$, as in scalar multiplication; $V(\alpha \circ X) = \alpha^2 \times V(X) + \alpha(1-\alpha) \times E(X)$, unlike in scalar multiplication.

Several models for time series of counts have been proposed based on the binomial thinning operation. These models are in general obtained as discrete analogues of the standard linear time series models. For example, the first-order integer-valued autoregressive (INAR(1)) model, introduced by McKenzie (1985) and Al-Osh and Alzaid (1987), mimics the structure and the autocorrelation function of the real-valued first-order autoregressive AR(1) model.

Definition 1.2. Let $\{\epsilon_t : t \in \mathbb{Z}\}$ be a sequence of nonnegative integer-valued independent and identically distributed (i.i.d.) r.v. with range $\mathbb{N}_0$, mean μ_ϵ and variance σ^2_ϵ, and α a scalar in $(0, 1)$. Then $\{X_t : t \in \mathbb{Z}\}$ is said to be a INAR(1) process if it satisfies the recursion

$$X_t = \alpha \circ X_{t-1} + \epsilon_t \,, \tag{1.2}$$

where: $\circ$ represents the binomial thinning operator; all thinning operations are performed independently of each other and of $\{\epsilon_t : t \in \mathbb{Z}\}$; the thinning operations at time t are independent of $\{..., X_{t-2}, X_{t-1}\}$; and ϵ_t and X_{t-1} are assumed to be independent r.v.

Besides taking only nonnegative integer values, the INAR(1) model also differs from the real-valued AR(1) model because the r.v. ϵ_t in this latter are usually interpreted as random noise, whereas in the INAR(1) model they introduce innovation to the process by *keeping the system alive* (Weiss, 2009d, p. 283) with *arrivals*.

Furthermore, the marginal distribution of the INAR(1) process can be expressed in terms of the r.v. ϵ_t (Silva, 2005, p. 35):

$$X_t \overset{d}{=} \sum_{j=0}^{+\infty} \alpha^j \circ \epsilon_{t-j} \tag{1.3}$$

(Al-Osh and Alzaid, 1987), the analogue of the moving average representation of the real-valued AR(1) model.

In the INAR(1) model setting, choosing an adequate family of distributions for the r.v. ϵ_t, say $\mathcal{F}$, so that X_t has a distribution that also belongs to $\mathcal{F}$, leads us to the class of discrete self-decomposable distributions defined by Steutel and Van Harn (1979): the r.v. X, with range $\mathbb{N}_0$, is said to have a discrete self-decomposable distribution if $X = \alpha \circ X' + X_\alpha$, where $\alpha \circ X'$ and X_α are independent, and X' is distributed as X.

It is worth mentioning that if ϵ_t has a discrete self-decomposable distribution such that $E(\epsilon_t) = \mu_\epsilon$ and $V(\epsilon_t) = \sigma^2_\epsilon < +\infty$ then the INAR(1) process is

second order weakly stationary with constant mean and variance function given by $E(X_t) = \frac{\mu_\epsilon}{1-\alpha}$ and $V(X_t) = \frac{\alpha\,\mu_\epsilon + \sigma_\epsilon^2}{1-\alpha^2}$, respectively (Weiss, 2009d, p. 283). Moreover, the INAR(1) and the AR(1) processes have similar autocorrelation function: $\rho_k = corr(X_t, X_{t-k}) = \alpha^{|k|}$, $k \in \mathbb{Z}$ (Weiss, 2009d, p. 285).

The class of discrete self-decomposable distributions contains the family of Poisson distributions (Silva, 2005, p. 35) and the Poisson INAR(1) process can be defined and characterised.

Definition 1.3. If $\epsilon_t \overset{i.i.d.}{\sim} \text{Poisson}(\lambda)$, $t \in \mathbb{Z}$, then $\{X_t = \alpha \circ X_{t-1} + \epsilon_t : t \in \mathbb{Z}\}$ is said to be a Poisson INAR(1) process.

The Poisson INAR(1) process is a second order weakly stationary process with marginal distribution

$$X_t \sim \text{Poisson}\left(\frac{\lambda}{1-\alpha}\right), \qquad t \in \mathbb{Z}\,, \tag{1.4}$$

and can be characterized as follows, according to Weiss (2009d, p. 283).

Proposition 1.1. *The Poisson INAR*(1) *process is a (time-)homogeneous Markov chain, with state space* $\mathbb{N}_0$ *and one-step transition probability matrix (TPM)* $\mathbf{P}$*, which depends on the values of* λ *and* α *and whose entries are given by*

$$\begin{aligned} p_{i\,j} &\equiv p_{i\,j}(\lambda, \alpha) \\ &= P(X_t = j \mid X_{t-1} = i) \\ &= \sum_{m=0}^{i} P(\alpha \circ X_{t-1} = m \mid X_{t-1} = i) \times P(\epsilon_t = j - m) \\ &= \sum_{m=0}^{\min\{i,j\}} \binom{i}{m} \alpha^m\,(1-\alpha)^{i-m} \times e^{-\lambda} \frac{\lambda^{j-m}}{(j-m)!}\,, \qquad i, j \in \mathbb{N}_0\,. \end{aligned} \tag{1.5}$$

The calculation of $\mathbf{P}$, for a few values of λ and α, led us to believe that no particular features are apparent in this matrix. For instance, even though X_t is a nonnegative r.v., $\mathbf{P}$ has no triangular block of zeros in the lower left hand corner or equal values along a line parallel to the main diagonal, such as the TPM Brook and Evans (1972) or Morais (2002) dealt with in a quality control setting.

Nevertheless, we managed to identify a peculiar and important feature of the TPM associated with a Poisson INAR(1) process: $\mathbf{P}$ is totally positive of order 2 (TP_2), i.e., it is a nonnegative matrix whose 2×2 minors are all nonnegative —

$$p_{i\,j} \times p_{i'\,j'} \;\geq\; p_{i'\,j} \times p_{i\,j'}\,, \qquad i < i',\;\; j < j' \tag{1.6}$$

—, as proved in the next section.

2. DISTINCTIVE FEATURES OF THE POISSON INAR(1) PROCESS

It is well known that the Poisson and binomial probability functions (p.f.),

- $f_{Poi(\lambda)}(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x \in \mathbb{N}_0$,
- $f_{Bin(n,p)}(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, ..., n$,

are log-concave in the sense that the likelihood ratio functions

$$\frac{f_{Poi(\lambda)}(x)}{f_{Poi(\lambda)}(x+1)} = \frac{x+1}{\lambda}$$

$$\frac{f_{Bin(n,p)}(x)}{f_{Bin(n,p)}(x+1)} = \frac{(x+1)(1-p)}{(n-x)p}$$

are nondecreasing functions of x over the supports of these p.f. That is to say, the Poisson and binomial distributions have what is also termed the Pólya frequency of order 2 (PF_2) property (Li and Shaked, 1997) or an increasing likelihood ratio (ILR), [1] the strongest ageing property that we consider here.

Furthermore, according to Casella and Berger (2002, p. 391), the families of Poisson and binomial p.f. have monotone likelihood ratio, in particular the following ones:

- $\{f_{Poi(\xi)}(x) : \xi > 0\}$;
- $\{f_{Bin(\xi,p)}(x) : \xi \in \mathbb{N}\}$ (here p is held fixed in $(0,1)$);
- $\{f_{Bin(n,\xi)}(x) : \xi \in (0,1)\}$ (n is fixed in $\mathbb{N}$).

For example, for $\xi_1 \leq \xi_2$,

$$\frac{f_{Poi(\xi_1)}(x)}{f_{Poi(\xi_2)}(x)} = e^{-(\xi_1-\xi_2)}\,(\xi_1/\xi_2)^x\,, \qquad x \in \mathbb{N}_0\,, \tag{2.1}$$

is a monotone — in this case nonincreasing — function of x. Interestingly enough, if we consider $P(x,\xi) \equiv f_{Poi(\xi)}(x)$ (or $\equiv f_{Bin(\xi,p)}(x)$, $f_{Bin(n,\xi)}(x)$) then $P(x,\xi)$ is a TP_2 function in x and ξ, i.e., the determinant

$$\begin{vmatrix} P(x_1,\xi_1) & P(x_1,\xi_2) \\ P(x_2,\xi_1) & P(x_2,\xi_2) \end{vmatrix} \geq 0\,, \qquad x_1 < x_2,\ \ \xi_1 < \xi_2 \tag{2.2}$$

(Karlin and Proschan, 1960).

[1] If you define the likelihood ratio function as $\frac{P(X=x+1)}{P(X=x)}$ instead, like Kijima (1997, p. 114) did, then the PF_2 property is equivalent to a decreasing likelihood ratio (DLR), as noted by Kijima (1997, p. 115).

Incidentally, the monotone likelihood ratio character — or TP_2 property — of a family of p.f. is related to the notion of stochastically smaller in the likelihood ratio sense (Ross, 1983, p. 281) stated below.

Definition 2.1. Let X and Y be two discrete r.v. with p.f. $P(X = x)$ and $P(Y = x)$. Then X is said to be stochastically smaller than Y in the likelihood ratio sense — denoted by $X \leq_{lr} Y$ — iff $\frac{P(X=x)}{P(Y=x)}$ is a nonincreasing function of x over the union of the supports of the r.v. X and Y (Shaked and Shanthikumar, 1994, pp. 27–28).

Expectedly, if a family of p.f. has monotone nonincreasing (resp. nondecreasing) likelihood ratio then the associated r.v. stochastically increase (resp. decrease) in the likelihood ratio sense — i.e., if $\xi_1 \leq \xi_2$ then $X(\xi_1) \leq_{lr} X(\xi_2)$ (resp. $X(\xi_1) \geq_{lr} X(\xi_2)$), in short $X(\xi) \uparrow_{lr}$ with ξ (resp. $X(\xi) \downarrow_{lr}$ with ξ). For the families of Poisson and binomial p.f. we have considered:

- $X(\xi) \sim \mathrm{Poi}(\xi) \uparrow_{lr}$ with ξ $(\xi > 0)$;
- $X(\xi) \sim \mathrm{Bin}(\xi, \alpha) \uparrow_{lr}$ with ξ $(\xi \in \mathbb{N}$, here α is held fixed in $(0, 1))$;
- $X(\xi) \sim \mathrm{Bin}(n, \xi) \uparrow_{lr}$ with ξ (n fixed in $\mathbb{N}$, $\xi \in (0, 1))$.

After these preliminary notions we can state that $X_t \equiv X_t(\lambda, \alpha) \sim \mathrm{Poi}\left(\frac{\lambda}{1-\alpha}\right)$ has the PF_2 property and

$$X_t \equiv X_t(\lambda, \alpha) \uparrow_{lr} \quad \text{with } \lambda,\ \alpha\ . \tag{2.3}$$

But what can be said about the Poisson INAR(1) process $\{X_t \equiv X_t(\lambda, \alpha) : t \in \mathbb{Z}\}$?

- *Is the PF_2 (resp. TP_2) property of the (resp. families of) Poisson and binomial distributions somehow inherited by a Poisson INAR(1) process (resp. a family of Poisson INAR(1) processes)?*
- *If that is the case what are the consequences?*

Proper replies to these queries are provided in this and the following sections.

Proposition 2.1. *The Poisson INAR(1) process $\{X_t : t \in \mathbb{Z}\}$ satisfies*

$$(X_t \mid X_{t-1} = i) \leq_{lr} (X_t \mid X_{t-1} = m)\ , \qquad i \leq m\ , \tag{2.4}$$

for any $t \in \mathbb{Z}$. Equivalently, $(X_t \mid X_{t-1} = i) \uparrow_{lr}$ with i, for $t \in \mathbb{Z}$, and we write

$$\{X_t : t \in \mathbb{Z}\} \in \mathcal{M}_{lr}\ , \tag{2.5}$$

where $\mathcal{M}_{lr}$ stands for the class of stochastic processes that are stochastically monotone in the likelihood ratio sense.

We defer the proof of Proposition 2.1 until a few remarks are made.

$\{X_t : t \in \mathbb{Z}\} \in \mathcal{M}_{lr}$ can be written as $\mathbf{P} \in \mathcal{M}_{lr}$, where $\mathbf{P}$ obviously denotes the TPM of this Markov chain. This feature of $\mathbf{P}$ obviously means that, if we associate a p.f. of a discrete r.v. to one of its rows, the corresponding r.v. stochastically increase in the likelihood ratio sense as we progress along the rows of this stochastic matrix. It also means that

$$\mathbf{P} \in \mathrm{TP}_2 \,, \tag{2.6}$$

as mentioned by Kijima (1997, p. 129, Definition 3.11).

Bearing in mind that the i^{th} row of $\mathbf{P}$ corresponds to the probability (row vector) of the r.v. $(X_{t+1} \mid X_t = i)$ and taking advantage of $\leq_{lr}$ ordering, we are tempted to investigate whether $\mathbf{P} \in \mathcal{M}_{lr}$ by checking if $\frac{p_{i\,j}}{p_{i+1\,j}} \downarrow_j$, over $\mathbb{N}_0$, for any fixed $i \in \mathbb{N}_0$; another possibility of proving Proposition 2.1 would be to check whether $\mathbf{P} \in \mathrm{TP}_2$.

This is not the easiest way of proving that $\mathbf{P} \in \mathrm{TP}_2$, thus the proof of Proposition 2.1 relies on a different reasoning.

Proof: Let us first note that, for $i \in \mathbb{N}_0$, $(X_t \mid X_{t-1} = i) \stackrel{st}{=} Z(i) + \epsilon_t$, where: $Z(0) \stackrel{st}{=} 0$; $Z(i) \sim \mathrm{Bin}(i, \alpha)$, $i \in \mathbb{N}$; $\epsilon_t \sim \mathrm{Poisson}(\lambda)$; $Z(i)$ and ϵ_t are independent r.v.

Now, capitalizing not only on the fact that, for $i \leq m$ $(i, m \in \mathbb{N}_0)$ and α (held fixed in the interval $(0, 1)$), $Z(i) \leq_{lr} Z(m)$, but also on the log-concave (or PF_2) character of the p.f. of the summands $Z(i)$ and the independence between $Z(n)$ and ϵ_t $(n = i, m)$, we can invoke the basic decomposition formula (Karlin, 1968, p. 17) or the closure of the stochastic order $\leq_{lr}$ under the sum of independent r.v. with log-concave densities (Shaked and Shanthikumar, 1994, p. 30) [2] to conclude that

$$Z(i) + \epsilon_t \leq_{lr} Z(m) + \epsilon_t,\ i \leq m,$$

thus proving the result. □

The stochastic ordering result in the next proposition may be thought as an extension of the notion of monotone likelihood ratio to the family of Poisson INAR(1) processes, $\{\{X_t \equiv X_t(\lambda, \alpha) : t \in \mathbb{Z}\} : (\lambda, \alpha) \in \mathbb{R}^+ \times (0, 1)\}$.

Proposition 2.2. *Let $\{X_t(\lambda, \alpha) : t \in \mathbb{Z}\}$ be a Poisson INAR(1) process such that $X_t(\lambda, \alpha) = \alpha \circ X_{t-1}(\lambda, \alpha) + \epsilon_t(\lambda)$, for $(\lambda, \alpha) \in \mathbb{R}^+ \times (0, 1)$. Then*

$$(X_t(\lambda_1, \alpha_1) \mid X_{t-1}(\lambda_1, \alpha_1) = i) \leq_{lr} (X_t(\lambda_2, \alpha_2) \mid X_{t-1}(\lambda_2, \alpha_2) = m)\,, \quad i \leq m\,, \tag{2.7}$$

for any $0 < \lambda_1 \leq \lambda_2$, $0 < \alpha_1 \leq \alpha_2 < 1$ and $t \in \mathbb{Z}$.

[2] For a slightly stronger result, please refer to Shaked and Shanthikumar (1994, p. 30, Theorem 1.C.5).

Proof: This proposition can be proved in a similar fashion to Proposition 2.1. Thus, let us consider $(X_t(\lambda,\alpha) \mid X_{t-1}(\lambda,\alpha)=i) \stackrel{st}{=} Z(i,\alpha)+\epsilon_t(\lambda)$, where: $Z(0,\alpha) \stackrel{st}{=} 0$; $Z(i,\alpha) \sim \text{Bin}(i,\alpha)$, $i \in \mathbb{N}$; $\epsilon_t(\lambda) \sim \text{Poisson}(\lambda)$; $Z(i,\alpha)$ and $\epsilon_t(\lambda)$ are independent r.v.

By taking into account the monotone likelihood ratio of the Poisson and binomial families, we can add that, for $0<\lambda_1 \leq \lambda_2$, $0<\alpha_1 \leq \alpha_2<1$: $Z(i,\alpha_1) \leq_{lr} Z(i,\alpha_2)$; $\epsilon_t(\lambda_1) \leq_{lr} \epsilon_t(\lambda_2)$.

If we add to these stochastic ordering results the PF_2 character of all the summands involved and the independence between $Z(i,\alpha_j)$ and $\epsilon_t(\lambda_j)$, for $j=1,2$, we can use once again the closure of $\leq_{lr}$ under the sum of independent r.v. with log-concave densities to assert that

$$(X_t(\lambda_1,\alpha_1) \mid X_{t-1}(\lambda_1,\alpha_1)=i) \leq_{lr} (X_t(\lambda_2,\alpha_2) \mid X_{t-1}(\lambda_2,\alpha_2)=i)\,.$$

Finally, take notice that $\{X_t(\lambda_j,\alpha_j): t\in\mathbb{Z}\} \in \mathcal{M}_{lr}$, for $j=1,2$, as a consequence, $(X_t(\lambda_2,\alpha_2) \mid X_{t-1}(\lambda_2,\alpha_2)=i) \leq_{lr} (X_t(\lambda_2,\alpha_2) \mid X_{t-1}(\lambda_2,\alpha_2)=m)$, for $i \leq m$, and

$$(X_t(\lambda_1,\alpha_1) \mid X_{t-1}(\lambda_1,\alpha_1)=i) \leq_{lr} (X_t(\lambda_2,\alpha_2) \mid X_{t-1}(\lambda_2,\alpha_2)=m)\,, \quad i \leq m\,.$$

This ends the proof. □

Corollary 2.1. *Let $P(n,\lambda,\alpha) \equiv P[X_t(\lambda,\alpha)=n \mid X_{t-1}(\lambda,\alpha)=i]$. Then $P(n,\lambda,\alpha)$ is TP_2 both as a function of n and λ (with α held fixed) and as a function of n and α (for fixed λ).*

As for the implications of propositions 2.1 and 2.2 — in particular on what the random time the Poisson INAR(1) process needs to exceed a critical level x is concerned — we refer the reader to the next sections.

3. VITAL PROPERTIES OF THE HITTING TIMES FOR POISSON INAR(1) PROCESSES

Hitting times (HT) arise naturally in level-crossing problems in several areas:

- in reliability theory, HT of appropriate stochastic processes often represent the time to failure of a device subjected to shocks (and wear), which fails when its damage level crosses a critical value (Li and Shaked, 1995);

- in queueing systems, the identity of the first customer whose waiting time exceeds a critical threshold is a HT and a relevant performance measure (Greenberg, 1997);
- HT become also apparent while dealing with the problem of the first detection of words in random sequences of letters from a finite alphabet (De Santis and Spizzichino, 2014).

Considering the applications of Poisson INAR(1) processes, studying the HT of these stochastic processes is surely of vital importance.

More than on the distribution of HT, in this section we are interested in assessing the ageing properties of the HT and the impact of an increase in

- the critical level,
- the initial state, and
- the parameters λ and α

on the associated HT. Needless to say that dealing with a stochastic process with a TP_2 TPM will play a major role in the derivation of all the results.

The conditions under which HT possess specific ageing properties have been extensively studied by many authors (see e.g.: Brown and Chaganty, 1983; Assaf *et al.*, 1985; Karasu and Özekici, 1989), and rigorously reported by Li and Shaked (1997). Furthermore, these conditions are closely related to the stochastic monotonicity character of the underlying process, as noted by Li and Shaked (1995).

The next result can be translated as follows in our specific setting: the PF_2 property of the Poisson and binomial distributions is shared with a particular HT. It is a consequence of an important result that can be traced back to Karlin (1964).

Proposition 3.1. *Let:* $\{X_t : t \in \mathbb{N}_0\}$ *be a Poisson INAR(1) process with initial state* $X_0 = 0$*;* $HT^0 = \min\{t \in \mathbb{N} : X_t > x \mid X_0 = 0\}$ *be the random number of transitions needed to exceed the critical level* x $(x \in \mathbb{N}_0)$ *starting from the initial state* 0*. Then* $HT^0 \in PF_2$*.*

Proof: This proposition follows from Theorem 3.1 by Assaf *et al.* (1985), who pointed out that their result was essentially proved by Karlin (1964, pp. 93–94). □

Let us remind the reader that, since $HT^0 \in \mathrm{PF}_2$, $\{X_t : t \in \mathbb{N}_0\}$ is said to be a PF_2 process (Shaked and Li, 1997, p. 12). We ought to also mention that Proposition 3.1 can be referred to as the *PF_2 Theorem* (Shaked and Li, 1997, p. 12) for the Poisson INAR(1) process.

The next result translates the stochastic impact of an increase of the critical value x.

Proposition 3.2. *Let:* $\{X_t : t \in \mathbb{N}\}$ *be a Poisson INAR(1) process, where* $X_1 \sim$ *Poisson* $\left(\frac{\lambda}{1-\alpha}\right)$*;* $HT_x = \min\{t \in \mathbb{N} : X_t > x\}$ *be the random time at which the process exceeds the critical level* x $(x \in \mathbb{N}_0)$*. Then* $HT_x \uparrow_{lr}$ *with* x*.*

Proof: Since X_t can be written as a sum of r.v. with PF_2 p.f.,

$$X_t = \sum_{j=0}^{t-1} \alpha^j \circ \epsilon_{t-j} + \alpha^t \circ X_1 \ , \qquad t \in \mathbb{N}\backslash\{1\} \ ,$$

we can apply Theorem 2 from Karlin and Proschan (1960) and conclude that the p.f. of HT_x,

$$P(n,x) \equiv P(HT_x = n) = P(X_n > x;\, X_j \leq x,\, j = 1,2,...,n-1) \ , \qquad n \in \mathbb{N} \ ,$$

as a function of n and x, is TP_2. Consequently, $HT_x \uparrow_{lr}$ with x. □

The next proposition shows how the TP_2 character of the Poisson INAR(1) process is crucial to guarantee a specific stochastic decrease of the HT with respect to the initial value of this process.

Proposition 3.3. *Let:* $\{X_t : t \in \mathbb{N}_0\}$ *be a Poisson INAR(1) process with initial state* $X_0 = i$ $(i \in \mathbb{N}_0)$*;* $HT^i = \min\{t \in \mathbb{N} : X_t > x \mid X_0 = i\}$ *be the random number of transitions needed to exceed the critical level* x $(x \in \mathbb{N}_0)$ *starting from the initial state* i*. Then* $HT^i \downarrow_{lr}$ *with* i*.*

Proof: Since $\mathbf{P} \in \mathrm{TP}_2$, we are allowed to invoke Theorem 2.1 from Karlin (1964, pp. 42–43) and assert that the p.f. of HT^i,

$$\begin{aligned} P(n,i) &\equiv P(HT^i = n) \\ &= P(X_n > x; X_j \leq x,\, j = 1,2,...,n-1) \ , \qquad n \in \mathbb{N} \ , \end{aligned}$$

as a function of n and i, is sign reverse rule of order 2 (RR_2), i.e.,

$$P(n,i) \times P(n',i') \leq P(n',i) \times P(n,i') \ , \qquad n \leq n', \ \ i \leq i' \ .$$

This inequality is equivalent to

$$\frac{P(HT^i = n)}{P(HT^{i'} = n)} \leq \frac{P(HT^i = n')}{P(HT^{i'} = n')} \ , \qquad n < n', \ \ i < i' \ ,$$

that is, $HT^i \geq_{lr} HT^{i'}$, for $i \leq i'$. □

So far we were not able to prove the following conjecture regarding a stochastic implication of an increase in parameter λ.

- Let $\{X_t(\lambda,\alpha) : t \in \mathbb{N}_0\}$ be a Poisson INAR(1) process with initial state $X_0(\lambda,\alpha) = 0$ and $HT^0(\lambda,\alpha) = \min\{t \in \mathbb{N} : X_t(\lambda,\alpha) > x \mid X_0(\lambda,\alpha) = 0\}$. Then $HT^0(\lambda,\alpha) \downarrow_{lr}$ with λ.

Morais (2002, p. 47) discusses thoroughly the problems that arise when we try to prove results such as the one stated in previous conjecture while dealing with hitting times for discrete-time Markov chains arising in quality control. As a consequence we have to content ourselves with further — yet weaker — stochastic ordering results; they are stated in Section 5 and are particularly relevant in the performance analysis of a quality control chart, described in Section 4 and meant to detect changes in the mean of a Poisson INAR(1) process.

4. CONTROLLING THE MEAN OF A POISSON INAR(1) PROCESS

Although quality has long been considered absolutely relevant, we have to leap to the beginning of the 20th century to meet the founder of Statistical Process Control (SPC) (Ramos, 2013, p. 2). When Walter A. Shewhart joined the Western Electric Company, industrial quality exclusively relied on the inspection of end products and the removal of defective items; however, this physicist, engineer and statistician soon realized that it was important to control not only the finished product but also the process responsible for its production (ASQ, n.d.).

Shewhart essentially suggested that we should monitor a (production) process by:

- choosing a measurable characteristic, say X, of this process;
- selecting a relevant parameter;
- collecting data on a regular basis;
- plotting the observed value of a control statistic against time and comparing it with appropriate control limits;
- triggering a signal if the observed value of the statistic is beyond these control limits.

The resulting graphic device is called a quality control chart, undoubtedly one of the most important tools of SPC.

Control charts are used with the purpose of establishing whether the process is operating within its limits of expected variation (Nelson, 1982, p. 176), and to

detect changes in process parameters which may indicate a deterioration in quality. The control chart should be set in such way that a change in the parameter is detected as fast as possible without triggering false alarms too frequently.

The detection of changes in the mean of an i.i.d. process of Poisson counts can be done by making use of quality control charts such as the c-chart pioneered by Shewhart (Montgomery, 2009, p. 309), the CUSUM chart (Brook and Evans, 1972; Gan, 1993) or the EWMA chart (Gan, 1990). However, autocorrelation often arises, severely changing the performance of all quality control charts relying on the assumption that the observations refer to i.i.d. r.v., hence the use of the charts such like the ones proposed by Weiss (2009, Chap. 20).

Throughout the remainder of this paper, we assume that:

- the target value of the process mean is $\frac{\lambda_0}{1-\alpha_0}$;
- the purpose of using an upper one-sided control chart is to detect an aggravation in the mean number of defects, from its target value $\frac{\lambda_0}{1-\alpha_0}$ to $\frac{\lambda}{1-\alpha}$, due to a change either from λ_0 to λ or from α_0 to α.

Consequently, we proceed to describe the upper one-sided version of the c-control chart found in Weiss (2007) and Weiss (2009d, p. 419).

Definition 4.1. Let $\{X_t \equiv X_t(\lambda, \alpha) : t \in \mathbb{N}_0\}$ be a Poisson INAR(1) process, where denotes the number of defects in sample t, for $t \in \mathbb{N}$, given that the process mean is at level $\frac{\lambda}{1-\alpha}$. Then $x_t \equiv x_t(\lambda, \alpha)$ is the observed value of the control statistic of the upper one-sided c-chart for the mean of this process and this chart triggers a signal at time t ($t \in \mathbb{N}$) if

$$x_t > UCL = \frac{\lambda_0}{1-\alpha_0} + k \times \sqrt{\frac{\lambda_0}{1-\alpha_0}}\,, \tag{4.1}$$

where k is a positive constant chosen in such way that increases in the process mean $\frac{\lambda}{1-\alpha}$ are detected as quickly as possible and false alarms are rather unfrequent.

Since points lying above the upper control limit (UCL) indicate a potential increase in the process mean that should be investigated and eliminated, the performance of this control chart is unsurprisingly assessed by making use of the run length (RL), the random number of samples collected before a signal (either false or a valid alarm) is triggered by the chart. Hence the RL coincides with the following HT for the Poisson INAR(1) process

$$HT(\lambda, \alpha) = \min\{t \in \mathbb{N} : X_t(\lambda, \alpha) > x\}\,, \tag{4.2}$$

where $x = \lfloor UCL \rfloor$ is the integer part of the upper control limit defined in (4.1).

In the next section we shall start by addressing a few weaker ageing notions with more tangible interpretations/implications than the PF_2 property of hitting times such as the RL of the upper one-sided c-chart for the mean of a Poisson INAR(1) process.

5. OTHER PROPERTIES OF THE HITTING TIMES FOR THE POISSON INAR(1) PROCESS

Let us start this section by reminding the reader of the ageing notions of increasing failure rate (IFR), new better than used (NBU) and new better than used in expectation (NBUE).

Definition 5.1. The nonnegative integer valued r.v. Y is said to be:

- increasing failure rate — $Y \in \mathrm{IFR}$ — if $h_Y(m) = \frac{P(Y=m)}{P(Y\geq m)} \uparrow_{m\in\mathbb{N}}$;
- new better than used — $Y \in \mathrm{NBU}$ — if $P(Y>j) \geq P(Y-m>j \mid Y>m)$, $m, j \in \mathbb{N}_0$;
- new better than used in expectation — $Y \in \mathrm{NBUE}$ — if $E(Y) \geq E(Y-m \mid Y>m),\ m \in \mathbb{N}_0$.

Please note that $Y \in \mathrm{PF}_2 \Longrightarrow Y \in \mathrm{IFR} \Longrightarrow Y \in \mathrm{NBU}$ (Kijima, 1997, p. 118), and, clearly, $Y \in \mathrm{NBU} \Longrightarrow Y \in \mathrm{NBUE}$.

By capitalizing on the TP_2 character of the TPM of the Poisson INAR(1) process and on the fact that $Y \in \mathrm{PF}_2 \Longrightarrow Y \in \mathrm{IFR}$, we can immediately conclude that the RL of the upper one-sided c-chart starting with a zero value,

$$HT^0(\lambda,\alpha) \equiv \min\{t \in \mathbb{N} : X_t(\lambda,\alpha) > x \mid X_0(\lambda,\alpha) = 0\}\,, \tag{5.1}$$

has PF_2 character and therefore

$$HT^0(\lambda,\alpha) \in \mathrm{IFR}\,, \tag{5.2}$$

as illustrated by Example 5.1.

Note, however, that, according to the *IFR Theorem* (Shaked and Li, 1997, p. 12), this property is ensured by a weaker condition than $\mathbf{P} \in \mathrm{TP}_2$. In fact, if we let $\mathbf{Q} = [q_{ij}]_{i,j} \equiv [\sum_{k\leq j} p_{ik}]_{i,j}$ denote the matrix of left partial sums of $\mathbf{P}$ then $\mathbf{Q} \in \mathrm{TP}_2$ would have been sufficient to have $HT^0(\lambda,\alpha) \in \mathrm{IFR}$.

Example 5.1. Assume the number of defects in the t^{th} random sample of fixed size (say n) is modelled by a Poisson INAR(1) process $\{X_t \equiv X_t(\lambda,\alpha) : t \in \mathbb{N}_0\}$.

Consider that the detection of increases in the expected value of $X_t(\lambda, \alpha)$, $\mu = \frac{\lambda}{1-\alpha}$, is done by means of the upper one-sided c-chart Poisson chart in Definition 4.1.

The performance of this chart is measured via the HT

$$HT^i(\lambda, \alpha) \equiv \min\{t \in \mathbb{N} : X_t(\lambda, \alpha) > x \mid X_0(\lambda, \alpha) = i\} , \tag{5.3}$$

where i $(i = 0, 1, ..., x)$ is the fixed value assigned to $X_0(\lambda, \alpha)$ by the quality practitioner. If $i = 0$ (resp. $i > 0$) no head start (resp. a head start) has been given to the chart.

Moreover, assume the constant k in the expression of the upper control limit in (4.1) was set in such way that the average run length (ARL) when the values of λ and α are on-target, $E[HT^i(\lambda_0, \alpha_0)]$, is reasonably large, say larger than 100 samples.

It is well known that, for each x, $HT^i(\lambda, \alpha)$ has exactly the same distribution as the time to absorption of a Markov chain with state space $\{0, 1, ..., x+1\}$ and TPM represented in partitioned form,

$$\begin{bmatrix} \mathbf{Q} & (\mathbf{I} - \mathbf{Q})\,\underline{\mathbf{1}} \\ \underline{\mathbf{0}}^\top & 1 \end{bmatrix} , \tag{5.4}$$

where:

- $\mathbf{Q} \equiv \mathbf{Q}(\lambda, \alpha) = [p_{ij}(\lambda, \alpha)]_{i,j=0}^{x}$;
- $\mathbf{I}$ is the identity matrix with rank $x + 1$;
- $\underline{\mathbf{1}}$ (resp. $\underline{\mathbf{0}}^\top$) is a column vector (resp. row vector) of $x + 1$ ones (resp. zeros).

The associated expected value, survival function and failure (or alarm) rate function are given by

$$E[HT^i(\lambda_0, \alpha_0)] = \underline{\mathbf{e}}_i^\top \times [\mathbf{I} - \mathbf{Q}(\lambda, \alpha)]^{-1} \times \underline{\mathbf{1}} , \tag{5.5}$$

$$\overline{F}_{HT^i(\lambda,\alpha)}(m) = \underline{\mathbf{e}}_i^\top \times [\mathbf{Q}(\lambda, \alpha)]^m \times \underline{\mathbf{1}} , \qquad m \in \mathbb{N} , \tag{5.6}$$

$$\begin{aligned} h_{HT^i(\lambda,\alpha)}(m) &= \frac{P[HT^i(\lambda, \alpha) = m]}{P[HT^i(\lambda, \alpha) \geq m]} \\ &= \frac{\overline{F}_{HT^i(\lambda,\alpha)}(m-1) - \overline{F}_{HT^i(\lambda,\alpha)}(m)}{\overline{F}_{HT^i(\lambda,\alpha)}(m-1)} , \qquad m \in \mathbb{N} \end{aligned} \tag{5.7}$$

(respectively), where $\underline{\mathbf{e}}_i$ represents the $(i+1)^{\text{th}}$ vector of the orthonormal basis of $\mathbb{R}^{x+1}$.

The failure (or alarm) rate function was proposed by Margavio *et al.* (1995) and represents the conditional probability that the critical level x has been

exceeded at time m, given that this threshold has not been crossed before. Even though the alarm rate function is defined in terms of HT probabilities, it will bring forth insights into the chart detection capability, as we progress with the sampling procedure, insights that cannot be provided by the ARL $E[HT^i(\lambda_0, \alpha_0)]$, as illustrated by Margavio *et al.* (1995) and Morais and Pacheco (2012).

The parameters $\lambda_0 = 1$, $\alpha_0 = 0.4$ and $k = 3$ yield an upper one-sided c-chart for the mean of the Poisson INAR(1) process with $x = 5$ and in-control ARL equal to $E[HT^0(\lambda_0, \alpha_0)] = 157.457$ and $E[HT^3(\lambda_0, \alpha_0)] = 153.971$.

$E[HT^i(\lambda_0, \alpha_0)]$ should be calculated for a wide range of changes in the parameters λ and α in order to assess the chart detection ability to several out-of-control conditions. For instance, an increase of 10% in λ leads to out-of-control ARL of $E[HT^0(1.1\,\lambda_0, \alpha_0)] = 104.554$ and $E[HT^3(1.1\,\lambda_0, \alpha_0)] = 101.548$, whereas an increase of the same magnitude in α yields $E[HT^0(\lambda_0, 1.1\,\alpha_0)] = 120.560$ and $E[HT^3(1.1\,\lambda_0, 1.1\,\alpha_0)] = 117.018$.

The values and graphs of the alarm rate function in Table 1 and Figure 1 give additional insights to the performance of the chart as we proceed with the sampling, and to the impact of the adoption of a head start.

Table 1: Values of: the alarm rate function $h_{HT^i(\lambda,\alpha)}(m)$, for $\lambda_0 = 1$, $\alpha_0 = 0.4$, $x = 5$, $i = 0, 2$ and several values of m; the associated ARL values.

m	$h_{HT}(m)$					
	$HT^0(\lambda_0,\alpha_0)$	$HT^3(\lambda_0,\alpha_0)$	$HT^0(1.1\,\lambda_0,\alpha_0)$	$HT^3(1.1\,\lambda_0,\alpha_0)$	$HT^0(\lambda_0,1.1\,\alpha_0)$	$HT^3(\lambda_0,1.1\,\alpha_0)$
1	0.000594	0.012317	0.000968	0.016344	0.000594	0.014636
2	0.003143	0.009673	0.004939	0.013533	0.003591	0.012668
3	0.005033	0.007672	0.007755	0.011176	0.006166	0.010237
4	0.005884	0.006891	0.008980	0.010259	0.007468	0.009171
5	0.006220	0.006598	0.009449	0.009919	0.008035	0.008733
10	0.006422	0.006424	0.009722	0.009725	0.008427	0.008435
20	0.006423	0.006423	0.009724	0.009724	0.008432	0.008432
30	0.006423	0.006423	0.009724	0.009724	0.008432	0.008432
40	0.006423	0.006423	0.009724	0.009724	0.008432	0.008432
50	0.006423	0.006423	0.009724	0.009724	0.008432	0.008432
$E(HT)$	157.457	153.971	104.554	101.548	120.560	117.018

When no head start has been adopted, the IFR character of HT means that signaling, given that no observation has previously exceeded the upper control limit, becomes more likely as we proceed with the collection of samples, as previously noted by Morais and Pacheco (2012) for other control charts for i.i.d. output, contributing to a considerable decrease of the inconvenient initial inertia of this chart in the out-of-control situation.

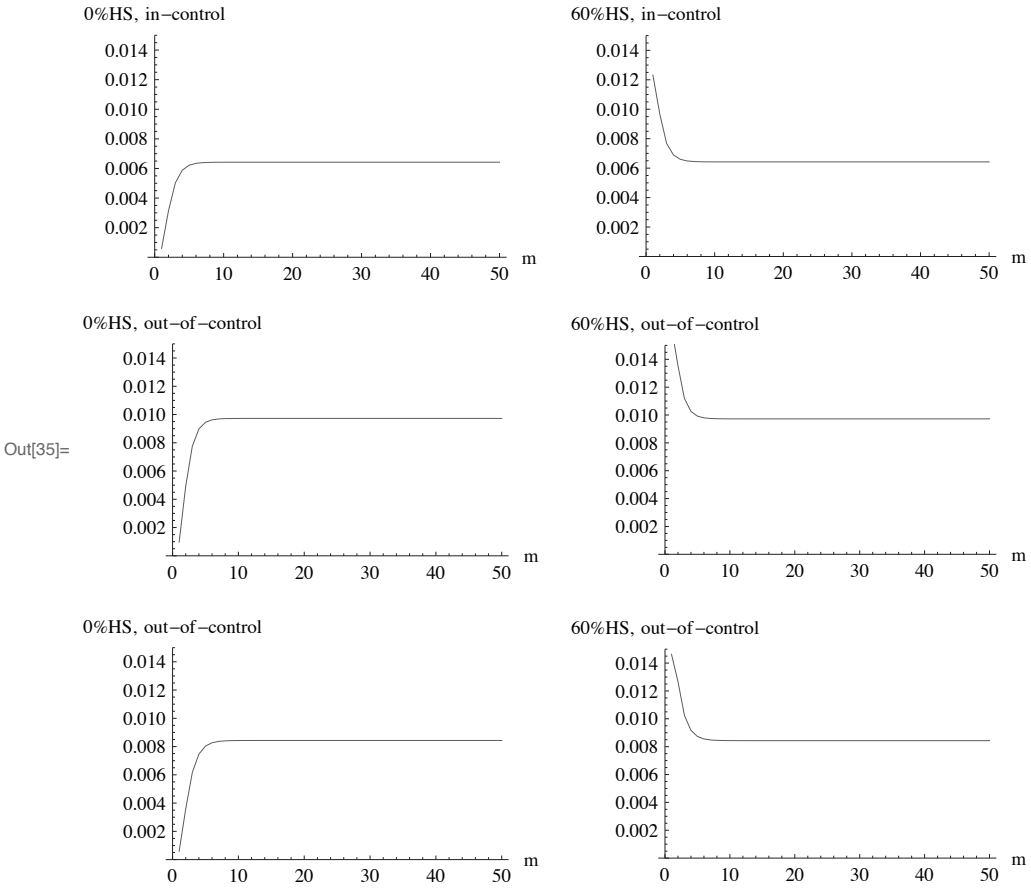

Figure 1: Alarm rates of $HT^i(\lambda, \alpha)$, for $i = 0$ (on the left) and $i = 3$ (on the right), and $(\lambda, \alpha) = (\lambda_0, \alpha_0)$, $(1.1\,\lambda_0, \alpha_0), (\lambda_0, 1.1\,\alpha_0)$ (top, center, bottom).

We ought to also note that, although the adoption of a 60% head start is responsible for mild reductions in the in-control and out-of-control ARL, adding this head-start radically changes the monotone behaviour of the alarm rate function, as shown by Figure 1: $HT^3(\lambda_0, \alpha_0)$, $HT^3(1.1\,\lambda_0, \alpha_0)$, $HT^3(\lambda_0, 1.1\,\alpha_0) \notin IFR$. Figure 1 also suggests a practical meaning of the impact of the adoption of a head start in the absence and in the presence of assignable causes: the false alarm (resp. valid signal) rate conveniently (resp. inconveniently) increases at the first samples.

We strongly believe that the results in this example show that the alarm rate function provides a more insightful portrait of the performance of the control chart than the one based on the ARL.

Even though $HT^i(\lambda, \alpha)$ may not be IFR for $i \neq 0$, it has a weaker ageing property:

$$HT^i(\lambda, \alpha) \in \text{NBU}\,, \qquad i = 0, 1, ..., x\,. \tag{5.8}$$

Brown and Chaganty (1983) devised a sufficient condition to deal with a HT with such property. However, stating this condition requires the definition of another stochastic order, a related class of stochastically monotone matrices/processes and a ordering between stochastic matrices.

Definition 5.2. Let X and Y be two nonnegative integer r.v. Then X is said to be stochastically smaller than Y in the usual sense — $X \leq_{st} Y$ — if

$$P(X > m) \leq P(Y > m)\,, \qquad m \in \mathbb{N}_0 \tag{5.9}$$

(Shaked and Shanthikumar, 1994, p. 3).

If the Markov chain $\{X_t : t \in \mathbb{Z}\}$ with TPM $\mathbf{P}$ satisfies

$$(X_t \mid X_{t-1} = i) \leq_{st} (X_t \mid X_{t-1} = m)\,, \qquad i \leq m\,, \tag{5.10}$$

for any $t \in \mathbb{Z}$, then it is said to be stochastically monotone in the usual sense (Kijima, 1995, p. 129). In this case we write $\{X_t : t \in \mathbb{Z}\} \in \mathcal{M}_{st}$ or $\mathbf{P} \in \mathcal{M}_{st}$, where $\mathcal{M}_{st}$ denotes the class of stochastic processes that are stochastically monotone in the usual sense.

Let $\mathbf{P}$ and $\mathbf{P'}$ two stochastic matrices governing two Markov chains $\{X_t : t \in \mathbb{Z}\}$ and $\{X'_t : t \in \mathbb{Z}\}$ defined in the same state space. Then $\mathbf{P}$ is said to be smaller than $\mathbf{P'}$ in the usual sense (or in the Kalmykov sense) — $\mathbf{P} \leq_{st} \mathbf{P'}$ — if

$$(X_t \mid X_{t-1} = i) \leq_{st} \left(X'_t \mid X'_{t-1} = m\right)\,, \qquad i \leq m\,. \tag{5.11}$$

Since the stochastic orders $\leq_{st}$ and $\leq_{lr}$ can be related — after all Theorem 1.C.1 of Shaked and Shanthikumar (2007, p. 42) leads to $X \leq_{lr} Y \Longrightarrow X \leq_{st} Y$ —, we naturally have $\mathbf{P} \in \mathcal{M}_{lr} \Longrightarrow \mathbf{P} \in \mathcal{M}_{st}$. Furthermore, Brown and Chaganty (1983) proved that $\mathbf{P} \in \mathcal{M}_{st}$ is sufficient to be dealing with NBU HT. Consequently, the Poisson INAR(1) process satisfies what Shaked and Li (1997, p. 13) called the *NBU Theorem*:

$$HT^i(\lambda, \alpha) \in \text{NBU}\,, \qquad i = 0, 1, ..., x\,. \tag{5.12}$$

Consequently,

$$HT^i(\lambda, \alpha) \in \text{NBUE}\,, \qquad i = 0, 1, ..., x\,. \tag{5.13}$$

By invoking Corollary 2.1 from Morais and Pacheco (2012),[3] we can add an implication of (5.13):

$$V[HT^i(\lambda, \alpha)] \leq V(Y)\,, \tag{5.14}$$

[3]This result reads as follows: if X is a discrete NBUE r.v. and Y is a geometric r.v. such that $E(X) \leq E(Y)$, then $V(X) \leq V(Y)$.

where is Y has a geometric distribution with parameter $p \leq \frac{1}{E[HT^i(\lambda,\alpha)]}$. In other words, if we hypothetically replace the upper one-sided c-chart by a chart with a geometrically distributed RL and this results in an aggravation of the ARL, then an increase in the standard deviation of the RL will also follow.

As put by Morais and Pacheco (2012), quality control practitioners should be reminded of Chebyshev's inequality and that considerable benefit it is to be gained by adopting a chart with a smaller standard deviation of the RL, thus diminishing the possibility of having observations beyond the UCL much sooner or much later than expected.

Now, we turn our attention to the HT for a Poisson INAR(1) process whose initial value is a r.v. $X_0(\lambda,\alpha) \sim \text{Poisson}\left(\frac{\lambda}{1-\alpha}\right)$. Following Weiss (2009d, p. 422), this could be called *overall RL* of the upper one-sided c-chart for the mean of such a process. This HT is a mixture of $(x+1)$ r.v. $HT^i(\lambda,\alpha)$, $i = 0, 1, ..., x$, and a zero-valued r.v. because any value of $X_0(\lambda,\alpha)$ beyond the UCL would lead to a null RL. The associated weights are $P[X_0(\lambda,\alpha) = i]$, $i = 0, 1, ..., x$, and $P[X_0(\lambda,\alpha) > x]$.

The next proposition provides a thorough characterization of this HT, represented from now on by $HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)$.

Proposition 5.1. *Let* $\{X_t(\lambda,\alpha) : t \in \mathbb{N}_0\}$ *be a Poisson INAR(1) process, where* $X_0(\lambda,\alpha) \sim$ *Poisson* $\left(\frac{\lambda}{1-\alpha}\right)$. *Then the HT*

$$HT^{X_0(\lambda,\alpha)}(\lambda,\alpha) \;=\; \min\{t \in \mathbb{N}_0 : X_t(\lambda,\alpha) > x\} \tag{5.15}$$

has expected value, survival function and failure rate function given by

$$E[HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)] \;=\; \sum_{i=0}^{x} E[HT^i(\lambda,\alpha)] \times P\left[X_0(\lambda,\alpha) = i\right], \tag{5.16}$$

$$\overline{F}_{HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)}(m) \;=\; \begin{cases} 1 - P[X_0(\lambda,\alpha) > x]\,, & m = 0\,, \\ 1 - \sum_{u=0}^{x} F_{HT^i(\lambda,\alpha)}(m) \times P\left[X_0(\lambda,\alpha) = i\right] & \\ \quad - P[X_0(\lambda,\alpha) > x]\,, & m \in \mathbb{N}\,, \end{cases} \tag{5.17}$$

$$\begin{aligned} h_{HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)}(m) \;&=\; \frac{P[HT^{X_0(\lambda,\alpha)} = m]}{P[HT^{X_0(\lambda,\alpha)} \geq m]} \\ &=\; \begin{cases} P[X_0(\lambda,\alpha) > x]\,, & m = 0\,, \\ 1 - \frac{\overline{F}_{HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)}(m)}{\overline{F}_{HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)}(m-1)}\,, & m \in \mathbb{N}_0 \end{cases} \end{aligned} \tag{5.18}$$

(respectively).

Not much can be said about the ageing properties of $HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)$ because the classes of NBU and NBUE r.v. are not closed under mixtures even though they are closed under convolutions (Barlow and Proschan, 1975/1981, pp. 104 and 187).

Table 2: Values of: the alarm rate function $h_{HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)}(m)$, for $\lambda_0 = 1$, $\alpha_0 = 0.4$, $x = 5$ and several values of m; the associated ARL values.

m	$h_{HT^{X_0}}(m)$		
	$HT^{X_0(\lambda_0,\alpha_0)}(\lambda_0,\alpha_0)$	$HT^{X_0(1.1\,\lambda_0,\alpha_0)}(1.1\,\lambda_0,\alpha_0)$	$HT^{X_0(\lambda_0,1.1\,\alpha_0)}(\lambda_0,1.1\,\alpha_0)$
0	0.007302	0.011272	0.010011
1	0.006551	0.009957	0.008677
2	0.006462	0.009795	0.008512
3	0.006437	0.009748	0.008462
4	0.006428	0.009732	0.008444
5	0.006425	0.009727	0.008437
10	0.006423	0.009724	0.008432
20	0.006423	0.009724	0.008432
30	0.006423	0.009724	0.008432
40	0.006423	0.009724	0.008432
50	0.006423	0.009724	0.008432
$E(HT^{X_0})$	154.525	101.648	70.147

Nonetheless, extensive numerical results, illustrated here by the values in Table 2 and the graphs in Figure 2, suggest that we are dealing with a HT with a decreasing failure rate.

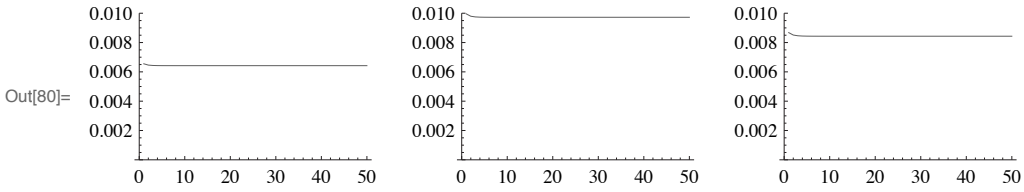

Figure 2: Alarm rates of $HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)$, for $(\lambda,\alpha) = (\lambda_0,\alpha_0), (1.1\,\lambda_0,\alpha_0), (\lambda_0,1.1\,\alpha_0)$ (left, center, right).

Finally, we qualitatively assess the impact of an increase in λ or α on $HT^{X_0(\lambda,\alpha)}(\lambda,\alpha)$ in the next proposition, stated without a proof since it follows from the fact that $\leq_{lr} \Longrightarrow \leq_{st}$ and an adaptation of Corollary 3.13 from Morais (2002, p. 46).

Proposition 5.2. *Let $\{X_t(\lambda_j,\alpha_j) : t \in \mathbb{N}_0\}$ be a Poisson INAR(1) process with initial state $X_0(\lambda_j,\alpha_j)$, for $j = 1, 2$. If $\lambda_1 \leq \lambda_2$ and $\alpha_1 \leq \alpha_2$ then*

$X_0(\lambda_1, \alpha_1) \leq_{st} X_0(\lambda_2, \alpha_2)$ *and more importantly:*

$$\mathbf{P}(\lambda_1, \alpha_1) \leq_{st} \mathbf{P}(\lambda_2, \alpha_2) \,, \tag{5.19}$$

$$HT^{X_0(\lambda_1,\alpha_1)}(\lambda_1, \alpha_1) \geq_{st} HT^{X_0(\lambda_2,\alpha_2)}(\lambda_2, \alpha_2) \,, \tag{5.20}$$

that is, $HT^{X_0(\lambda,\alpha)}(\lambda, \alpha) \downarrow_{st}$ *with* λ, α.

The stochastic ordering result (5.20) from Proposition 5.2 can be interpreted as follows: the upper one-sided c-chart for the mean of a Poisson INAR(1) process stochastically increases its detection speed as the increase in λ or α becomes more severe. This result parallels with the notion of a sequentially repeated test possessing what Ramachandran (1958) called the *monotonicity property.*

Results such as (5.20) also remind us of the notion of the level crossing ordering introduced by Irle and Gani (2001). For instance, a Markov chain $\{Y_t : t \in \mathbb{N}_0\}$ is slower in level crossing than a Markov chain $\{Z_t : t \in \mathbb{N}_0\}$ if it takes $\{Y_t : t \in \mathbb{N}_0\}$ stochastically longer than $\{Z_t : t \in \mathbb{N}_0\}$ to exceed any given level. Thus, instead of comparing two stochastic processes through all their finite dimensional distributions as for *st*-ordering, the *lc*-ordering compare two stochastic processes through their hitting times (Ferreira and Pacheco, 2007).

In light of this definition we can add that result (5.20) translates as follows: for $\lambda_1 \leq \lambda_2$ and $\alpha_1 \leq \alpha_2$, the Poisson INAR(1) process $\{X_t(\lambda_1, \alpha_1) : t \in \mathbb{N}_0\}$ is said to be slower in level-crossing in the *st*-sense than $\{X_t(\lambda_2, \alpha_2) : t \in \mathbb{N}_0\}$.

6. ON GOING AND FURTHER WORK

More than 50 years after Samuel Karlin's first and astounding contributions on total positivity, we illustrate how this concept and its implications provide insights on the performance of quality control charts for the mean of the Poisson INAR(1) process.

Directions for future work include trying to prove the conjecture $HT^0(\lambda, \alpha) \downarrow_{lr}$ with λ.

So far we can add that extensive numerical results, such as the ones shown in Figure 3, suggest this conjecture is valid. In this figure, we can find the likelihood ratio functions $\frac{P[HT^0((j+0.1)\,\lambda_0,\alpha_0)=m]}{P[HT^0(j\,\lambda_0,\alpha_0)=m]}$, for $j = 1, 1.1, 1.2, 1.3$, when $\lambda_0 = 1$, $\alpha_0 = 0.4$ and $k = 3$, as in Example 5.1. All these likelihood ratios are nonincreasing functions suggesting that

$$HT^0((j + 0.1)\,\lambda_0, \alpha_0) = m] \leq_{lr} HT^0(j\,\lambda_0, \alpha_0) \,, \quad j = 1, 1.1, 1.2, 1.3 \,. \tag{6.1}$$

Interestingly, additional numerical results led to the conclusion that $HT^0(\lambda,\alpha) \not\nearrow_{lr}$ with α.

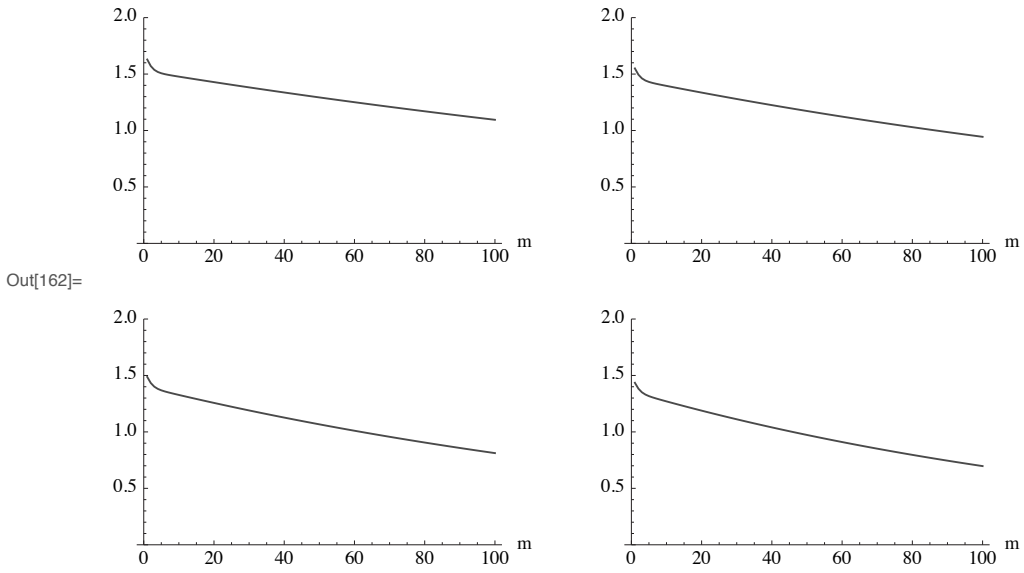

Figure 3: Likelihood ratios functions: $\frac{P[HT^0((j+0.1)\,\lambda_0,\alpha_0)=m]}{P[HT^0(j\,\lambda_0,\alpha_0)=m]}$, for $\lambda_0 = 1$, $\alpha_0 = 0.4$, $k = 3$, and $j = 1$ (top left), 1.1 (top right), 1.2 (bottom left), 1.3 (bottom right).

The simplicity of the Shewhart control charts, such as the upper one-sided c-chart we used, was responsible for their widespread popularity among quality control practitioners. The fact that Shewhart-type charts only use the last observed value of their control statistics to trigger (or not) a signal is responsible for a serious limitation: they are not effective in the detection of small and moderate shifts in the parameter being monitored. As put by Ramos (2013, p. 5), this limitation led to the cumulative sum control chart (CUSUM) proposed by Page (1954) and the exponentially weighted moving average control chart (EWMA) introduced by Roberts (1959), originally designed to monitor the process mean. CUSUM and EWMA control charts make use of recursive control statistics that account for the information contained in every collected sample of the process and prove to be more sensitive to small and moderate shifts in the process mean.

As a consequence, we also plan to conduct a similar analysis on the HT of these more sophisticated quality control charts to monitor $\frac{\lambda}{1-\alpha}$. A few difficulties may arise in the derivation of a result such as (5.20), namely because the CUSUM and EWMA control statistics constitute Markov chains with two-dimensional state spaces, as noted by Weiss and Testik (2009) and Weiss (2009c).

ACKNOWLEDGMENTS

This work was partially supported by FCT (Fundação para a Ciência e a Tecnologia) through projects PEst-OE/MAT/UI0822/2014 and PEst-OE/MAT/UI4080/2014.

REFERENCES

[1] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, **8**, 261–275.

[2] Assaf, D.; Shaked, M. and Shanthikumar, J.G. (1985). First-passage times with PF_r densities, *Journal of Applied Probability*, **22**, 185–196.

[3] ASQ (n.d.) *The Early 20th Century*, American Society for Quality, accessed from `http://asq.org/learn-about-quality/history-of-quality/overview/20th-century.html` on 2012-07-02.

[4] Barlow, R.E. and Proschan, F. (1975/1981). *Statistical Theory of Reliability and Life Testing: Probability Models* (2nd printing), Silver Spring: Holt, Rinehart and Winston, Inc.

[5] Boucher, J.-P.; Denuit, M. and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions, *Variance*, **2**, 135–162.

[6] Brook, D. and Evans, D.A. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika*, **59**, 539–549.

[7] Brown, M. and Chaganty, N.R. (1983). On the first passage time distribution for a class of Markov chains, *The Annals of Probability*, **11**, 1000–1008.

[8] Casella, G. and Berger, R.L. (2002). *Statistical Inference* (second edition), Duxbury.

[9] De Santis, E. and Spizzichino, F. (2014). Stochastic Comparisons between hitting times for Markov Chains and words' occurrences, `http://arxiv.org/abs/1210.1116v2`

[10] Ferreira, F. and Pacheco, A. (2007). Comparison of level-crossing times for Markov and semi-Markov processes, *Statistics & Probability Letters*, **77**, 151–157.

[11] Gan, F.F. (1990). Monitoring Poisson observations using modified exponentially weighted moving average control charts, *Communications in Statistics: Simulation and Computation*, **19**, 103–124.

[12] Gan, F.F. (1993). An optimal design of CUSUM control charts for binomial counts, *Journal of Applied Statistics*, **20**, 445–460.

[13] Greenberg, I. (1997). Markov chain approximation methods in a class of level-crossing problems, *Operations Research Letters*, **21**, 153–158.

[14] Irle, A. and Gani, J. (2001). The detection of words and an ordering for Markov chains, *Journal of Applied Probability*, **38A**, 66–77.

[15] Karasu, I. and Özekici, S. (1989). NBUE and NWUE properties of increasing Markov processes, *Journal of Applied Probability*, **27**, 827–834.

[16] Karlin, S. (1964). Total positivity, absorption probabilities and applications, *Transactions of the American Mathematical Society*, **11**, 33–107.

[17] Karlin, S. (1968). *Total Positivity – Vol. I*, California: Stanford University Press.

[18] Karlin, S. and Proschan, F. (1960). Pólya type distributions of convolutions, *Annals of Mathematical Statistics*, **31**, 721–736.

[19] Kijima, M. (1997). *Markov Processes for Stochastic Modelling*, London: Chapman and Hall.

[20] Kim, J.S. and Proschan, F. (1982). *Total positivity*. In "Encyclopedia of Statistical Sciences" (S. Kotz, N.L. Johnson, N. Lloyd and C.B. Read, Eds.), Volume 7, 289–297, New York: John Wiley & Sons.

[21] Li, H. and Shaked, M. (1995). On the first passage times for Markov processes with monotone convex transition kernels, *Stochastic Processes and their Applications*, **58**, 205–216.

[22] Li, H. and Shaked, M. (1997). Ageing first-passage times of Markov processes: a matrix approach, *Journal of Applied Probability*, **34**, 1–13.

[23] Margavio, T.M.; Conerly, M.D.; Woodall, W.H. and Drake, L.G. (1995). Alarm rates for quality control charts, *Statistics & Probability Letters*, **24**, 219–224.

[24] McKenzie, E. (1985). Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 645–650.

[25] Montgomery, D.C. (2009). *Introduction to Statistical Quality Control* (6th. edition), New York: John Wiley & Sons.

[26] Morais, M.J.C. (2002). *Stochastic Ordering in the Performance Analysis of Quality Control Schemes*, PhD Thesis, Technical University of Lisbon, Portugal.

[27] Morais, M.C. and Pacheco, A. (2012). A note on the aging properties of the run length of Markov-type control charts, *Sequential Analysis*, **31**, 88–98.

[28] Nelson, L.S. (1982). *Control charts*. In "Encyclopedia of Statistical Sciences" (S. Kotz and N.L. Johnson, Eds.), Volume 2, pp. 176–183, John Wiley & Sons.

[29] Page, E.S. (1954). Continuous inspection schemes, *Biometrika*, **41**, 100–115.

[30] Ramachandran, K.V. (1958). A test of variances, *Journal of the American Statistical Society*, **53**, 741–747.

[31] Ramos, P.A.A.C.F. (2013). Performance analysis of simultaneous control schemes for the process mean (vector) and (co)variance (matrix), PhD Thesis, Technical University of Lisbon, Portugal.

[32] Roberts, S.W. (1959). Control charts tests based on geometric moving averages, *Technometrics*, **1**, 239–250.

[33] Ross, S.M. (1983). *Stochastic Processes*, New York: John Wiley & Sons.

[34] Shaked, M. and Li, H. (1997). Aging first-passage times. In "Encyclopedia of Statistical Sciences: Update Volume 1" (S. Kotz, C.B. Read and D.L. Banks, Eds.), pp. 11–20, John Wiley & Sons.

[35] Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications*, London: Academic Press.

[36] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, New York: Springer-Verlag.

[37] Silva, I. (2005). *Contributions to the Analysis of Discrete-valued Time Series*, PhD Thesis, Faculdade de Ciências da Universidade do Porto, Portugal.

[38] Silva, N.; Pereira, I. and Silva, M.E. (2009). Forecasting in INAR(1) model, *REVSTAT – Statistical Journal*, **7**, 119–134.

[39] Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability*, **7**, 893–899.

[40] Weiss, C.H. (2007). Controlling correlated processes of Poisson counts, *Quality Reliability Engineering International*, **23**, 741–754.

[41] Weiss, C.H. (2009a). Monitoring correlated processes with binomial marginals, *Journal of Applied Statistics*, **36**, 399–414.

[42] Weiss, C.H. (2009b). Controlling jumps in correlated processes of Poisson counts, *Applied Stochastic Models in Business and Industry*, **25**, 551–564.

[43] Weiss, C.H. (2009c). EWMA Monitoring of correlated processes of Poisson counts, *Quality Technology and Quantitative Management*, **6**, 137–153.

[44] Weiss, C.H. (2009d). *Categorical Time Series Analysis and Applications in Statistical Quality Control*, PhD Thesis, Fakultät für Mathematik und Informatik der Universität Würzburg, dissertation.de — Verlag im Internet GmbH.

[45] Weiss, C.H. and Testik, M.C. (2009). CUSUM monitoring of first-order integer-valued autoregressive processes of Poisson counts, *Journal of Quality Technology*, **41**, 389–400.

[46] Zeger, S.L. (1988). A regression model for time series of counts, *Biometrika*, **75**, 621–629.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in February, one in April, one in June and the other in October.

— On average, five articles will be published per issue

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Editorial Board