ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal

Volume 6, No.2
June 2008

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.: +351 218 426 100
  Fax: +351 218 426 364
  Web site: http://www.ine.pt
  Customer Support Service
  (National network): 808 201 808
  (Other networks): +351 226 050 748
- **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**
  - *350 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue …… **€ 10**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)… **€ 24**
- Annual subscription (No. 2, No. 3) …… **€ 16**

INDEX

REVSTAT – Statistical Journal
Volume 6, Number 2, June 2008, 101–121

ON A CLASS OF $\mathbb{Z}_+$-VALUED AUTOREGRESSIVE MOVING AVERAGE (ARMA) PROCESSES

Authors: Emad-Eldin A. A. Aly
– Department of Statistics and Operations Research, Kuwait University, Kuwait
emad@kuc01.kuniv.edu.kw

Nadjib Bouzar
– Department of Mathematics and Computer Science, University of Indianapolis, U.S.A.
nbouzar@uindy.edu

Received: April 2007 Revised: November 2007 Accepted: November 2007

Abstract:

- A convolution semigroup of probability generating functions and its related operator $\odot_F$ are used to construct a class of stationary $\mathbb{Z}_+$-valued autoregressive moving average ($ARMA$) processes. Several distributional and regression properties are obtained. A number of $ARMA$ processes with specific innovation sequences are presented.

Key-Words:

- *stationarity; semigroup of probability generating functions; Mittag–Leffler distribution; Linnik distribution; time-reversibility.*

AMS Subject Classification:

- 60G10, 62M10.

1. INTRODUCTION

Time series models for count data have been the object of growing interest in the last twenty years. Numerous articles dealing with the theoretical aspects of these models as well as their applicability have appeared in the literature. We refer to McKenzie (2003) for an overview of the recent work in this area,

Stationary time series with a given marginal distribution have been developed by several authors. Most notably, McKenzie (1986, 1988), Al-Osh and Alzaid (1988) and Pillai and Jayakumar (1995) constructed stationary integer-valued autoregressive moving average (INARMA) processes with Poisson, negative binomial and discrete Mittag–Leffler marginal distributions. These models are based on the binomial thinning operator $\odot$ of Steutel and van Harn (1979) which is defined as follows: if X is a $\mathbb{Z}_+$-valued random variable (rv) and $\alpha \in (0,1)$, then

$$\alpha \odot X = \sum_{i=1}^{X} X_i \ , \tag{1.1}$$

where $(X_i, i \geq 1)$ is a sequence of iid Bernoulli(α) rv's independent of X. The binomial thinning operator incorporates the discrete nature of the variates and replaces the multiplication used in the definition of standard *ARMA* processes. Related models that make use of a more general operator were introduced by Aly and Bouzar (1994) and Zhu and Joe (2003). Other aspects of the analysis of INARMA processes, such as parameter estimation and the study of extremal properties, can be found in Al-Osh and Alzaid (1987), McCormick and Park (1997), Park and Oh (1997), Kim and Park (2006), and Hall and Scotto (2006).

Aly and Bouzar (2005) used a convolution semigroup of probability generating functions (pgf's) and the related operator $\odot_F$ (see definitions below) to construct a class of stationary $\mathbb{Z}_+$-valued INAR(p) processes. They developed a number of models with specific marginals which were shown to generalize several existing INAR(p) models. The aim of this paper is to use the semigroup approach to construct a family of stationary F-INMA(1), F-INMA(q), and F-INARMA$(1,q)$ processes. These processes can be seen as extensions of the classical branching processes of Galton–Watson–Bienaymé (Arthreya and Ney, 1972). We obtain various distributional and regression properties of F-INARMA$(1,q)$ processes. We establish in particular that a stationary F-INMA(1) process has the property of linear regression if and only if its marginal distribution is (discrete) F-stable. F-INARMA$(1,q)$ processes with F-stable, F-Mittag–Leffler, and compound discrete Linnik innovation sequences are studied. Examples are developed throughout the paper.

In the remainder of this section we recall some definitions and results that will be needed throughout the paper. For proofs and further details we refer to Athreya and Ney (1972, Chapter 3), van Harn *et al.* (1982) and van Harn and Steutel (1993).

$F := (F_t; t \geq 0)$ will denote a continuous composition semigroup of pgf's such that $F_t \not\equiv 1$ and $\delta_F = -\ln F_1'(1) > 0$. For any $|z| \leq 1$,

$$F_s \circ F_t(z) = F_{s+t}(z), \quad (s,t \geq 0)\,; \qquad \lim_{t \downarrow 0} F_t(z) = z\,; \qquad \lim_{t \to \infty} F_t(z) = 1\,. \tag{1.2}$$

The infinitesimal generator U of the semigroup F is defined by

$$U(z) = \lim_{t \downarrow 0} \big(F_t(z) - z\big)/t \qquad (|z| \leq 1)\ , \tag{1.3}$$

and satisfies $U(z) > 0$ for $0 \leq z < 1$. There exists a constant $a > 0$ and a distribution $(h_n, n \geq 0)$ on $\mathbb{Z}_+$ with pgf $H(z)$ such that $h_1 = 0$,

$$H'(1) = \sum_{n=1}^{\infty} n\, h_n \leq 1\ , \tag{1.4}$$

and

$$U(z) = a\big\{H(z) - z\big\}\ , \qquad |z| \leq 1\ , \tag{1.5}$$

The related A-function is defined by

$$A(z) = \exp\left\{ -\int_0^z \big(U(x)\big)^{-1}\, dx \right\}\ , \qquad z \in [0,1]\ . \tag{1.6}$$

$A(z)$ is strictly decreasing over $[0,1]$, with $A(0) = 1$ and $A(1) = 0$. The functions $U(z)$ and $A(z)$ satisfy

$$U\big(F_t(z)\big) = U(z)\, F_t'(z) \ \text{ and } \ A\big(F_t(z)\big) = e^{-t} A(z) \qquad (t \geq 0;\ 0 \leq z \leq 1)\ . \tag{1.7}$$

Moreover,

$$\delta_F = a\big(1 - H'(1)\big) = -U'(1) \ \text{ and } \ F_t'(1) = e^{-\delta_F t} \qquad (t \geq 0)\ . \tag{1.8}$$

The function $B(z)$ defined by

$$B(z) = \lim_{t \to \infty} \frac{F_t(z) - F_t(0)}{1 - F_t(0)} \tag{1.9}$$

is a pgf such that $B(0) = 0$ and takes the form

$$B(z) = 1 - A(z)^{\delta_F}\,. \tag{1.10}$$

For a $\mathbb{Z}_+$-valued rv X and $\eta \in (0,1)$, the generalized multiplication $\eta \odot_F X$ is defined by

$$\eta \odot_F X = \sum_{i=1}^{X} Y_i\ , \tag{1.11}$$

where $(Y_i, i \geq 1)$ is a sequence of iid rv's independent of X, with common pgf F_t, $t = -\ln \eta$.

A distribution on $\mathbb{Z}_+$ with pgf $P(z)$ is said to be F-self-decomposable if for any $t > 0$, there exists a pgf $P_t(z)$ such that

$$P(z) = P\big(F_t(z)\big)\, P_t(z) \ , \qquad |z| \leq 1 \ . \tag{1.12}$$

F-self-decomposable distributions are infinitely divisible.

Throughout the paper, stationarity of a stochastic process is considered to be in the strict sense. Finally, P_X will denote the pgf of the distribution of the $\mathbb{Z}_+$-valued rv X.

2. F-INMA(1) PROCESSES

Definition 2.1. A sequence $(X_n, n \in \mathbb{Z})$ of $\mathbb{Z}_+$-valued rv's is said to be an F-INMA(1) process if for any $n \in \mathbb{Z}$,

$$X_n \ = \ \eta \odot_F \epsilon_{n-1} + \epsilon_n \ , \tag{2.1}$$

where $0 < \eta < 1$ and $(\epsilon_n, n \in \mathbb{Z})$ is a sequence of iid, $\mathbb{Z}_+$-valued rv's. $(\epsilon_n, n \in \mathbb{Z})$ is called the innovation sequence.

The generalized multiplication $\eta \odot_F \epsilon_{n-1}$ in (2.1) is performed independently for each n. More precisely, we assume the existence of an array $(Y_{i,n}, i \geq 0, n \in \mathbb{Z})$ of iid $\mathbb{Z}_+$-valued rv's, independent of $(\epsilon_n, n \in \mathbb{Z})$, such that the array's common pgf is $F_t(z)$, $t = -\ln \eta$, and

$$\eta \odot_F \epsilon_{n-1} \ = \ \sum_{i=1}^{\epsilon_{n-1}} Y_{i,n-1} \ . \tag{2.2}$$

It is clear that model (2.1) is not a Galton–Watson process. However, one can give a branching process-like interpretation as follows. In the time interval $(n-1, n]$, each element of ϵ_{n-1} brings into the system k new elements (offspring) according to the probability distribution with pgf $F_t(z)$. This gives rise to a total of $U_{n-1} = \eta \odot_F \epsilon_{n-1}$ new elements in the system by time n. The variable X_n is then obtained by superposing U_{n-1} and ϵ_n. For any n, elements of ϵ_n can only be present at time n and their offspring at time $n+1$. In other words, elements of ϵ_n and their offspring remain in the system for at most two units of time.

The transformed version of (2.1) in terms of pgf's is given by

$$P_{X_n}(z) = P_\epsilon(z)\, P_\epsilon\big[F_t(z)\big] \ , \qquad |z| \leq 1 \ , \tag{2.3}$$

where P_ϵ is the common pgf of the ϵ_n's and $t = -\ln\eta$. Furthermore, it can be easily shown that for an F-INMA(1) process $(X_n,\, n\in\mathbb{Z})$ the joint pgf Φ_k of $(X_{n+1}, X_{n+2}, ..., X_{n+k})$ for any $n\in\mathbb{Z}$ and $k\geq 2$ is

$$\Phi_k(z_1, z_2, ..., z_k) \;=\; P_\epsilon(z_k)\,\prod_{i=1}^{k} P_\epsilon\big(z_{i-1}F_t(z_i)\big)\,, \tag{2.4}$$

where $z_0 = 1,\ |z_i|\leq 1,\ i = 1, 2, ..., k,$ and $t = -\ln\eta$. It follows from (2.4) that any F-INMA(1) process is stationary.

Further distributional and correlation properties of F-INMA(1) processes are gathered in the following proposition.

Proposition 2.1. *Let $(X_n,\, n\in\mathbb{Z})$ be an F-INMA(1) process with coefficient $\eta\in(0,1)$. Assume further that the mean μ_ϵ and the variance σ_ϵ^2 of ϵ_n are finite and that $\sum_{n=2}^{\infty} n\,(n-1)\,h_n < \infty$. Then*

(**i**) $E(X_n) = \mu_\epsilon(\eta^{\delta_F}+1)$;

(**ii**) $\mathrm{Var}(X_n) = \sigma_\epsilon^2(1+\eta^{2\delta_F}) + \mu_\epsilon\left(1-\dfrac{U''(1)}{U'(1)}\right)\eta^{\delta_F}(1-\eta^{\delta_F})$;

(**iii**) *for any* $n\in\mathbb{Z}$, $\mathrm{Cov}(X_{n-1}, X_n) = \eta^{\delta_F}\,\sigma_\epsilon^2$;

(**iv**) *the autocorrelation function (ACRF) of $(X_n,\, n\in\mathbb{Z})$ at lag k is*

$$\rho(k) = \begin{cases} \eta^{\delta_F}\sigma_\epsilon^2\big/\big(\sigma_\epsilon^2(1+\eta^{2\delta_F})\big) + \mu_\epsilon\left(\left(1-\frac{U''(1)}{U'(1)}\right)\eta^{\delta_F}(1-\eta^{\delta_F})\right), & \text{if } k=1\,, \\ 0\,, & \text{if } k>1\,. \end{cases} \tag{2.5}$$

Proof: We first note that $\sum_{n=2}^{\infty} n\,(n-1)\,h_n < \infty$ implies $U''(1)$ exists (see (1.4) and (1.5)). By (1.8) and (2.2), we have $E\big(\eta\odot_F\epsilon_{n-1}\,|\,\epsilon_{n-1}\big) = \eta^{\delta_F}\,\epsilon_{n-1}$. Therefore, $E(X_n) = E(\epsilon_n) + E(\eta\odot_F\epsilon_{n-1}) = E(\epsilon_n) + \eta^{\delta_F}E(\epsilon_{n-1}) = \mu_\epsilon(\eta^{\delta_F}+1)$, and thus (**i**) holds. By differentiating twice the expression $U(F_t(z)) = F_t'(z)\,U(z)$ $(t = -\ln\eta)$ with respect to z and letting $z\to 1$, we obtain $F_t''(1) = \eta^{\delta_F}(\eta^{\delta_F}-1)\cdot$ $\cdot\, U''(1)/U'(1)$. By (1.8) and (2.2), $E\big((\eta\odot_F\epsilon_{n-1})^2\,|\,\epsilon_{n-1}\big) = \mathrm{Var}(Y_{1,n-1})\,\epsilon_{n-1} +$ $\eta^{2\delta_F}\,\epsilon_{n-1}^2$. Noting that

$$\mathrm{Var}(Y_{1,n-1}) \;=\; F_t''(1) + F_t'(1) - F_t'(1)^2 \;=\; \eta^{\delta_F}(1-\eta^{\delta_F})\left(1-\frac{U''(1)}{U'(1)}\right),$$

(**ii**) follows by direct calculations. By (2.1) and independence,

$$E(X_{n-1}X_n) = E(X_{n-1})E(\epsilon_n) + E\big(\epsilon_{n-1}(\eta\odot_F\epsilon_{n-1})\big) + E(\eta\odot_F\epsilon_{n-2})\,E(\eta\odot_F\epsilon_{n-1})\,.$$

Since $E\big(\epsilon_{n-1}(\eta\odot_F\epsilon_{n-1})\,|\,\epsilon_{n-1}\big) = \eta^{\delta_F}\,\epsilon_{n-1}^2$, again direct calculations (and (i)) yield (**iii**). (**iv**) results from (ii) and (iii) combined with the fact that X_{n-k} and X_n are independent for $k>1$. □

A stochastic process $(Z_n, n \in \mathbb{Z})$ is time reversible if for all n, $(Z_1, Z_2, ..., Z_n)$ and $(Z_n, Z_{n-1}, ..., Z_1)$ have the same distribution. By a result of McKenzie (1988), an F-INMA(1) process is time reversible if and only if for any n, (X_{n-1}, X_n) has the same distribution as (X_n, X_{n-1}). The following result gives a necessary condition for the time reversibility of an F-INMA(1) process.

Theorem 2.1. *Assume that $(X_n, n \in \mathbb{Z})$ is a time reversible F-INMA(1) process with coefficient $\eta \in (0,1)$. Then the pgf $P_\epsilon(z)$ of the marginal distribution of the innovation sequence $(\epsilon_n, n \in \mathbb{Z})$ admits the representation*

$$P_\epsilon(z) \,=\, C\, z^m \exp\left\{-\lambda \int_0^{z/F_t(0)} \frac{F_t(x) - F_t(0)}{x}\, dx\right\}, \qquad z \in [0, F_t(0)]\ , \tag{2.6}$$

where $t = -\ln \eta$, m is a nonnegative integer, $\lambda > 0$ and $0 < C < 1$ are real numbers.

Proof: Since $A(z)$ is strictly decreasing on $[0,1]$ (with $A(0) = 1$ and $A(1) = 0$), it is invertible. It follows by (1.7) that $F_t(0) = A^{-1}(e^{-t}) > 0$ for any $t > 0$. Assume $P_\epsilon(0) \neq 0$. By (2.4) (applied to $k = 2$) and the property of time reversibility, we have

$$P_\epsilon(z_2)\, P_\epsilon\big(F_t(z_1)\big)\, P_\epsilon\big(z_1 F_t(z_2)\big) \,=\, P_\epsilon(z_1)\, P_\epsilon\big(F_t(z_2)\big)\, P_\epsilon\big(z_2\, F_t(z_1)\big)\ , \tag{2.7}$$

where $t = -\ln \eta$ and $|z_i| \leq 1$, $i = 1, 2$. Setting $z_1 = 0$ and $z_2 = z$ in (2.7) yields

$$P_\epsilon(z) \,=\, \frac{P_\epsilon\big(F_t(z)\big)\, P_\epsilon\big(z\, F_t(0)\big)}{P_\epsilon\big(F_t(0)\big)}\ . \tag{2.8}$$

Moreover, differentiating with respect to z_1 in (2.7), setting $z_1 = 0$ and $z_2 = z$ in the resulting equation, and using (2.8), we obtain

$$P_\epsilon\big(z\, F_t(0)\big)\left[\frac{F_t'(0)\, P_\epsilon'\big(F_t(0)\big)}{P_\epsilon\big(F_t(0)\big)} + \frac{P_\epsilon'(0)}{P_\epsilon(0)}\big(F_t(z) - 1\big)\right] \,=\, F_t'(0)\; z\, P_\epsilon'\big(z\, F_t(0)\big)\ . \tag{2.9}$$

Setting $z = 0$ in (2.9) gives $\frac{F_t'(0)\, P_\epsilon'(F_t(0))}{P_\epsilon(F_t(0))} = \frac{P_\epsilon'(0)}{P_\epsilon(0)}\,(1 - F_t(0))$. Therefore,

$$F_t(0)\, \frac{P_\epsilon'\big(z\, F_t(0)\big)}{P_\epsilon\big(z\, F_t(0)\big)} \,=\, \lambda\, \frac{F_t(z) - F_t(0)}{z}\ , \tag{2.10}$$

where $\lambda = \frac{P_\epsilon'(0)}{P_\epsilon(0)}\, \frac{F_t(0)}{F_t'(0)}$. The solution to the differential equation (2.10) is easily seen to be

$$P_\epsilon\big(z\, F_t(0)\big) \,=\, P_\epsilon(0)\, \exp\left\{\lambda \int_0^z \frac{F_t(x) - F_t(0)}{x}\, dx\right\}, \qquad z \in [0,1]\ . \tag{2.11}$$

If $P_\epsilon(0) = 0$, then let m be the smallest positive integer such that $P_\epsilon^*(0) \neq 0$, where $P_\epsilon^*(z) = P_\epsilon(z)/z^m$. It is easily seen that $P_\epsilon^*(z)$ satisfies (2.7) and thus admits the representation (2.11). This leads to (2.6) with $C = P_\epsilon^*(0) = P_\epsilon^{(m)}(0)$. $\square$

A process $(Z_n, n \in \mathbb{Z})$ has the property of (forward) linear regression if for any $n \in \mathbb{Z}$,

$$E(Z_n \,|\, Z_{n-1}) = a + b\, Z_{n-1} \ . \tag{2.12}$$

Aly and Bouzar (2005) showed that F-INAR(1) processes do possess this property. This is not true in general for F-INMA(1) models. In fact the following result gives a characterization of those F-INMA(1) processes that possess the property of linear regression. Recall that F-stable distributions (see van Harn *et al.*, 1982) have a pgf of the form

$$P(z) = \exp\big\{-\lambda\, A(z)^{\gamma}\big\} \ , \qquad \lambda > 0, \ \ |z| \le 1 \ , \tag{2.13}$$

where γ, called the exponent of the distribution, must satisfy $0 < \gamma \le \delta_F$. F-stable distributions are F-self-decomposable (see (1.12)).

Theorem 2.2. *Assume that the distribution $(h_n, n \ge 0)$ of (1.4)–(1.5) satisfies*

$$\sum_{n=2}^{\infty} h_n\, n \ln n \ < \ \infty \ . \tag{2.14}$$

Let $(X_n, n \in \mathbb{Z})$ be an F-INMA(1) process such that $0 < P_\epsilon(0) < 1$ and $\mu_\epsilon = P'_\epsilon(1) < \infty$. Then $(X_n, n \in \mathbb{Z})$ has the property of linear regression if and only if ϵ_n has an F-stable distribution with exponent δ_F, and in this case

$$E(X_n \,|\, X_{n-1}) = \mu_\epsilon + \frac{\eta^{\delta_F}}{1+\eta^{\delta_F}}\, X_{n-1} \ . \tag{2.15}$$

Proof: Assume that (2.12) holds for some real numbers a and b. By (2.4) (for $k = 2$), the joint pgf of (X_{n-1}, X_n), $n \in \mathbb{Z}$, is

$$\Phi_2(z_1, z_2) \ = \ P_\epsilon(z_2)\, P_\epsilon\big(F_t(z_1)\big)\, P_\epsilon\big(z_1\, F_t(z_2)\big) \ , \qquad t = -\ln \eta \ .$$

Differentiating Φ_2 with respect to z_2 and then setting $z_2 = 1$ and $z_1 = z$, we obtain

$$E\big(X_n\, z^{X_{n-1}}\big) \ = \ P_\epsilon\big(F_t(z)\big)\, \Big[P'_\epsilon(1)\, P_\epsilon(z) + F'_t(1)\, z\, P'_\epsilon(z)\Big] \ , \qquad n \in \mathbb{Z} \ . \tag{2.16}$$

By (2.12), we have for any $n \in \mathbb{Z}$,

$$\begin{aligned} E(X_n\, z^{X_{n-1}}) \ &= \ E\big(z^{X_{n-1}}\, E(X_n \,|\, X_{n-1})\big) \\ &= \ b\, z\, E\big(X_{n-1}\, z^{X_{n-1}-1}\big) \ + \ a\, E(z^{X_{n-1}}) \ . \end{aligned} \tag{2.17}$$

Note that $E(X_{n-1}\, z^{X_{n-1}-1}) = P'(z)$, where $P(z)$ is the pgf of X_{n-1}. It follows by (2.17) that $E(X_n\, z^{X_{n-1}}) = a\, P(z) + b\, z\, P'(z)$ which, combined with (2.3), implies

$$\begin{aligned} E(X_n z^{X_{n-1}}) \ &= \ a\, P_\epsilon\big(F_t(z)\big)\, P_\epsilon(z) \\ &\quad + \ b\, z\, \Big[F'_t(z)\, P'_\epsilon\big(F_t(z)\big)\, P_\epsilon(z) + P'_\epsilon(z)\, P_\epsilon\big(F_t(z)\big)\Big] \ . \end{aligned} \tag{2.18}$$

Letting $Q(z) = P'_\epsilon(z)/P_\epsilon(z)$ and noting, by (1.8), $F'_t(1) = \eta^{\delta_F}$, it follows by (2.16) and (2.18) that

$$P'_\epsilon(1) + \eta^{\delta_F} z\, Q(z) \;=\; a + b z\Big[F'_t(z)\, Q\big(F_t(z)\big) + Q(z)\Big]\,. \tag{2.19}$$

Setting $z = 0$ and $z = 1$ in (2.19) (recall $Q(1) = P'_\epsilon(1) \neq 0$), we deduce that $a = P'_\epsilon(1)$ and $b = \eta^{\delta_F}/(\eta^{\delta_F}+1)$. Therefore, (2.19) reduces to

$$F'_t(z)\, Q\big(F_t(z)\big) \;=\; \eta^{\delta_F}\, Q(z)\;,$$

or, by (1.7),

$$Q(z) \;=\; \eta^{-\delta_F}\, \frac{U\big(F_t(z)\big)}{U(z)}\; Q\big(F_t(z)\big)\;.$$

The additivity property $F_t(F_{jt}(z)) = F_{(j+1)t}(z)$ and an induction argument yield for any $n \geq 1$,

$$Q(z) \;=\; e^{n\delta_F t}\, \frac{U\big(F_{nt}(z)\big)}{U(z)}\; Q\big(F_{nt}(z)\big)\;. \tag{2.20}$$

From the semigroup properties (1.2), (1.8), and (1.9) we have

$$\lim_{n\to\infty} F_{nt}(z) = 1\,, \quad \lim_{n\to\infty} \frac{U\big(F_{nt}(z)\big)}{F_{nt}(z)-1} = U'(1) = -\delta_F\,, \quad \lim_{n\to\infty} \frac{F_{nt}(z)-1}{F_{nt}(0)-1} = 1-B(z)\;.$$

Moreover, (2.14) implies (see van Harn *et al.*, 1982)

$$\lim_{n\to\infty} e^{n\delta_F t}\big(F_{nt}(0)-1\big) = -1\;.$$

By letting $n \to \infty$ in (2.20), we obtain

$$Q(z) \;=\; \frac{P'_\epsilon(z)}{P_\epsilon(z)} \;=\; \delta_F\, Q(1)\, \frac{1-B(z)}{U(z)}\;. \tag{2.21}$$

Since (by (1.6) and (1.10)) $1/U(z) = -A'(z)/A(z)$ and $1 - B(z) = A(z)^{\delta_F}$, it follows from (2.21)

$$\ln P_\epsilon(z) \;=\; -\delta_F\, Q(1) \int_1^z A'(x)\, A(z)^{\delta_F - 1}\, dx \;=\; -\delta_F\, Q(1)\, A(z)^{\delta_F}\;.$$

This proves the necessary part. To prove sufficiency, assume that $P_\epsilon(z) = \exp\big\{-\lambda\, A(z)^{\delta_F}\big\}$ for some $\lambda > 0$. Since by (1.9) $P_\epsilon(z) = \exp(B(z)-1)$, assumption (2.14), which is equivalent to $B'(1) < \infty$ (see Athreya and Ney (1972), Chapter 3, or van Harn *et al.* (1982), Remark 7.3), implies $\mu_\epsilon = E(\epsilon_n) < \infty$, and thus $E(X_n) < \infty$. We have for any $n \in \mathbb{Z}$

$$E(X_n \,|\, X_{n-1}) \;=\; E\bigg(E\Big(X_n \,|\, \epsilon_{n-1},\, \epsilon_{n-2},\, \big(Y_{i,n-1}, Y_{i,n-2},\, i \geq 1\big)\Big)\;\Big|\; X_{n-1}\bigg)$$

and, by independence and (2.2),

$$E\Big(X_n \,|\, \epsilon_{n-1},\, \epsilon_{n-2},\, \big(Y_{i,n-1}, Y_{i,n-2},\, i \geq 1\big)\Big) \;=\; \mu_\epsilon + \eta^{\delta_F}\, \epsilon_{n-1}\;.$$

Therefore,

$$E(X_n\,|\,X_{n-1}) \,=\, \mu_\epsilon + \eta^{\delta_F}\, E(\epsilon_{n-1}\,|\,X_{n-1})\ . \tag{2.22}$$

The joint pgf $g(z_1,z_2) = E(z_1^{\epsilon_n}\, z_2^{X_n})$ of (ϵ_n, X_n) is independent of n and is given by

$$g(z_1,z_2) \,=\, P_\epsilon(z_1 z_2)\, P_\epsilon\big(F_t(z_2)\big)\ . \tag{2.23}$$

By (2.3) and (1.7), the pgf $P(z)$ of X_n is

$$P(z) \,=\, \exp\Big\{-\lambda\Big(A\big(F_t(z)\big)^{\delta_F} + A(z)^{\delta_F}\Big)\Big\} \,=\, \exp\Big\{-\lambda\big(1+\eta^{\delta_F}\big)\, A(z)^{\delta_F}\Big\}\ . \tag{2.24}$$

Moreover,

$$\frac{d}{dz_1} g(z_1,z_2)\bigg|_{z_1=1,\,z_2=z} \,=\, E(\epsilon_n z^{X_n}) \,=\, \sum_{k=0}^{\infty} z^k\, E\big(\epsilon_n\,|\,X_n=k\big)\, p_k\ , \tag{2.25}$$

where $(p_k,\, k\geq 0)$ is the distribution of X_n. By (2.23),

$$\frac{d}{dz_1} g(z_1,z_2) \,=\, z_2\, P_\epsilon\big(F_t(z_2)\big)\, P'_\epsilon(z_1 z_2)\ .$$

Direct calculations, combined with (1.7) and the equation $A'(z)/A(z) = -1/U(z)$ (from (1.6)), yield

$$\frac{d}{dz_1} g(z_1,z_2)\bigg|_{z_1=1,\,z_2=z} \,=\, \lambda\,\delta_F\, \frac{z\,A(z)^{\delta_F}}{U(z)}\, \exp\Big\{-\lambda\big(1+\eta^{\delta_F}\big)\, A(z)^{\delta_F}\Big\}\ .$$

We deduce (in view of (2.24))

$$\frac{d}{dz_1} g(z_1,z_2)\bigg|_{z_1=1,\,z_2=z} \,=\, \big(1+\eta^{\delta_F}\big)^{-1}\, z\,P'(z) \,=\, \big(1+\eta^{\delta_F}\big)^{-1}\, \sum_{k=0}^{\infty} k\, p_k\, z^k\,. \tag{2.26}$$

Since $P(z)$ is infinitely divisible and $p_1 = P'(0) = \lambda\,\delta_F\, e^{-\lambda} > 0$, it follows by Corollary 8.3, p. 51, in Steutel and van Harn (2004) that $p_k > 0$ for all $k \geq 0$. Uniqueness of the power series coefficients in (2.25) and (2.26) implies that for any $n \in \mathbb{Z}$

$$E(\epsilon_n\,|\,X_n) \,=\, (1+\eta^{\delta_F})^{-1}\, X_n\ . \tag{2.27}$$

Equation (2.15) follows then from (2.22) and (2.27). □

van Harn *et al.* (1982) (see also Zhu and Joe, 2003) give some rich examples of continuous composition semigroups of pgf's from which one can generate F-INMA(1) processes. We mention the parameterized family of semigroups $\big(F^{(\theta)},\, \theta\in[0,1)\big)$ described by

$$F_t^{(\theta)}(z) \,=\, 1 - \frac{\overline{\theta}\, e^{-\overline{\theta}t}\,(1-z)}{\overline{\theta} + \theta\,(1-e^{-\overline{\theta}t})\,(1-z)}\ , \qquad t\geq 0,\ \ |z|\leq 1,\ \ \overline{\theta}=1-\theta\ . \tag{2.28}$$

In this case we have $\delta_{F^{(\theta)}} = \overline{\theta}$, $U^{(\theta)}(z) = (1-z)(1-\theta z)$ and $A^{(\theta)}(z) = \left(\frac{1-z}{1-\theta z}\right)^{\frac{1}{\theta}}$. We note that for $\theta = 0$, $F^{(\theta)}$ corresponds to the standard semigroup $F_t^{(0)}(z) = 1 - e^{-t} + e^{-t}z$ and $\odot_{F^{(0)}}$ is the binomial thinning operator of Steutel and van Harn (1979) (see (1.1)).

For the family of semigroups $\left(F^{(\theta)}, \theta \in [0,1)\right)$ of (2.28), the pgf $P_\epsilon(z)$ of (2.6) is shown to be (via analytic continuation):

$$P_\epsilon(z) = \begin{cases} z^m\, e^{-\lambda(1-z)}, & \text{if} \quad \theta = 0 \quad (\lambda > 0), \\ z^m \left(\dfrac{\overline{\theta}}{1-\theta z}\right)^r, & \text{if} \quad 0 < \theta < 1 \quad (r > 0)\ , \end{cases} \tag{2.29}$$

for some nonnegative integer m. Therefore, by Theorem 2.1, for time reversibility for an $F^{(0)}$-INMA(1) (resp. $F^{(\theta)}$-INMA(1), $0 < \theta < 1$) to hold it is necessary that $\epsilon_n \stackrel{d}{=} \epsilon + m$ where ϵ has a Poisson distribution with some mean $\lambda > 0$ (resp. a negative binomial distribution with probability of success θ). In this case the converse holds as well, as shown by Al-Osh and Alzaid (1988), for $\theta = 0$, and by Aly and Bouzar (1994), for $0 < \theta < 1$.

The family of semigroups $\left(F^{(\theta)}, \theta \in [0,1)\right)$ of (2.28) necessarily satisfies condition (2.14) (since $h_n = 0$ for $n \geq 3$). By Theorem 2.2, an $F^{(\theta)}$-INMA(1) process has the property of (forward) linear regression if and only if its innovation sequence has a Poisson geometric distribution with pgf

$$P_\epsilon(z) = \exp\left\{-\lambda\, \frac{1-z}{1-\theta z}\right\} \qquad (\lambda > 0)\ . \tag{2.30}$$

The version of Theorem 2.2 for the semigroup $F^{(\theta)}$ was established Al-Osh and Alzaid (1988) (for $\theta = 0$) and by Aly and Bouzar (1994) (for $0 < \theta < 1$).

3. F-INMA(1) PROCESSES WITH A DISCRETE STABLE INNOVATION SEQUENCE

Aly and Bouzar (2005) introduced a stationary F-INAR(1) process with an F-stable marginal. In this section, we construct its F-INMA(1) counterpart.

Let $(X_n, n \in \mathbb{Z})$ be an F-INMA(1) process such that ϵ_n has the F-stable distribution with exponent γ, $0 < \gamma \leq \delta_F$, and pgf (2.13). Then by (1.3), (2.3), and (2.13), the marginal distribution of $(X_n, n \in \mathbb{Z})$ is F-stable with the same exponent and with pgf

$$P(z) = \exp\left\{-\lambda(1+\eta^\gamma)\, A(z)^\gamma\right\}\ , \qquad \lambda > 0, \quad |z| \leq 1\ . \tag{3.1}$$

The joint pgf of $(X_1, X_2, ..., X_k)$ is (by way of (2.4) and (1.3))

$$\Phi_k(z_1, z_2, ..., z_k) = \exp\left\{-\lambda\left(\eta^\gamma A(z_1)^\gamma + \sum_{i=2}^{k} A\big(z_{i-1}\,F_t(z_i)\big)^\gamma + A(z_k)^\gamma\right)\right\}, \tag{3.2}$$

where $t = -\ln\eta$.

By van Harn *et al.* (1982), an F-stable distribution with exponent γ has a finite mean if and only if $\gamma = \delta_F$ and $B'(1) < \infty$ (or, equivalently, (2.14) holds). Therefore, a finite mean F-INMA(1) process with an F-stable marginal distribution exists only if $\gamma = \delta_F$ and $B'(1) < \infty$. In this case $\mu_\epsilon = \lambda B'(1)$. If we further assume that $B''(1) < \infty$, the variance of ϵ_n is $\sigma_\epsilon^2 = \lambda(B''(1) + B'(1))$. The mean and variance of X_n as well as the correlation coefficient of (X_n, X_{n+1}), follow from Proposition 2.1, under the further assumption $\sum_{n=2}^{\infty} n(n-1)h_n < \infty$.

The branching process-like interpretation of an F-INMA(1) process (described in Section 2) leads naturally to consider the variable $T_k = \sum_{i=1}^{k} X_i$. T_k represents the total number of elements that were present in the system during the time interval $[0, k]$. It can be easily seen that the pgf of T_k is $P_{T_k}(z) = \Phi(z, z, ..., z)$ (see (2.4)), or

$$P_{T_k}(z) = \exp\left\{-\lambda\Big[(1+\eta^\gamma)\big(A(z)\big)^\gamma + (k-1)\,A\big(z\,F_t(z)\big)^\gamma\Big]\right\}, \quad t = -\ln\eta\,. \tag{3.3}$$

It is easily shown from (3.3) that $T_k \stackrel{d}{=} Y_1 + Z_1$, where Y_1 is F-stable with exponent γ, Z_1 is an F-stable compounding (with exponent γ) of the distribution with pgf $z\,F_t(z)$, and Y_1 and Z_1 are independent.

Considering the family of semigroups $\big(F^{(\theta)}, \theta \in [0,1)\big)$ of (2.28), we note that the Poisson INMA(1) process of McKenzie (1988) is the finite mean $F^{(0)}$-INMA(1) process with an $F^{(0)}$-stable marginal. The Poisson geometric INMA(1) process of Aly and Bouzar (1994) (with pgf (2.30)) arises as the finite mean $F^{(\theta)}$-INMA(1) process with an $F^{(\theta)}$-stable marginal.

4. F-INMA(1) PROCESSES WITH A DISCRETE MITTAG–LEFFLER INNOVATION SEQUENCE

A distribution on $\mathbb{Z}_+$ is said to have an F-Mittag–Leffler (or F-ML) distribution with exponent γ, $0 < \gamma \le \delta_F$, if its pgf is of the form

$$P(z) = \big(1 + c\,A(z)^\gamma\big)^{-1} \qquad \text{for some} \quad c > 0\,. \tag{4.1}$$

F-ML distributions are F-self-decomposable (van Harn and Steutel, 1993). Aly and Bouzar (2005) presented a stationary F-INAR(1) process with an F-ML marginal.

If $(X_n,\, n\in\mathbb{Z})$ is an F-INMA$\,(1)$ process such that ϵ_n has the F-ML distribution of (4.1), then X_n admits the following representation:

$$X_n \stackrel{d}{=} \sum_{i=1}^{Y+Z} W_i\ , \tag{4.2}$$

where the W_i's are iid $\mathbb{Z}_+$-valued rv's with common pgf $B(z)$ of (1.9)–(1.10), and Y and Z are independent $\mathbb{Z}_+$-valued rv's (also independent of the W_i's) and with respective pgf's

$$P_Y(z) = \Big(1 + c(1-z)^{\gamma/\delta_F}\Big)^{-1} \quad \text{and} \quad P_Z(z) = \Big(1 + c\,\eta^\gamma(1-z)^{\gamma/\delta_F}\Big)^{-1}. \tag{4.3}$$

This is shown as follows. Let $P(z)$ be the pgf of $\sum_{i=1}^{Y+Z} W_i$. By (1.10) and (4.3),

$$P(z) \,=\, P_{Y+Z}\big(B(z)\big) \,=\, P_Y\big(B(z)\big)\, P_Z\big(B(z)\big) \,=\, \big(1+c\,A(z)^\gamma\big)^{-1}\big(1+c\,\eta^\gamma A(z)^\gamma\big)^{-1},$$

or, $P(z) = P_\epsilon(z)\, P_\epsilon(F_t(z))$, $t = -\ln\eta$. The representation (4.2) follows then from (2.3).

The joint pgf of $(X_1, X_2, ..., X_k)$ is (by way of (2.4) and (1.3))

$$\begin{aligned}\Phi_k(z_1,...,z_k) \;&=\; \big(1 + c\,\eta^\gamma A(z_1)^\gamma\big)^{-1} \\ &\quad\times \left[\prod_{i=2}^{k}\Big(1 + c\,A\big(z_{i-1}F_t(z_i)\big)^\gamma\Big)^{-1}\right]\big(1 + c\,\eta^\gamma A(z_k)^\gamma\big)^{-1},\end{aligned} \tag{4.4}$$

where $t = -\ln\eta$.

Similarly to the discrete stable case of Section 3, an F-ML distribution with exponent γ has a finite mean if and only if $\gamma = \delta_F$ and $B'(1) < \infty$ (or, equivalently, (2.14) holds). Therefore, a finite mean F-INMA$\,(1)$ process with an F-ML innovation exists only if $\gamma = \delta_F$ and $B'(1) < \infty$. In this case $\mu_\epsilon = c\,B'(1)$. If we further assume that $B''(1) < \infty$, the variance of ϵ_n is $\sigma_\epsilon^2 = c\big(B''(1) + c\,B'(1)^2 + B'(1)\big)$. The mean and variance of X_n as well as the correlation coefficient of (X_n, X_{n+1}) follow from Proposition 2.1, under the further assumption $\sum_{n=2}^{\infty} n\,(n-1)\,h_n < \infty$. We note that when $\gamma = \delta_F$, the distributions of the rv's Y and Z of (4.2) and (4.3) simplify respectively to a Geometric $\big(\frac{c}{1+c}\big)$ and a Geometric $\big(\frac{c\,\eta^\gamma}{1+c\,\eta^\gamma}\big)$.

The total number of elements, $T_k = \sum_{i=1}^{k} X_i$, that were present in the system during the time interval $[0, k]$ for an F-INMA$\,(1)$ process with an F-ML marginal has pgf

$$P_{T_k}(z) = \Big[\big(1+c\,A(z)^\gamma\big)\big(1+c\,\eta^\gamma A(z)^\gamma\big)\Big]^{-1}\Big(1+c\,A\big(z\,F_t(z)\big)^\gamma\Big)^{1-k}, \quad t = -\ln\eta\,. \tag{4.5}$$

By (4.5), T_k admits the representation $T_k \stackrel{d}{=} Y_2 + W_2 + Z_2$, where Y_2, W_2 and Z_2 are independent, Y_2 and W_2 have F-ML distributions with exponent γ, and Z_2 is

a compounding of the distribution with pgf $zF_t(z)$ by the $(k-1)$-th convolution of the distribution of ϵ_n.

Following McKenzie (1986), an F-INMA(1) process with an F-ML marginal distribution can be obtained by modifying (2.1) as follows:

$$X_n = \eta \odot \epsilon_n + B_n \epsilon_{n-1} \ , \tag{4.6}$$

where $0<\eta<1$, $(\epsilon_n, n\in\mathbb{Z})$ is a sequence of iid rv's with a common F-ML distribution with exponent $0<\gamma\leq\delta_F$, $(B_n, n\in\mathbb{Z})$ is a sequence of iid Bernoulli$(1-\eta^\gamma)$ rv's, and $(\epsilon_n, n\in\mathbb{Z})$ and $(B_n, n\in\mathbb{Z})$ are independent. By (4.6), (4.1), and (1.7), the pgf $P(z)$ of X_n is shown to be

$$P(z) \ = \ \big(1+c\,\eta^\gamma A(z)^\gamma\big)^{-1}\Big(\eta^\gamma+(1-\eta^\gamma)\big(1+c\,A(z)^\gamma\big)^{-1}\Big) \ = \ \big(1+c\,A(z)^\gamma\big)^{-1}.$$

The finite mean $F^{(0)}$-ML innovation sequence corresponding to the $F^{(0)}$-INMA(1) process of (2.1) (with $F^{(0)}$ as in (2.28)) reduces to a geometric innovation with probability of success $1/(1+c)$. Likewise, for the semigroup $F^{(\theta)}$, $0<\theta<1$, of (2.28), the finite mean $F^{(\theta)}$-ML innovation process for an $F^{(\theta)}$-INMA(1) process admits the representation $\epsilon_n \stackrel{d}{=} I_n\,\epsilon'_n$ where $(I_n, n\in\mathbb{Z})$ and $(\epsilon'_n, n\in\mathbb{Z})$ are independent sequences of $\mathbb{Z}_+$-valued iid rv's, I_n is Bernoulli$(c/(1+c))$, and ϵ'_n has a (truncated at zero) geometric distribution with probability of success $\overline{\theta}/(1+c)$. Finally, the geometric INMA(1) process of McKenzie (1986) corresponds to the modified finite mean $(\gamma=1)$ F-INMA(1) process of (4.6) with an $F^{(0)}$-ML marginal distribution.

5. F-INMA(1) PROCESSES WITH A COMPOUND DISCRETE LINNIK INNOVATION SEQUENCE

A $\mathbb{Z}_+$-valued rv X is said to have an F-compound discrete Linnik distribution if its pgf has the form

$$P(z) = \big(1+\lambda\,A(z)^\gamma\big)^{-r}, \tag{5.1}$$

for some $0<\gamma\leq\delta_F$, $\lambda>0$, and $r>0$. van Harn and Steutel (1993) showed that F-compound discrete Linnik distributions are F-self-decomposable and arise as solutions to stability equations for $\mathbb{Z}_+$-valued processes with stationary independent increments. Aly and Bouzar (2005) constructed a $\mathbb{Z}_+$-valued stationary INAR(1) process with an F-compound discrete Linnik distribution. Note the case $r=1$ corresponds to the F-ML distribution of the previous section.

If $(X_n, n\in\mathbb{Z})$ is an F-INMA(1) process such that ϵ_n has the F-compound discrete Linnik distribution, then the distribution of X_n has the following repre-

sentation:

$$X_n \stackrel{d}{=} \sum_{i=1}^{Y+Z} W_i \ , \tag{5.2}$$

where the W_i's are iid with common pgf $B(z)$, and Y and Z are $\mathbb{Z}_+$-valued independent rv's (also independent of the W_i's) with respective pgf's

$$P_Y(z) = \Big(1 + c(1-z)^{-\gamma/\delta_F}\Big)^{-r} \quad \text{and} \quad P_Z(z) = \Big(1 + c\,\eta^\gamma(1-z)^{-\gamma/\delta_F}\Big)^{-r}. \tag{5.3}$$

The proof of (5.2)–(5.3) is identical to the one given in the case of the F-INMA(1) process with an F-ML innovation (see (4.2)–(4.3) of the previous section). The details are omitted.

Formulas for the joint pgf of $(X_1, X_2, ..., X_k)$ as well as the pgf of $T_k = \sum_{i=1}^k X_i$ can be derived similarly to the F-ML case of the previous section.

Furthermore, a finite mean F-INMA(1) process with a compound discrete Linnik innovation sequence exists only if $\gamma = \delta_F$ and $B'(1) < \infty$. In this case $\mu_\epsilon = rc\,B'(1)$. If we further assume then $B''(1) < \infty$, the variance of ϵ_n is $\sigma_\epsilon^2 = rc\big(B''(1) + cB'(1)^2 + B'(1)\big)$. The mean and variance of X_n as well as the correlation coefficient of (X_n, X_{n+1}) follow from Proposition 2.1, under the further assumption $\sum_{n=2}^\infty n\,(n-1)\,h_n < \infty$. We note that when $\gamma = \delta_F$, the distributions of the rv's Y and Z of (5.2) simplify respectively to a negative binomial $\big(\frac{c}{1+c},\ r\big)$ and a negative binomial $\big(\frac{c\eta^\gamma}{1+c\eta^\gamma},\ r\big)$.

6. F-INMA(q) PROCESSES

Definition 6.1. A sequence $(X_n, n \in \mathbb{Z})$ of $\mathbb{Z}_+$-valued rv's is said to be an F-INMA(q) process if for any $n \in \mathbb{Z}$,

$$X_n \ = \ \epsilon_n + \sum_{i=1}^q \eta_i \odot_F \epsilon_{n-i} \ , \tag{6.1}$$

where $(\epsilon_n, n \in \mathbb{Z})$ is a sequence of iid, $\mathbb{Z}_+$-valued rv's, and $0 < \eta_i < 1$, $i = 1, 2, ..., q$.

The generalized multiplications $\eta_i \odot_F \epsilon_{n-i}$, $i = 1, ..., q$, in (6.1) are performed independently. More precisely, we assume the existence of q independent arrays $\big(Y_{j,n}^{(i)},\ j \geq 0,\ n \in \mathbb{Z}\big)$, $i = 1, 2, ..., q$, of iid $\mathbb{Z}_+$-valued rv's, independent of $(\epsilon_n, n \in \mathbb{Z})$, such that for each $i = 1, 2, ..., q$, the array's common pgf is $F_{t_i}(z)$, $t_i = -\log \eta_i$, and

$$\eta_i \odot_F \epsilon_{n-i} \stackrel{d}{=} \sum_{j=1}^{\epsilon_{n-i}} Y_{j,n-j}^{(i)} \ . \tag{6.2}$$

Equation (6.2) can be interpreted as follows. Subsequent to time n, each element of ϵ_n has q nonoverlapping reproduction periods: $(n+i,\, n+i+1]$, $i = 0, 1, 2, ..., q-1$ with the distribution of offspring having pgf $F_{t_i}(z)$ over $(n+i,\, n+i+1]$. Each $\eta_i \odot_F \epsilon_n$ represents then the total number of offspring brought into the system by all ϵ_n elements. The offspring survive one unit of time and are replaced at time $n+1$ by the offspring from the next reproduction period. The offspring of ϵ_n are phased out of the system after q units of time. It is important to note that for all $n \in \mathbb{Z}$ and all $i, j = 1..., q$, $i \neq j$, $\eta_i \odot \epsilon_n$ and $\eta_j \odot \epsilon_n$ are independent, given ϵ_n.

The process $(X_n, n \in \mathbb{Z})$ of (6.1) is necessarily stationary and its marginal distribution has pgf

$$P_X(z) \,=\, P_\epsilon(z) \prod_{i=1}^{q} P_\epsilon\big[F_{t_i}(z)\big]\,, \tag{6.3}$$

where $t_i = -\log \eta_i,\ i = 1, ..., q$.

Distributional properties of an F-INMA(q) process are given in the following proposition. The proof is similar to the one given in the case $q = 1$ in section 2 (Proposition 2.1). The details are omitted.

Proposition 6.1. *Let $(X_n, n \in \mathbb{Z})$ be an F-INMA(q) process. Assume further that the mean μ_ϵ and the variance σ_ϵ^2 of ϵ_n are finite and that $\sum_{n=2}^{\infty} n(n-1)h_n < \infty$. Then (with $\eta_0 = 1$)*

(**1**) $E(X_n) = \mu_\epsilon \sum_{i=0}^{q} \eta_i^{\delta_F}$;

(**2**) $\mathrm{Var}(X_n) = \sigma_\epsilon^2 \left(\sum_{i=0}^{q} \eta_i^{\delta_F} \right) + \mu_\epsilon \Big(1 - U''(1)/U'(1)\Big) \left(\sum_{i=0}^{q} \eta_i^{\delta_F} \big(1 - \eta_i^{\delta_F}\big) \right)$;

(**3**) *the ACRF of $(X_n, n \in \mathbb{Z})$ at lag k is*

$$\rho(k) = \begin{cases} \left[\left(\sum_{i=0}^{q-k} \eta_i^{\delta_F} \eta_{i+k}^{\delta_F} \right) \sigma_\epsilon^2 \right] \Big/ \left[\left(\sum_{i=0}^{q} \eta_i\, \delta_F \right) \sigma_\epsilon^2 \right. \\ \qquad \left. + \,\mu_\epsilon \left(1 - \dfrac{U''(1)}{U'(1)}\right) \left(\sum_{i=0}^{q} \eta_i^{\delta_F} \big(1 - \eta_i^{\delta_F}\big) \right) \right], & 0 \le k \le q\,, \\ 0, & k > q\,. \end{cases} \tag{6.4}$$

It is clear from (6.4) that an F-INMA(q) process has the same correlation structure as the standard $MA(q)$ processes.

An F-INMA(q) process with an F-stable innovation sequence has finite mean only if $\gamma = \delta_F$ and $B'(1) < \infty$. In addition, if $B''(1) < \infty$, then ϵ_n has

finite mean and finite variance (recall, $\mu_\epsilon = \lambda B'(1)$ and $\sigma_\epsilon^2 = \lambda(B''(1) + B'(1))$). The mean and variance of X_n as well as the correlation coefficient of (X_n, X_{n+1}), follow from Proposition 6.1, under the further assumption $\sum_{n=2}^{\infty} n(n-1)h_n < \infty$.

If $(X_n, n \in \mathbb{Z})$ is an F-INMA(q) process such that ϵ_n has an F-stable distribution with pgf given by (2.13), then by (6.3) and (1.7) its marginal is also F-stable with pgf

$$P_X(z) = \exp\left\{-\lambda\left(\sum_{i=0}^{q} \eta_i^\gamma\right)\big(A(z)\big)^\gamma\right\}, \qquad \eta_0 = 1 \ . \tag{6.5}$$

We note that the Poisson INMA(q) process of McKenzie (1988) and the Poisson Geometric INMA(q) process of Aly and Bouzar (1994) are special cases of F-INMA(q) processes with a stable marginal for the semigroups $F^{(0)}$ and $F^{(\theta)}$ $(0 < \theta < 1)$ of (2.28), respectively.

If $(X_n, n \in \mathbb{Z})$ is an F-INMA(q) process such that ϵ_n has the F-ML distribution of (4.1), then X_n admits the following representation:

$$X_n \stackrel{d}{=} \sum_{i=1}^{Y+Z_1+\cdots+Z_{q-1}} W_i \ , \tag{6.6}$$

where the W_i's are iid $\mathbb{Z}_+$-valued rv's with common pgf $B(z)$ of (1.9)–(1.10), and Y and Z_i, $i = 1, ..., q-1$, are independent $\mathbb{Z}_+$-valued rv's (also independent of the W_i's) and with respective pgf's

$$\begin{aligned} P_Y(z) &= \left(1 + c(1-z)^{\gamma/\delta_F}\right)^{-1} , \\ P_{Z_i}(z) &= \left(1 + c\,\eta_i^\gamma (1-z)^{\gamma/\delta_F}\right)^{-1}, \qquad i = 1, ..., q-1 \ . \end{aligned} \tag{6.7}$$

F-INMA(q) processes with compound discrete Linnik innovation sequences can be constructed in similar fashion. The details are omitted.

We note next the existence of an F-INMA process of infinite order (F-INMA(∞)). Let $X_n, n \in \mathbb{Z}$ be a stationary F-INAR(1) process, i.e.,

$$X_n = \eta \odot_F X_{n-1} + \epsilon_n \ , \qquad n \in \mathbb{Z} \ , \tag{6.8}$$

for some innovation sequence $(\epsilon_n, n \in \mathbb{Z})$ and some $0 < \eta < 1$. Then (see Aly and Bouzar, 2005) $X_n, n \in \mathbb{Z}$ admits the following F-INMA(∞) representation:

$$X_n = \sum_{i=0}^{\infty} \eta^i \odot_F \epsilon_{n-i} \ , \qquad n \in \mathbb{Z} \ . \tag{6.9}$$

We conclude this section by mentioning that classes of F-INMA(q) processes with an autocorrelation structure different from (6.4) may result by assuming

some form of dependence between the generalized multiplications in equation (6.1). Al-Osh and Alzaid (1988) and Brännäs and Hall (2001) proposed several INMA(q) processes where dependence between the binomial thinnings in the governing equation was allowed.

7. F-INARMA$(1,q)$ PROCESSES

In this section the F-INAR(1) process of Aly and Bouzar (2005) is combined with the F-INMA(q) process of the previous section to obtain a mixed process. Let $(\epsilon_n, n\in\mathbb{Z})$ be a sequence of iid rv's and define the F-INAR(1) process $(Y_n, n\in\mathbb{Z})$ by

$$Y_n = \eta \odot_F Y_{n-1} + \epsilon_n \ , \tag{7.1}$$

The F-INARMA$(1,q)$ process is defined as

$$X_n = Y_{n-q} + \sum_{i=1}^{q} \eta_i \odot_F \epsilon_{n+1-i} \ . \tag{7.2}$$

Note that both the $AR(1)$ and the $MA(q)$ components in (7.1)–(7.2) share the same innovation sequence $(\epsilon_n, n\in\mathbb{Z})$. Moreover, the generalized multiplications $\eta_i \odot_F \epsilon_{n+1-i},\ i=1,...,q$, in (7.2) are performed independently. A representation of $(X_n, n\in\mathbb{Z})$ of (7.1)–(7.2) in terms of sequences of iid rv's can be easily obtained from the representations of its $AR(1)$ and $MA(q)$ components. The details are left out. If $(Y_n, n\in\mathbb{Z})$ is stationary (see Aly and Bouzar (2005) for sufficient conditions), then $(X_n, n\in\mathbb{Z})$ is also stationary. We will assume throughout the section that $(Y_n, n\in\mathbb{Z})$ is stationary. The joint pgf of higher order distributions of $(X_n, n\in\mathbb{Z})$ can be expressed in terms of the pgf's $P_Y(z)$ of Y_n, $P_\epsilon(z)$ of ϵ_n, $F_t(z)$ $(t=-\ln\eta)$, and $F_{t_i}(z)$ $(t_i=-\ln\eta_i,\ i=1,...,q$. For example, the joint pgf of (X_{n-1}, X_n) is shown to be

$$\phi_2(z_1,z_2) \ = \ P_Y\big[z_1 F_t(z)\big]\ P_\epsilon\big[F_{t_1}(z_2)\big]\ \prod_{i=1}^{q} P_\epsilon\big[F_{t_i}(z_1)\, F_{t_{i+1}}(z_2)\big] \ , \tag{7.3}$$

where $F_{t_{q+1}}(z_2) = z_2$.

Assume that Y_n and ϵ_n have finite means (μ_ϵ and μ, respectively) and finite variances (σ_ϵ^2 and σ^2 respectively). Assume further that $\sum_{n=2}^{\infty} n(n-1)h_n < \infty$. It can be shown that

$$\mathrm{Cov}\big(\eta \odot_F \epsilon_n,\ \eta' \odot_F \epsilon_n\big) = (\eta\,\eta')^{\delta_F}\,\sigma_\epsilon^2 \ , \qquad \eta,\eta' \in (0,1) \ , \tag{7.4}$$

$$\mathrm{Cov}(Y_{n-k}, Y_n) = \eta^{k\delta_F}\,\sigma^2 \ , \tag{7.5}$$

and

$$Y_{n-q} \stackrel{d}{=} \eta^k \odot_F Y_{n-q-k} + \sum_{i=0}^{k-1} \eta^i \odot_F \epsilon_{n-q-i} \ . \tag{7.6}$$

By (7.4)–(7.6) and the independence assumptions we obtain the $ACRF$ at lag k of $(X_n, n \in \mathbb{Z})$:

$$\rho(k) = \begin{cases} \dfrac{\eta^{k\delta_F}\,\sigma^2_{Y_0} + \eta^{(k-q)\delta_F}\left(\sum_{l=1}^{q}\left(\eta^{l-1}\,\eta_l\right)^{\delta_F}\right)\sigma^2_\epsilon}{\sigma^2_{Y_0} + \left(\sum_{k=1}^{q}\eta_k^{2\delta_F}\right)\sigma^2_\epsilon + \mu_\epsilon\left(\sum_{k=1}^{q}\eta_k^{\delta_F}\left(1-\eta_k^{\delta_F}\right)\right)\left(1-U''(1)/U'(1)\right)}\,, & k > q \\ \dfrac{\eta^{k\delta_F}\,\sigma^2_{Y_0} + \left(\sum_{l=1}^{q-k}\left(\eta^l\,\eta_l\,\eta_k\right)^{\delta_F} + \eta^{(k-q)\delta_F}\sum_{l=q-k+1}^{q}\left(\eta^{l-1}\,\eta_l\right)^{\delta_F}\right)\sigma^2_\epsilon}{\sigma^2_{Y_0} + \left(\sum_{k=1}^{q}\eta_k^{2\delta_F}\right)\sigma^2_\epsilon + \mu_\epsilon\left(\sum_{k=1}^{q}\eta_k^{\delta_F}\left(1-\eta_k^{\delta_F}\right)\right)\left(1-U''(1)/U'(1)\right)}\,, & k \le q \end{cases} \tag{7.7}$$

where $t = \ln\eta$ and $t_i = \ln\eta_i$, $i = 1, 2, ..., q$.

Let $(X_n, n \in \mathbb{Z})$ be a F-INARMA$(1, q)$ process. Assume that its F-INAR(1) component $(Y_n, n \in \mathbb{Z})$ of (7.1) is stationary with an F-stable marginal distribution with pgf (2.13). Then the innovation sequence $(\epsilon_n, n \in \mathbb{Z})$ has also an F-stable marginal distribution with pgf (see Aly and Bouzar, 2005)

$$P_\epsilon(z) = \exp\Big[-\lambda\,(1-\eta^\gamma)\,A(z)^\gamma\Big] \ .$$

It follows that the associated F-INMA(q) component in (7.2) has also an F-stable marginal with pgf

$$P_1(z) = \exp\left\{-\lambda\,(1-\eta^\gamma)\left(\sum_{i=1}^{q}\eta_i^\gamma\right)\left(A(z)\right)^\gamma\right\} .$$

Therefore, $(X_n, n \in \mathbb{Z})$ is stationary with an F-stable marginal distribution with pgf

$$P_X(z) = \exp\left\{-\lambda\left[1 + (1-\eta^\gamma)\left(\sum_{i=1}^{q}\eta_i^\gamma\right)\right]\left(A(z)\right)^\gamma\right\} . \tag{7.8}$$

If $(X_n, n \in \mathbb{Z})$ is an F-INARMA$(1, q)$ process such that its F-INAR(1) component $(Y_n, n \in \mathbb{Z})$ has an F-ML marginal (with pgf (4.1)), then the innovation sequence $(\epsilon_n, n \in \mathbb{Z})$ admits the representation (see Aly and Bouzar, 2005)

$$\epsilon_n = I_n\,E_n \ , \tag{7.9}$$

where $(I_n, n \in \mathbb{Z})$ and $(E_n, n \in \mathbb{Z})$ are independent sequences of iid rv's such that I_n is Bernoulli$(1-\eta^\gamma)$ and E_n has the same distribution as Y_n. It follows from (4.1), (7.2) and (7.9) that $(X_n, n \in \mathbb{Z})$ is stationary with marginal pgf

$$P_X(z) = \frac{1}{1+dA(z)^\gamma}\ \prod_{i=1}^{q}\left(\eta^\gamma + \frac{1-\eta^\gamma}{1+d\eta_i^\gamma A(z)^\gamma}\right) . \tag{7.10}$$

8. CONCLUSION

We have presented a class of integer-valued time series that can be used to model count data. The models introduced in this paper may be seen as extensions of the classical branching processes of Galton–Watson–Bienaymé. Various distributional and regression properties were shown to be similar to those of the standard real-valued ARMA processes. Models with specific marginals such as stable distributions and Mittag–Leffler distributions were discussed in some detail and some examples were developed.

ACKNOWLEDGMENTS

The authors are grateful to a referee for helpful comments and suggestions. The research of E.-E.A.A. Aly was done while the author was on sabbatical leave from Kuwait University.

REFERENCES

[1] Al-Osh, M.A. and Alzaid, A.A. (1988). Integer-valued moving average ($INMA$) process, *Stat. Hefte*, **29**, 281–300.

[2] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive $INAR(1)$ process, *J. Time Ser. Anal.*, **8**, 261–275.

[3] Aly, E.-E.A.A. and Bouzar, N. (1994). On some integer-valued autoregressive moving average models, *J. of Multivariate Analysis*, **50**, 132–151.

[4] Aly, E.-E.A.A. and Bouzar, N. (2005). Stationary solutions for integer-valued autoregressive processes, *International J. Math. and Math. Sci*, **2005**(1), 1–18.

[5] Athreya, K.B. and Ney, P.E. (1972). *Branching Processes*, Springer, Berlin–Heidelberg–New York.

[6] Brännäs, K. and Hall, A. (2001). Estimation in integer-valued moving average models, *Appl. Stochastic Models Bus. Ind.*, **17**, 277–291.

[7] Hall, A. and Scotto, M.G. (2006). Extremes of periodic integer-valued sequences with exponential type tails, *REVSTAT*, **4**, 249–273.

[8] Harn, K. van; Steutel, F.W. and Vervaat, W. (1982). Self-decomposable discrete distributions and branching processes, *Z. Wahrsch. Verw. Gebiete*, **61**, 97–108.

[9] Harn, K. van and Steutel, F.W. (1993). Stability equations for processes with stationary independent increments using branching processes and Poisson mixtures, *Stochastic Processes Appl.*, **45**, 209–230.

[10] Kim, Y.H. and Park, Y.S. (2006). Prediction mean squared error of the Poisson INAR(1) process with estimated parameters, *J. Korean Stat. Soc.*, **35**, 37–47.

[11] McCormick, W.P. and Park, Y.S. (1997). An Analysis of Poisson moving-average processes, *Probab. Engin. Informa. Sci.*, **11**, 487–507.

[12] McKenzie, E. (1986). Autoregressive-moving average processes with negative binomial and geometric marginal distributions, *Adv. Appl. Prob.*, **18**, 679–705.

[13] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts, *Adv. Appl. Probab.*, **20**, 822–835.

[14] McKenzie, E. (2003). *Discrete variate time series.* In "Stochastic processes: Modelling and simulation" (D.N. Shanbhag *et al.*, Ed.), Amsterdam: North-Holland, *Handb. Stat.*, **21**, 573–606.

[15] Park, Y.S. and Oh, C.W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals, *Stat. Pap.*, **38**, 287–302.

[16] Pillai, R.N. and Jayakumar, K. (1995). Discrete Mittag–Leffler distributions, *Stat. Probab. Lett.*, **23**, 271–274.

[17] Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *Ann. Probab.* **7**, 893–899.

[18] Steutel, F.W. and van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*, Marcel Dekker, Inc., New York–Basel.

[19] Zhu, R. and Joe, H. (2003). A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series, *Stoch. Models*, **19**, 235–254.

REVSTAT – Statistical Journal
Volume 6, Number 2, June 2008, 123–150

MINIMALLY BIASED NONPARAMETRIC REGRESSION AND AUTOREGRESSION

Authors: Timothy L. McMurry
– Department of Mathematical Sciences, DePaul University, Chicago, IL, USA
tmcmurry@depaul.edu

Dimitris N. Politis
– Department of Mathematics, University of California, San Diego, La Jolla, CA, USA
politis@math.ucsd.edu

Received: September 2007 Revised: November 2007 Accepted: December 2007

Abstract:

- A nonparametric regression estimator is introduced which adapts to the smoothness of the unknown function being estimated. This property allows the new estimator to automatically achieve minimal bias over a large class of locally smooth functions without changing the rate at which the variance converges. Optimal convergence rates are shown to hold for both i.i.d. data and autoregressive processes satisfying strong mixing conditions.

Key-Words:

- *nonparametric regression; autoregression; Fourier transform.*

AMS Subject Classification:

- 62G08.

1. INTRODUCTION

Suppose the data $(X_1, Y_1), ..., (X_n, Y_n)$ are observations from a general real valued bivariate random process. The simplest example is when the data are generated by a model of the form $Y_i = r(X_i) + \epsilon_i$ where the ϵ_i are mean zero random errors satisfying some conditions; in general, the ϵ_i will not be independent. A second example of interest is nonparametric autoregression, where $Y_t = X_{t+1}$. The function r, the conditional mean of Y given X, is unknown and will be estimated from the data. There are many nonparametric approaches to estimating r, including various kernel methods proposed by Nadaraya [15] and Watson [25], Gasser and Müller [7, 8], and local polynomial estimators, Fan [6]. In each of these techniques $r(x)$ is in essence estimated through weighted local averaging on the data near x. The smoothness of the function r and properties of the weights used in this averaging determine the performance of the estimator. In this paper we propose a new class of kernels which allow the Nadaraya–Watson estimator to automatically achieve asymptotically optimal performance no matter how smooth r happens to be.

The Nadaraya–Watson estimator is defined to be

$$\hat{r}(x) := \frac{\sum_{i=1}^{n} Y_i \, K\big((X_i - x)/h\big)}{\sum_{i=1}^{n} K\big((X_i - x)/h\big)} \, . \tag{1.1}$$

The function $K(x)$ is the kernel; it is used to weight the observations. The denominator ensures the weights sum to 1. The parameter h is the bandwidth, or smoothing parameter. It balances a tradeoff between bias and variance. Small values of h concentrate the mass of the kernel near x, giving heavy weight to nearby observations and relatively little or no weight to more distant observations, resulting in a relatively unbiased but highly variable estimate. By contrast, large values of h average over many data points, resulting in an estimate with relatively low variance, but potentially large bias, as observations which are quite distant from x are included in the average. Since the number of data points included in the average is proportional to nh, each of these estimators has pointwise variance proportional to $1/(nh)$. For these reasons, we require that as $n \to \infty$, $h \to 0$ in such a way that $nh \to \infty$.

It is well known that the asymptotic bias of such nonparametric regression estimators is proportional to h^p, where p depends on the smoothness of r, the smoothness of the marginal density of the X_i, and the properties of the kernel, or in the case of local polynomials, the polynomial degree of the local fit. In this paper we show that through appropriate choice of kernel, the rate at which the bias converges to zero will only be limited by properties of the unknown function, and not the kernel.

Sections 2 and 3 contain some important definitions and background. The case where the pairs of data $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d. will be studied in Section 4; the case where the data satisfy strong mixing conditions will be presented in Section 5; a small simulation study is presented in Section 6. Technical proofs have been placed in Section 7.

2. INFINITE ORDER KERNELS

If the kernel K has finite moments up to order q and its first $q-1$ moments are 0, then K is said to be of *order* q. The most frequently used kernels are second order; common examples include the Epanechnikov kernel, $K_e(x) := (3/4)\,(1-x^2)\,1_{[-1,1]}(x)$, and the scaled normal density.

In general, if r is k times differentiable, with $k \geq 2$, the bias of a second order kernel estimate is $O(h^2)$. This rate of convergence can be improved up to $O(h^k)$ by choosing a kernel of order greater than or equal to k. However, the degree of smoothness in the underlying function is unknown and difficult to estimate, so it is difficult to know what order kernel to use.

In order to alleviate this difficulty we focus on a class of kernels that effectively have infinite order. These kernels automatically reduce the bias to $o(h^k)$ no matter how large k happens to be. As in Politis and Romano [18, 19, 20] and Politis [16, 17], we now state the following general definition.

Definition 2.1. A general flat-top kernel K is defined in terms of its Fourier transform λ, which in turn is defined as follows. Fix a constant $c > 0$. Let

$$\lambda(s) = \begin{cases} 1 & \text{if } |s| \leq c\,, \\ g\big(|s|\big) & \text{if } |s| > c\,, \end{cases} \tag{2.1}$$

where the function g is chosen to make $\lambda(s)$, $\lambda^2(s)$, and $s\,\lambda(s)$ integrable. The flat top kernel is now given by

$$K(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda(s)\, e^{-isx}\, ds\,, \tag{2.2}$$

i.e., the inverse Fourier transform of $\lambda(s)$.

Note that in the preceding definition, the choice of g is not unique. The function λ, and hence the kernel K, depend on the function g and the parameter c although this dependence will not be explicitly denoted.

Kernels satisfying this definition do not necessarily satisfy the moment conditions $\int z^k K(z)\, dz = 0$ for all integers k, as some of these integrals may not

be defined in either the Riemann or Lebesgue sense. However, the Cauchy principal value of each of these integrals is zero, and in many cases this is sufficient for optimal asymptotic performance.

The simplest kernel satisfying Definition 2.1 is determined by

$$\lambda_D(s) = \begin{cases} 1 & \text{if } |s| \le 1\,, \\ 0 & \text{if } |s| > 1\,. \end{cases}$$

This is the example studied in the case of density estimation by Davis [2, 3], Devroye [4], and Ibragimov and Hasminksii [12], and it generates the Dirichlet kernel, $K(x) := \sin(x)/(\pi\, x)$. Both $\lambda(s)$ and the resulting kernel are shown in Figure 1. We can see that the tails of this kernel are very wiggly. This is problematic in two ways. First, the slow decay in the tails and the large negative oscillations increase $\int K^2(z)\, dz$, which will be shown to increase the variance of the estimate. Secondly, the large wiggles distant from 0 generate a finite sample bias because they allow observations which are relatively distant from x to have a substantial influence on the estimate at x. These difficulties make density estimators using this kernel relatively uncompetitive for all but extremely large sample sizes.

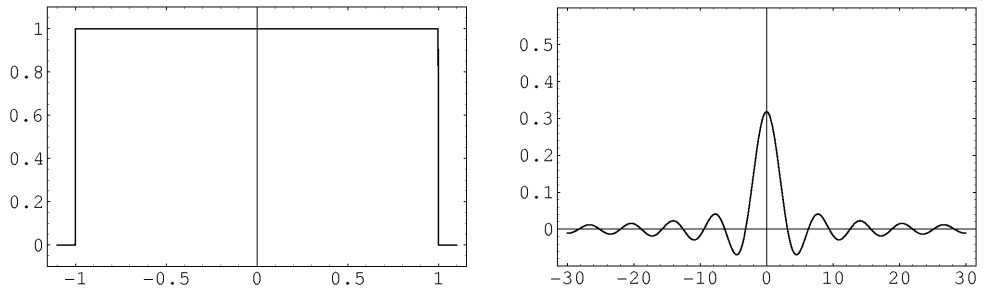

Figure 1: $\lambda(s)$ and the resulting Dirichlet kernel.

These problems can be substantially remedied by making the transition from 0 to 1 in the Fourier domain less abrupt. For example, Devroye and Gyorfi [5], Hall and Marron [11], and in the case of spectral density estimation, Politis and Romano [19], studied the kernel whose Fourier transform is given by

$$\lambda_{T,1/2}(s) = \begin{cases} 1 & \text{if } |s| \le 1/2\,, \\ 2\,\big(1-|s|\big) & \text{if } 1/2 < s \le 1\,, \\ 0 & \text{if } |s| > 1\,. \end{cases} \tag{2.3}$$

The corresponding kernel is

$$K(x) = \frac{2\,\big(\cos(x/2) - \cos(x)\big)}{\pi\, x^2}\,.$$

These are shown in Figure 2. Note the substantial improvement in the tails of the kernel.

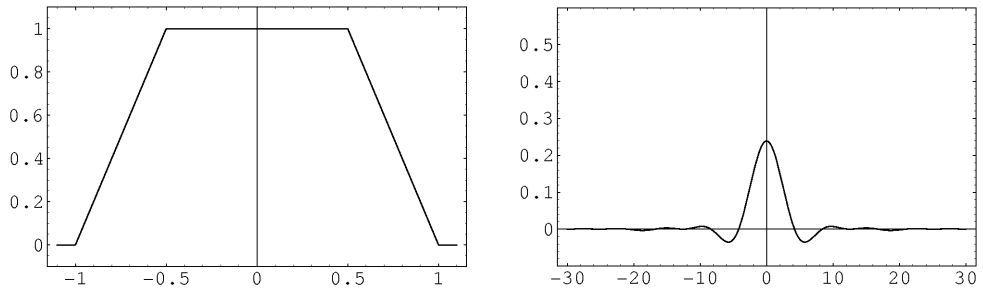

Figure 2: $\lambda(s)$ and the resulting improved kernel.

Unfortunately, in the case of regression, it is often only reasonable to assume that the function being estimated is smooth over some interval rather than over its entire domain; if the marginal density of the X_i has compact support, then the endpoints often generate discontinuities. Since infinite order kernels do not have compact support, the effects caused by these breaks get spread across the whole region of interest, potentially worsening the rate of convergence. For this reason, as discussed in the case of discontinuous density estimation in Politis [16], it is important that the tails of K decay as quickly as possible, to minimize the effect on the interior of the interval. This can be ensured by requiring the Fourier transform of the kernel to be very smooth. If λ is infinitely differentiable, then the tails of $K(x)$ decay faster than x^{-m} for any positive m. In addition, $\lambda(s)$ as defined in equation (2.1) clearly has an infinite number of zero derivatives at $s = 0$. Together, these two conditions ensure that all moments of K are zero in the Lebesgue sense. For these reasons, for the remainder of this work, we will restrict ourselves to kernels satisfying the following stronger definition.

Definition 2.2. An infinitely differentiable flat-top kernel K is a flat-top kernel (as in Definition 2.1) with the added caveat that the function g is chosen to make $\lambda(s)$ infinitely differentiable for all s.

We now provide an example of such a kernel, which was first introduced in McMurry and Politis [14], where the case of fixed design regression was studied. Let b and c be constants satisfying $b > 0$ and $0 < c < 1$. Define $\lambda(s)$ by

$$\lambda_{IO}(s) = \begin{cases} 1 & \text{if } |s| \le c\,, \\ \exp\Big[-b\,\exp\Big[-b/\big(|s|-c\big)^2\Big]/\big(|s|-1\big)^2\Big] & \text{if } c<|s|<1\,, \\ 0 & \text{if } |s| \ge 1\ . \end{cases} \tag{2.4}$$

The parameter c determines the region over which the kernel is identically 1; the parameter b allows the shape of λ to be altered, making the transition from 0 to 1 less abrupt. Figure 3 show plots of λ (as defined above) and the resulting kernel K for $c = 0.05$ and $b = 1/4$.

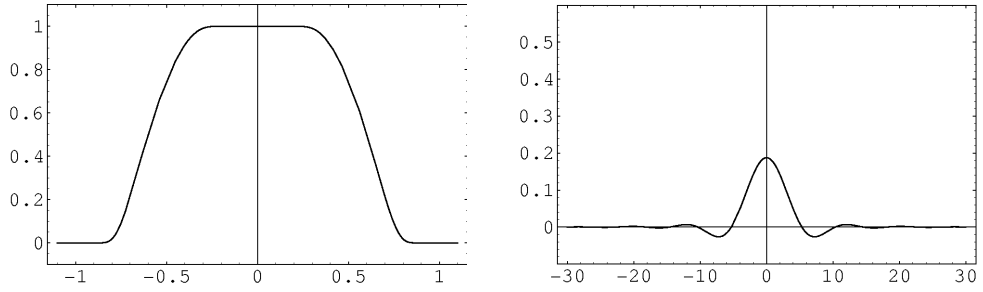

Figure 3: Smooth $\lambda(s)$ and the resulting kernel with $b = 1/4$ and $c = 0.05$.

The function $\exp\big[-b\,\exp[-b/(|s|-c)^2]/(|s|-1)^2\big]$ was chosen because it connects the regions where λ is 0 and the region where λ is 1 in a manner such that $\lambda(s)$ is infinitely differentiable for all s, including where $|s| = c$, and $|s| = 1$.

3. BACKGROUND AND NOTATION

We examine the performance of the Nadaraya–Watson estimator when using infinite order kernels. The observed data is assumed to take the form of identically distributed pairs $(X_1, Y_1), ..., (X_n, Y_n)$, which satisfy $Y_i = r(X_i) + \epsilon_i$. Further restrictions will be necessary, but their discussion will be postponed for the moment. The Nadaraya–Watson estimator introduced in equation (1.1) can be written as

$$\hat{r}(x) := \frac{(1/n)\,\sum_{i=1}^{n} Y_i\,K_h(X_i - x)}{(1/n)\,\sum_{i=1}^{n} K_h(X_i - x)}\ ,$$

where

$$K_h(x) := (1/h)\,K(x/h)\ .$$

This estimator should be viewed in two different ways. As mentioned before, it is a weighted local average of the Y_i's, where the denominator normalizes the weights so they sum to 1. It is also an explicit estimator of conditional expectation. The denominator is the standard kernel estimate of the design density, the marginal density of the X_i's. The numerator is an approximation to $\int_{-\infty}^{\infty} y\,f(x,y)\,dy$, where $f(x,y)$ is the joint density of (X_i, Y_i). Put together, this is an approximation to $r(x) = E\big[Y|X = x\big]$.

In order to simplify notation, define

$$\hat{g}(x) := (1/n) \sum_{i=1}^{n} Y_i \, K_h(X_i - x)$$

and

$$\hat{f}(x) := (1/n) \sum_{i=1}^{n} K_h(X_i - x) \ ,$$

which are the finite sample approximations to

$$g(x) := \int_{-\infty}^{\infty} y \, f(x,y) \ dy$$

and

$$f(x) := \int_{-\infty}^{\infty} f(x,y) \ dy \ .$$

We note that $r(x) = g(x)/f(x)$.

4. THE INFINITE ORDER NADARAYA-WATSON ESTIMATOR FOR I.I.D. DATA

We first examine the behavior of the Nadaraya–Watson estimator when the observed pairs of data, $(X_1, Y_1), ..., (X_n, Y_n)$, are i.i.d. In order to understand the estimator as a whole, we begin with lemmas quantifying the asymptotic performance of the numerator and denominator, $\hat{g}(x)$ and $\hat{f}(x)$, as they approximate $g(x)$ and $f(x)$. In the process, it will be necessary to impose some assumptions, which will be introduced and discussed as needed. We first place some reasonable restrictions on the behavior of the bandwidth h as the sample size grows large and on the conditional distribution of the errors.

Assumption 1. As the sample size $n \to \infty$, the bandwidth $h \to 0$ in such a way that $nh \to \infty$.

Assumption 2. $E\big[\epsilon_i | X_i = x\big] = 0$, and $E\big[\epsilon_i^2 | X_i = x\big] := \sigma^2(x) < \infty$.

Under this assumption, $\hat{f}(x)$ and $\hat{g}(x)$ are infinite order estimators of $f(x)$ and $g(x)$; this is quantified in the following lemma.

Lemma 4.1. *If x is contained in an open interval on which $f(x)$ has p bounded continuous derivatives and $r(x)$ has q bounded continuous derivatives, then under Assumptions 1 and 2,*

$$E\big[\hat{f}(x)\big] - f(x) \ = \ o(h^p) \tag{4.1}$$

and

$$E\big[\hat{g}(x)\big] - g(x) \;=\; o(h^k)\;, \tag{4.2}$$

where $k = \min\{p, q\}$. *If both* $f(x)$ *and* $g(x)$ *are infinitely differentiable, then each of these biases become* $o(h^m)$ *for all positive real* m.

If we impose the additional assumptions that the observed pairs of data are i.i.d., and that f, g, and $\sigma^2(x)$ are reasonably well behaved, then the variance of $\hat{f}$ and $\hat{g}$ also behaves as expected.

Assumption 3. $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d.

The next assumption is necessary to ensure that the asymptotic approximations we use are valid, and to avoid division by zero.

Assumption 4. The point x is a continuity point of $\sigma^2(x)$, $f(x) > C$ for some $C > 0$, and r and f are each differentiable in a neighborhood of x.

Lemma 4.2. *Under Assumptions 1–4,*

$$\operatorname{var}\big[\hat{f}(x)\big] \;=\; \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(z)\, dz \;+\; o\!\left(\frac{1}{nh}\right) + O\!\left(\frac{1}{n}\right), \tag{4.3}$$

$$\operatorname{var}\big[\hat{g}(x)\big] \;=\; \frac{\Big(r^2(x) + \sigma^2(x)\Big) f(x)}{nh} \int_{-\infty}^{\infty} K^2(z)\, dz \;+\; o\!\left(\frac{1}{nh}\right) + O\!\left(\frac{1}{n}\right) \tag{4.4}$$

and

$$\operatorname{cov}\big[\hat{f}(x), \hat{g}(x)\big] \;=\; \frac{r(x)\, f(x)}{nh} \int_{-\infty}^{\infty} K^2(z)\, dz \;+\; o\!\left(\frac{1}{nh}\right) + O\!\left(\frac{1}{n}\right). \tag{4.5}$$

Now that the behaviors of $\hat{f}$ and $\hat{g}$ are understood independently, the analysis will proceed by establishing their joint asymptotic normality. Once this has been shown, a Taylor series argument can be employed to show that $\hat{r}$ also has an asymptotic normal distribution with optimal bias and the standard variance. The joint asymptotic normality of $\hat{f}$ and $\hat{g}$ will be established via the Liapunov condition, which implies the Lindeberg–Feller central limit theorem. This requires a uniform bound on the $2+\delta$'th moments of the Y_i, for some $\delta > 0$.

Assumption 5. There exists a positive constants M and δ such that

$$E\Big[\,|Y_i|^{2+\delta}\,\Big|X_i = x\Big] < M\;,$$

for all x.

The final assumption forces the conditional variance of the errors to be bounded above and below for all x. The bound from below is assumed for technical simplicity.

Assumption 6. There exist strictly positive constants b and B such that $b < \sigma^2(x) < B$ for all x.

Lemma 4.3. *Under Assumptions 1–6, for all real c_1 and c_2 (not both zero),*

$$\sqrt{nh}\left[c_1\Big(\hat{f}(x) - E\big[\hat{f}(x)\big]\Big) + c_2\Big(\hat{g}(x) - E\big[\hat{g}(x)\big]\Big)\right] \xrightarrow{\mathcal{D}} N\big(0, \theta(x)\big)\,. \tag{4.6}$$

where $\theta(x) := \Big(c_1^2 + 2\,c_1\,c_2\,r(x) + c_2^2\big[r^2(x) + \sigma^2(x)\big]\Big)\,f(x)\,\int_{-\infty}^{\infty} K^2(z)\,dz$. *This implies the joint asymptotic normality of $\hat{f}$ and $\hat{g}$.*

The consequence of the preceding Lemma is the asymptotic normality of our estimator.

Theorem 4.1. *If x is contained in an open interval on which $f(x)$ has p bounded continuous derivatives and $r(x)$ has q bounded continuous derivatives, then under Assumptions 1–6,*

$$\sqrt{nh}\left(\hat{r}(x) - r(x) + o(h^k)\right) \xrightarrow{\mathcal{D}} N\left(0\,,\ \frac{\sigma^2(x)}{f(x)}\int_{-\infty}^{\infty} K^2(z)\,dz\right), \tag{4.7}$$

where $k = \min\{p, q\}$.

Remark 4.1. Letting h proportional to $n^{-1/(2k+1)}$, the mean square optimal rate, we get $\hat{r}(x) = r(x) + O_p\big(n^{-k/(2k+1)}\big)$, and $\sqrt{nh}\big(\hat{r}(x) - r(x)\big) \xrightarrow{\mathcal{D}} N\Big(0,\ \frac{\sigma^2(x)}{f(x)}\int_{-\infty}^{\infty} K^2(z)\,dz\Big)$, which demonstrates the higher order accuracy provided by infinite order kernels.

5. DEPENDENT DATA AND NONPARAMETRIC AUTOREGRESSION

It is desirable to weaken the condition that $(X_1, Y_1), ..., (X_n, Y_n)$ are i.i.d. In particular we wish to be able to estimate nonparametric autoregression, where X_t may be an unknown function of X_{t-1}. Mathematically, autoregressive processes are assumed to satisfy a model of the form, $X_t = r(X_{t-1}) + \sigma(X_{t-1})\,\epsilon_t$, where r and σ are unknown, and the ϵ_t are mean zero errors; further restrictions

similar to those in Section 4 will be imposed as necessary. The problem of interest is the estimation of $r(x) := E\big[X_t | X_{t-1} = x\big]$; this can be done by pairing consecutive observations, $(X_1, X_2), (X_2, X_3), ..., (X_{n-1}, X_n)$, and then performing a standard nonparametric regression.

Since nothing is gained by restricting ourselves to the case of autoregression, we will study the infinite order kernel estimator in the case where $(X_1, Y_1), ..., (X_n, Y_n)$ satisfy the same type of asymptotic dependence conditions that we wish the autoregressive process to satisfy. The results in this more general situation will then imply the desired result in the specific case of interest.

Any meaningful analysis will require conditions that ensure some sort of asymptotic independence; that is, random samples at times which are very distant from each other should behave as if they are independent. We will focus on the case of α-mixing because it is the weakest of the most commonly studied conditions.

Definition 5.1. Let $\mathcal{F}_l^m$ be the σ-field generated by $U_l, U_{l+1}, ..., U_m$. A stationary time series $\{U_n\}_{n\in\mathbb{Z}}$ is said to be α-mixing (or strong-mixing), if

$$\sup_{k\in\mathbb{Z}} \sup_{A\in\mathcal{F}_{-\infty}^k, B\in\mathcal{F}_{k+i}^\infty} \Big| P(A)\,P(B) - P(AB) \Big| := \alpha(i) \;\to\; 0 \;,$$

as $i \to \infty$. The $\alpha(i)$'s are called the α-mixing coefficients.

The analysis proceeds as in the i.i.d. case. We begin by proving, under some conditions, the joint asymptotic normality of $\hat{f}(x)$ and $\hat{g}(x)$. Once this has been established, we will be able to use the same argument used in the proof of Theorem 4.1 to show the asymptotic normality of $\hat{g}/\hat{f}$. Similar results for local polynomial regression and for Nadaraya–Watson estimators with finite order kernels have been obtained by Masry and Fan [13] and Robinson [21] respectively. Although their arguments are similar, the central limit theorem we prove here will be more closely related to that of Masry and Fan [13]. Let c_1 and c_2 be real numbers (not both zero). Define

$$Z_i \;:=\; c_1\Big[K_h(X_i - x) - E\big[K_h(X_i - x)\big]\Big] + c_2\Big[Y_i\,K_h(X_i - x) - E\big[Y_i\,K_h(X_i - x)\big]\Big]$$

and

$$Q_n := \frac{1}{n}\sum_{i=1}^{n} Z_i \;.$$

We will establish asymptotic normality for $\sqrt{nh}\,Q_n$. In order to do so, we impose further assumptions on the marginal distributions of (X_1, Y_1), and on the α-mixing coefficients associated with the time series defined by $U_i = (X_i, Y_i)$.

Assumption 7. There exist finite positive bounds M_1, M_2, and M_3, such that

(**i**) $f_i(u,v) \le M_1$, where $f_i(u,v)$ is the joint density of (X_1, X_i);

(**ii**) $E\big[Y_1^2 + Y_i^2 | X_1,\, X_i\big] \le M_2$;

(**iii**) There exists $\delta > 2$ and $\beta > 1 - 2/\delta$ such that $E\big[|Y_1|^\delta \,|\, X\big] \le M_3$ and $\sum_{i=1}^{\infty} i^\beta [\alpha(i)]^{1-2/\delta} < \infty$.

Lemma 5.1. *Let x be a continuity point of conditional mean and variance functions, $r(\cdot)$ and $\sigma^2(\cdot)$. In addition, suppose that the marginal density of the X_i, $f(\cdot)$, and the product of $r(\cdot)$ and $f(\cdot)$, $g(\cdot)$, have k bounded continuous derivatives in a neighborhood of x, where $k \ge 1$. Under Assumptions 2, 4, 6, and 7, we have the following convergences as $n \to \infty$, $h \to 0$, and $nh \to \infty$:*

(**a**) $h \operatorname{var}[Z_1] \to \theta(x)$,

(**b**) $h \sum_{i=1}^{n-1} \big|\operatorname{cov}[Z_1, Z_{i+1}]\big| \to 0$,

(**c**) $nh \operatorname{var}[Q_n] \to \theta(x)$,

where, as before, $\theta(x) := \Big(c_1^2 + 2\, c_1\, c_2\, r(x) + c_2^2 \big[r^2(x) + \sigma^2(x)\big]\Big)\, f(x) \int_{-\infty}^{\infty} K^2(z)\, dz$.

In order to establish the central limit theorem, we need one final condition on the α-mixing coefficients.

Assumption 8. There exists a sequence of positive integers satisfying $s_n \to \infty$ and $s_n = o(\sqrt{nh})$ such that $\sqrt{n/h}\, \alpha(s_n) \to 0$.

Remark 5.1. Assumption 8 is a technical assumption that may dictate some particular rates at which $h \to 0$; nevertheless, Assumption 8 is weak enough to allow for a wide range of useful rates. To elaborate, Assumption 7 (iii) requires that the mixing coefficients decay at a polynomial rate depending on δ. In particular, it requires that there exist $C > 0$, $\epsilon > 0$, and $n_0 > 0$ such that for all $n > n_0$, $\alpha(n) < C n^{-(\delta/(\delta-2)+1+\epsilon)}$. For example, if $\delta = 3$, then the mixing coefficients need to decay slightly faster than n^{-4}. Assuming that the function being estimated is at least twice differentiable, h will optimally decrease at a rate equal to or slower than $n^{-1/5}$. This means that $\sqrt{n/h} \le C n^{3/5}$ for some constant C. Similarly, $\sqrt{nh} \ge C n^{2/5}$. Under the strongest moment assumptions, $\alpha(n)$ is required to decay faster than n^{-2}. If we put these together, $\sqrt{n/h}\, \alpha(s_n) \le C n^{3/5} s_n^{-2}$. From this expression it is easily seen that as long as s_n grows faster than $n^{3/10}$, the second requirement of Assumption 8 will be satisfied. By the preceding argument the first requirement is satisfied if $s_n = o(n^{2/5})$. Since these conditions can be met simultaneously, Assumption 8 generally imposes no additional restrictions.

Lemma 5.2. *Under Assumptions 2, 4, and 6–8, we have as* $n \to \infty$, $h \to 0$, *and* $nh \to \infty$,

$$\sqrt{nh}\, Q_n \xrightarrow{\mathcal{D}} N\big(0, \theta(x)\big)\,.$$

The immediate consequence of this result is the following asymptotic normality for $\hat{r}$.

Theorem 5.1. *If* x *is contained in an open interval on which* $f(x)$ *has* p *bounded continuous derivatives and* $r(x)$ *has* q *bounded continuous derivatives, then under Assumptions 1, 2, 4, and 6–8,*

$$\sqrt{nh}\left(\hat{r}(x) - r(x) + o(h^k)\right) \xrightarrow{\mathcal{D}} N\left(0\,,\, \frac{\sigma^2(x)}{f(x)} \int_{-\infty}^{\infty} K^2(z)\, dz\right), \tag{5.1}$$

where $k = \min\{p, q\}$.

6. SIMULATIONS

An extensive simulation study was undertaken to investigate the performance of the proposed estimator. For each combination of regression function, design density, error variance, and sample size, 100 data sets were created and smoothed using the infinite order and local linear estimators. Finally the integrated square error was estimated using Simpson's rule. Bandwidths for the infinite order estimator were selected by the rule of thumb suggested in [14] and developed further in [17]. Bandwidths for the local linear estimator were selected using the direct plug-in method suggested by Ruppert, Sheather, and Wand [22] and implemented in the R package KernSmooth [24].

The first regression function was taken to be $r(x) = x + 4\exp(-2x^2)/\sqrt{2\pi}$, which includes sections of almost linear behavior and an exponential bump with more curvature. Design densities were uniform on $[-2, 2]$ and $N(0, 1)$, and the integrated square error is over the interval $[-2, 2]$. The resulting integrated square errors for one simulation are shown in Figure 4. A scatterplot along with the two smoothings is shown in Figure 5.

The second regression function was taken to be $r(x) = \sin(4\pi x)$ with uniform design density on $[0, 1]$. The integrated square error was calculated on both the entire interval $[0, 1]$ and over the interval $[0.15, 0.85]$ to exclude edge effects.

It is clear from the simulations that the two estimators have different strengths. The infinite order estimator is clearly superior in the interior of the data set, when the error variance is large, and when the sample size is moderate to large. Since the local linear estimator automatically adapts to the edges of the design and the infinite order estimator does not, it is unsurprising that the local linear estimator is superior in these regions.

Table 1: Comparison of infinite order and local linear estimators.

Function	Design	n	σ	Median Integrated Square Error	
				Infinite Order	Local Linear
Exponential	Normal	100	0.3	0.0870	0.0631
			0.5	0.1312	0.1627
			0.7	0.2101	0.2521
Exponential	Normal	200	0.3	0.0337	0.0324
			0.5	0.0633	0.0686
			0.7	0.1091	0.1375
Exponential	Normal	1000	0.3	0.0065	0.0077
			0.5	0.0132	0.0166
			0.7	0.0204	0.0274
Exponential	Uniform	100	0.3	0.0520	0.0384
			0.5	0.0954	0.1613
			0.7	0.0813	0.1481
Exponential	Uniform	200	0.3	0.0251	0.0190
			0.5	0.0481	0.0474
			0.7	0.0731	0.0823
Exponential	Uniform	1000	0.3	0.0066	0.0051
			0.5	0.0110	0.0112
			0.7	0.0175	0.0189
Sin (edges excluded)	Uniform	100	0.3	0.0065	0.0052
			0.5	0.0135	0.0139
			0.7	0.0232	0.0228
Sin (edges included)	Uniform	100	0.3	0.0120	0.0077
			0.5	0.0221	0.0191
			0.7	0.0412	0.0333
Sin (edges excluded)	Uniform	200	0.3	0.0032	0.0031
			0.5	0.0069	0.0068
			0.7	0.0108	0.0126
Sin (edges included)	Uniform	200	0.3	0.0066	0.0042
			0.5	0.0115	0.0091
			0.7	0.0191	0.0183
Sin (edges excluded)	Uniform	1000	0.3	0.0007	0.0008
			0.5	0.0013	0.0019
			0.7	0.0021	0.0030
Sin (edges included)	Uniform	1000	0.3	0.0029	0.0011
			0.5	0.0037	0.0026
			0.7	0.0050	0.0040

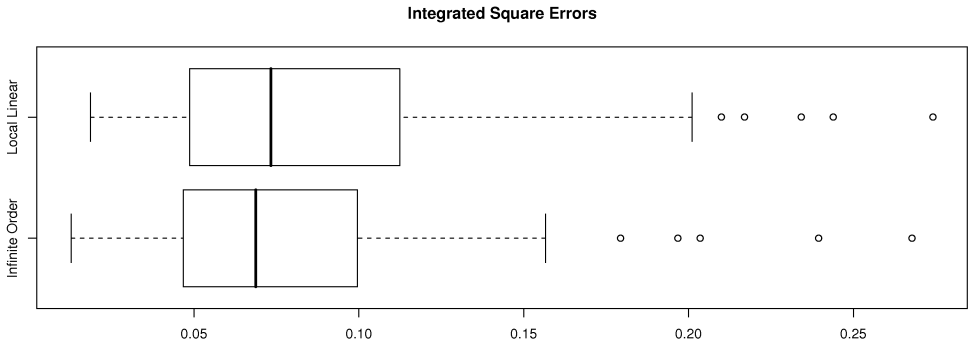

Figure 4: Comparison of integrated square errors for 100 simulations of the exponential function with $n = 200$ and $\sigma = 0.5$.

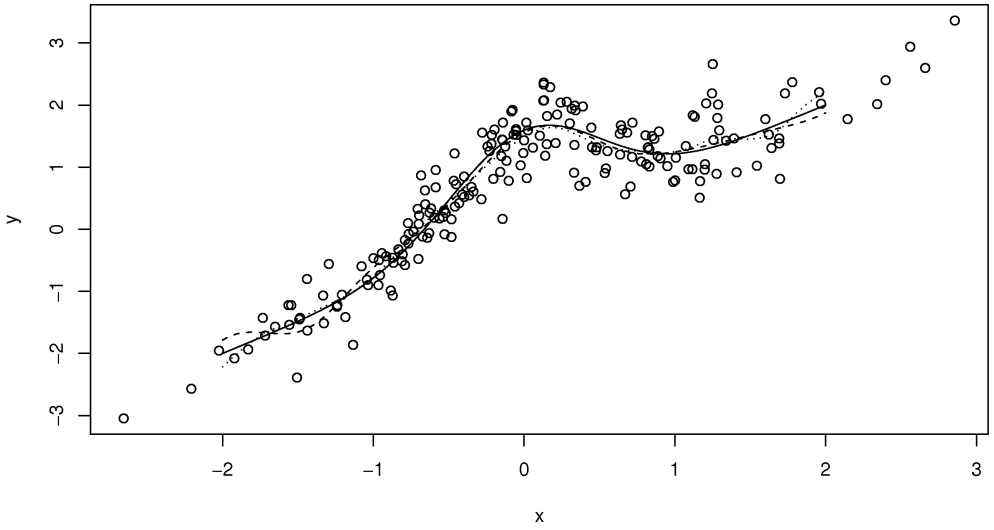

Figure 5: A sample regression. The solid line is the true function, the dashed line is the infinite order estimate, and the dotted line is the local linear estimate.

7. TECHNICAL PROOFS

Proof of Lemma 4.1: The proof of (4.2) is almost identical to, but slightly more complicated than the proof of (4.1). For this reason, only (4.2) will be shown. The proof is similar in spirit to the proof of Theorem 2 in McMurry and Politis [14], except this time it requires both r and f to be smooth. It will be proved using a different, although more standard, technique. This method of

proof is somewhat less elegant than the Fourier transform method used previously, but it has the advantage of showing that the same convergence rates hold even if f and g are only smooth on an open interval containing x. This proof technique does have a slight disadvantage. If f and g are smooth over all $\mathbb{R}$, then the Fourier transform technique can be employed to show the same convergence rates hold even if $\lambda(s)$, the Fourier transform of K, is not smooth. By conditioning on X_i,

$$\begin{aligned} E\big[\hat{g}(x)\big] - g(x) &= E\Big[E\big[Y_i\, K_h(X_i - x)\,\big|\, X_i\big]\Big] - g(x) \\ &= E\Big[r(X_i)\, K_h(X_i - x)\Big] - g(x) \\ &= \int_{-\infty}^{\infty} r(u)\, f(u)\, K_h(u-x)\, du \;-\; g(x)\;. \end{aligned}$$

Suppose (rf) has k bounded continuous derivatives on an interval (a,b) containing x. Should (rf) be smooth over all $\mathbb{R}$, then the proof can be simplified by taking $(a,b) = (-\infty, \infty)$:

$$\begin{aligned} E\big[\hat{g}(x)\big] - g(x) \;=\; & \int_a^b r(u)\, f(u)\, K_h(u-x)\, du \;+\; \int_{-\infty}^{a} r(u)\, f(u)\, K_h(u-x)\, du \\ & +\; \int_b^{\infty} r(u)\, f(u)\, K_h(u-x)\, du \;-\; g(x)\;. \end{aligned}$$

Since the tails of $K(x)$ decay faster than x^{-m} for all positive m, the two error terms are $o(h^m)$ for all positive m. At this point we perform a Taylor series expansion of the product $(rf)(z)$ around x:

$$\begin{aligned} & E\big[\hat{g}(x)\big] - g(x) \;= \\ & \quad = \int_a^b r(u)\, f(u)\, K_h(u-x)\, du \;-\; g(x) \;+\; o(h^m) \\ & \quad = \int_{(a-x)/h}^{(b-x)/h} r(x+hv)\, f(x+hv)\, K(v)\, dv \;-\; g(x) \;+\; o(h^m) \\ & \quad = \int_{(a-x)/h}^{(b-x)/h} \left[(rf)(x) + hv(rf)'(x) + \cdots + \frac{(hv)^k}{k!}\,(rf)^{(k)}\,(x+\xi)\right] K(v)\, dv \\ & \qquad -\; g(x) \;+\; o(h^m)\;, \end{aligned}$$

where ξ is between x and $x + hv$. Since K integrates to one, its moments are zero, and since $g(x) = r(x)\, f(x)$,

$$\begin{aligned} & E\big[\hat{g}(x)\big] - g(x) \;= \\ & \quad = \int_{-\infty}^{\infty} \left[(rf)(x) + hv(rf)'(x) + \cdots + \frac{(hv)^{k-1}}{(k-1)!}\,(rf)^{(k-1)}(x)\right] K(v)\, dv \\ & \qquad +\; \int_{(a-x)/h}^{(b-x)/h} \frac{(hv)^k}{k!}\,(rf)^{(k)}(x+\xi)\, K(v)\, dv \;-\; g(x) \;+\; o(h^m) \\ & \quad = \int_{(a-x)/h}^{(b-x)/h} \frac{(hv)^k}{k!}\,(rf)^{(k)}(x+\xi)\, K(v)\, dv \;+\; o(h^m)\;. \end{aligned}$$

Since (rf) has bounded continuous derivatives on (a,b), we can apply the dominated convergence theorem, yielding

$$\lim_{h\to 0} \int_{(a-x)/h}^{(b-x)/h} (rf)^{(k)}(x+\xi)\, v^k\, K(v)\, dv \;=\; \int_{-\infty}^{\infty} (rf)^{(k)}(x)\, v^k\, K(v)\, dv \;=\; 0\ .$$

Therefore,

$$E\big[\hat{g}(x)\big] - g(x) \;=\; o(h^k)\ . \qquad \square$$

Proof of Lemma 4.2: As $f(x)$ can be viewed as a special case of $g(x)$ with $Y_i = 1$ for all i, (4.3) will follow immediately from the proof of (4.4). The proof of (4.5) is almost identical to the proof of (4.4), so it is omitted.

By conditioning on X_1, and by Lemma 4.1,

$$\begin{aligned}
\mathrm{var}\big[\hat{g}(x)\big] \;&=\; \frac{1}{n}\, \mathrm{var}\big[Y_1 K_h(X_1 - x)\big] \\
&=\; \frac{1}{n}\left[\int_{-\infty}^{\infty} \Big(r^2(u) + \sigma^2(u)\Big)\, f(u)\, K_h^2(u-x)\, du \;-\; r^2(x) \;+\; o(h^k)\right] \\
&=\; \frac{1}{nh} \int_{-\infty}^{\infty} \Big(r^2(x+hz) + \sigma^2(x+hz)\Big)\, f(x+hz)\, K^2(z)\, dz \;+\; O\!\left(\frac{1}{n}\right),
\end{aligned}$$

since x is a continuity point of r, f, and σ^2, the dominated convergence theorem yields

$$\mathrm{var}\big[\hat{g}(x)\big] \;=\; \frac{\Big(r^2(x) + \sigma^2(x)\Big)\, f(x)}{nh} \int_{-\infty}^{\infty} K^2(z)\, dz \;+\; o\!\left(\frac{1}{nh}\right) \;+\; O\!\left(\frac{1}{n}\right). \qquad \square$$

Proof of Lemma 4.3: The proof proceeds by verifying that the Liapunov condition holds, which is sufficient for the Lindeberg–Feller central limit theorem. The result is trivial if $c_1 = c_2 = 0$, so assume that at least one of these constants is nonzero. Let C denote a positive constant.

$$\begin{aligned}
&\frac{\sum_{i=1}^n E\Big[\big|(c_1 + c_2 Y_i)\, K_h(X_i - x)\big|^{2+\delta}\Big]}{\mathrm{var}\Big[c_1 \sum_{i=1}^n K_h(X_i - x) \;+\; c_2 \sum_{i=1}^n Y_i K_h(X_i - x)\Big]^{(2+\delta)/2}} \;\le \\
&\qquad\qquad \le\; \frac{C n h^{1+\delta}}{n^{1+\delta/2}\Big[\theta(x) + o(1/h) + O(1)\Big]^{(2+\delta)/2}}\ ,
\end{aligned}$$

where the inequality follows from the proof of Lemma 4.2. After multiplying the numerator and denominator by $h^{(2+\delta)/2}$, it is clear that this quantity goes to zero as n goes to infinity. Therefore, the Liapunov condition is satisfied, and the Lemma follows immediately. $\square$

Proof of Theorem 4.1: We begin with a lemma which ensures that for large enough n, $\hat{f}(x) \geq c > 0$ for some constant c, as long as $f(x) > 0$.

Lemma 7.1. *Suppose $f(x) > c$ for some $c > 0$. Also suppose h decreases slowly enough that $n^{-2/7} + \delta = o(h)$ for some $\delta > 0$. Then for all $\epsilon > 0$, $P\big[|f(x) - \hat{f}(x)| > \epsilon \text{ i.o.}\big] = 0$.*

It should also be noted that much stronger results hold. Under additional conditions, rates of uniform almost sure convergence over compact sets can be established. See Bosq [1] or Györfi *et al.* [9].

Proof of Lemma 7.1: We make use of the following Bernstein inequality, which is Theorem 1.3 part (2) in Bosq [1].

Lemma 7.2. *Let $\{W_t\}_{t\in\mathbb{Z}}$ be a mean zero real valued random process such that $\sup_{1\leq t\leq n} \|W_t\|_\infty \leq b$, and let $S_n = \sum_{t=1}^n W_t$. Then for each integer $q \in \lfloor 1, n/2 \rfloor$ and each $\epsilon > 0$,*

$$\mathrm{P}\big[|S_n| > n\epsilon\big] \;\leq\; 4\exp\!\left(-\frac{\epsilon^2}{8\,v^2(q)}\,q\right) + 22\left(1+\frac{4\,b}{\epsilon}\right)^{1/2} q\,\alpha\!\left(\left\lfloor\frac{n}{2\,q}\right\rfloor\right),$$

where $v^2(q) = \frac{2}{p^2}\,\sigma^2(q) + \frac{b\epsilon}{2}$, $p = \frac{n}{2q}$, and

$$\sigma^2(q) \;=\; \max_{0\leq j\leq 2q-1} E\Bigg[\Big(\lfloor jp\rfloor + 1 - jp\Big)X_{\lfloor jp\rfloor+1} + X_{\lfloor jp\rfloor+2} + \cdots + X_{\lfloor (j+1)p\rfloor} + \Big[(j+1)p - \big\lfloor(j+1)p\big\rfloor\Big]X_{\lfloor(j+1)p+1\rfloor}\Bigg].$$

The Borel–Cantelli lemma will be used to show

$$\mathrm{P}\Big[\big|\hat{f}_n(x) - E\,\hat{f}_n(x)\big| > \epsilon \text{ i.o.}\Big] \;=\; 0\;.$$

Since $E\,\hat{f}_n(x) \to f(x)$, this will establish the desired result.

Let $W_i = K_h(X_i - x) - E\big[K_h(X_i - x)\big]$. Then $\|W_i\|_\infty \leq \hat{K}/h$ for all i, where $\hat{K} = 2\sup_{x\in\mathbb{R}} K(x)$. In addition, it can easily be seen that $\sigma^2(q) \leq \frac{(p+1)^2}{n^2h^2}\,\hat{K}^2$.

Therefore,

$$\begin{aligned} p_n \;&:=\; \mathrm{P}\left[\frac{1}{n}\left|\sum_{i=1}^n W_i\right| > \epsilon\right] \\ &\leq\; 4\exp\!\left(-\frac{C_1\,\epsilon^2}{\frac{(p+1)^2}{n^2h^2p^2} + \frac{1}{h}}\,q\right) + C_2\left(1+\frac{C_3}{h\epsilon}\right)^{1/2} q\,\alpha\!\left(\left\lfloor\frac{n}{2\,q}\right\rfloor\right). \end{aligned}$$

We need $\sum_n p_n < \infty$. In order for the first term in the sum to be fininte, it is necessary that $n^2 h^2 q/(n^2 h) \to \infty$ at a rate equal to or faster than n^{δ_1} for some $\delta_1 > 0$; this requires q grow at least as fast as $n^{\delta_1} h^{-1}$. On the other hand, the second term requires $\sum_n (q/\sqrt{h})\, \alpha(\lfloor n/2q \rfloor) < \infty$. As noted in the Remark 5.1, it is sufficient to choose q such that $\sum_n (q/\sqrt{h})(q/n)^2 < \infty$. The latter condition can be satisfied if $q^3 h^{-1/2}$ grows at a rate $n^{1-\delta_2}$, for some $\delta_2 > 0$. Equivalently, it suffices for q to grow at a rate equal to or slower than $n^{(1/3)-\delta_2} h^{1/6}$. It can easily be seen that these requirements can be met simultaneously as long as h satisfies $n^{-2/7+\delta} = o(h)$, for some $\delta > 0$; this includes all optimal rates. This result could be further strengthened by imposing additional assumptions. For example, in the case where $X_1, ..., X_n$ are i.i.d., the mixing coefficients are 0, and hence summable. In this situation, the only restriction on h is that it decrease slightly slower than $1/n$. In the case of a mixing process, the possible range of rates for h could be expanded if one were to assume that the joint density of (X_1, X_i) is differentiable with partial derivatives uniformly bounded in i. □

We now return to the proof of the main result. By the preceding lemma, for large enough n, we can assume that $\hat{f}_n(x) > c/2$. So, we can apply the intermediate value theorem to see,

$$\begin{aligned} \hat{r}(x) - r(x) &= \frac{\hat{g}(x)}{\hat{f}(x)} - \frac{g(x)}{f(x)} \\ &= \hat{g}(x) \left(\frac{1}{f(x)} - \frac{1}{\xi_n^2} \Big(\hat{f}(x) - f(x) \Big) \right) - \frac{g(x)}{f(x)} \ , \end{aligned}$$

where $|\xi_n - f(x)| \le |\hat{f}(x) - f(x)|$. This can be further simplified to

$$\begin{aligned} \hat{r}(x) - r(x) &= \frac{1}{f(x)} \Big(\hat{g}(x) - g(x) \Big) - \frac{\hat{g}(x)}{\xi_n^2} \Big(\hat{f}(x) - f(x) \Big) \\ &= \frac{1}{f(x)} \Big(\hat{g}(x) - E\big[\hat{g}(x)\big] \Big) - \frac{\hat{g}(x)}{\xi_n^2} \Big(\hat{f}(x) - E\big[\hat{f}(x)\big] \Big) + o(h^k) \ . \end{aligned}$$

By Lemmas 4.1 and 4.2, $\hat{g}(x)$ and $\hat{f}(x)$ converge in probability to $g(x)$ and $f(x)$ respectively. Therefore, ξ_n also converges to $f(x)$ in probability. By Slutsky's theorem, and Lemma 4.3,

$$\sqrt{n\,h} \Big(\hat{r}(x) - r(x) + o(h^k) \Big) \ \xrightarrow{\mathcal{D}} \ N\left(0, \frac{\sigma^2(x)}{f(x)} \int_{-\infty}^{\infty} K^2(z)\, dz \right) ,$$

the desired result. □

Proof of Lemma 5.1: The proof of part (a) follows from similar results for the i.i.d. case:

$$\begin{aligned} \mathrm{var}\big[Z_1\big] &= E\Big[(c_1 + c_2 Y_1)^2 K_h^2(X_1 - x) \Big] - \Big(c_1 f(x) + c_2\, r(x) f(x) + O(h^k) \Big)^2 \\ &= E\left[\Big(c_1^2 + 2\, c_1 c_2\, r(X_1) + c_2^2 \big[r^2(X_1) + \sigma^2(X_1) \big] \Big) K_h^2(X_1 - x) \right] + O(1) \\ &= \theta(x)/h + o(1/h) + O(1) \ . \end{aligned}$$

The proof of part (b) is more challenging. Let d_n be a sequence of integers such that $d_n \to \infty$ and $d_n h \to 0$. Define

$$J_1 := \sum_{i=1}^{d_n-1} \big|\mathrm{cov}[Z_1, Z_{i+1}]\big| \,,$$

and

$$J_2 := \sum_{i=d_n}^{n-1} \big|\mathrm{cov}[Z_1, Z_{i+1}]\big| \,.$$

We wish to show $J_1 = o(1/h)$ and $J_2 = o(1/h)$. We begin with J_1. By conditioning on (X_1, X_i),

$$\begin{aligned}
\big|\mathrm{cov}[Z_1, Z_i]\big| \;&\leq \\
&\leq \; \Big| E\Big[(c_1 + c_2 Y_1)\, K_h(X_1 - x)\, (c_1 + c_2 Y_i)\, K_h(X_i - x)\Big]\Big| + O(1) \\
&\leq \; E\Big[\big|K_h(X_1 - x)\, K_h(X_i - x)\big|\, E\Big[\big|(c_1 + c_2 Y_1)\, (c_1 + c_2 Y_i)\big|\,\Big|\, X_1, X_i\Big]\Big] + O(1) \\
&\leq \; E\Big[\big|K_h(X_1 - x)\, K_h(X_i - x)\big| \\
&\quad \times \Big(E\Big[(c_1 + c_2 Y_1)^2\,\big|\, X_1, X_i\Big]\, E\Big[(c_1 + c_2 Y_i)^2\,\big|\, X_1, X_i\Big]\Big)^{1/2}\Big] + O(1) \\
&\leq \; C\, E\Big[\big|K_h(X_1 - x)\, K_h(X_i - x)\big|\Big] + O(1) \\
&\leq \; C\left(\int_{-\infty}^{\infty} \big|K_h(u - x)\big|\, du\right)^2 + O(1) \,.
\end{aligned}$$

Since the $O(1)$ term is the same for all i, $\big|\mathrm{cov}[Z_1, Z_i]\big| < C$ for some positive C. Therefore $J_1 = o(1/h)$. For the second term, J_2, we employ Davydov's Lemma (see Hall and Heyde [10]), which tells us

$$\big|\mathrm{cov}[Z_1, Z_{i+1}]\big| \;\leq\; 8\,\big[\alpha(i)\big]^{1-2/\delta}\,\big[E|Z_1|^{\delta}\big]^{2/\delta} \,.$$

We now need to put a bound on $E|Z_i|^{\delta}$.

$$\begin{aligned}
E|Z_i|^{\delta} \;&=\; E\Bigg[\Big| c_1\Big[K_h(X_i - x) - E\big[K_h(X_i - x)\big]\Big] \\
&\qquad +\; c_2\Big[Y_i\, K_h(X_i - x) - E\big[Y_i\, K_h(X_i - x)\big]\Big]\Big|^{\delta}\Bigg] \\
&\leq\; 2\, E\Bigg[\Big| c_1\Big[K_h(X_i - x) - E\big[K_h(X_i - x)\big]\Big]\Big|^{\delta}\Bigg] \\
&\qquad +\; 2\, E\Bigg[\Big| c_2\Big[Y_i\, K_h(X_i - x) - E\big[Y_i\, K_h(X_i - x)\big]\Big]\Big|^{\delta}\Bigg] \,.
\end{aligned} \tag{7.1}$$

These two terms behave similarly, so it is sufficient to examine only the second. Let C denote a generic positive constant which may take on different values:

$$\begin{aligned} 2\,E\left[\,\left|\,c_2\Big[Y_i\,K_h(X_i-x) - E\big[Y_i\,K_h(X_i-x)\big]\Big]\,\right|^{\delta}\right] \;\leq \\ \leq\; 4\,E\Big[\,\big|\,c_2\,Y_i\,K_h(X_i-x)\,\big|^{\delta}\Big] \;+\; 4\Big|E\big[Y_i\,K_h(X_i-x)\big]\Big|^{\delta} \\ \leq\; C\,E\Big[\big|K_h(X_i-x)\big|^{\delta}\,E\big[|Y_i|^{\delta}\big|X_i\big]\Big] + C \\ \leq\; C\,h^{1-\delta} + C\;. \end{aligned}$$

An identical result holds for the first term in equation (7.1). Putting these two terms together yields

$$\begin{aligned} \big[E|Z_1|^{\delta}\big]^{2/\delta} \;&\leq\; \big[C\,h^{1-\delta} + C\big]^{2/\delta} \\ &\leq\; C\,h^{2/\delta-2} + C\;. \end{aligned}$$

Returning to J_2,

$$\begin{aligned} J_2 \;&\leq\; \sum_{i=d_n}^{\infty} 8\,\big[\alpha(i)\big]^{1-2/\delta}\,\big[E|Z_1|^{\delta}\big]^{2/\delta} \\ &\leq\; \sum_{i=d_n}^{\infty} \big[\alpha(i)\big]^{1-2/\delta}\big(Ch^{2/\delta-2} + C\big)\;. \end{aligned}$$

By Assumption 7, $\sum_{i=d_n}^{\infty}[\alpha(i)]^{1-2/\delta} \to 0$ as $d_n \to \infty$. So,

$$\begin{aligned} J_2 \;&\leq\; C\,h^{2/\delta-2}\,\sum_{i=d_n}^{\infty}\big[\alpha(i)\big]^{1-2/\delta} + o(1) \\ &\leq\; C\,h^{2/\delta-2}\,d_n^{-\beta}\,\sum_{i=d_n}^{\infty} i^{\beta}\big[\alpha(i)\big]^{1-2/\delta} + o(1)\;. \end{aligned}$$

By choosing d_n such that $h^{2/\delta-1}d_n^{-\beta} \to 1$, we see $J_2 = o(1/h)$ and $hd_n \to 0$, which ensures the convergence of J_1. This completes the proof of (b). The proof of (c) is an immediate consequence of (a) and (b). □

Proof of Lemma 5.2: The proof employs a small-block large-block argument. The set $\{1, ..., n\}$ is partitioned into $2k+1$ alternating large and small subsets. Let r_n be the size of the large blocks and s_n be the size of the small blocks. Then $k_n = \lfloor n/(r_n + s_n)\rfloor$. For $0 \leq j \leq k-1$, define

$$U_j \;:=\; \sqrt{h}\sum_{i=j(r+s)+1}^{j(r+s)+r} Z_i\;,$$

$$V_j \;:=\; \sqrt{h}\sum_{i=j(r+s)+r+1}^{(j+1)(r+s)} Z_i$$

and

$$W_j := \sqrt{h} \sum_{i=k(r+s)+1}^{n} Z_i \ .$$

We see immediately that U_j sums the Z_i over the blocks of size r, V_j sums the Z_i over the blocks of size s, and W_j accounts for the remaining terms that do not fit evenly into the first $2k$ blocks.

The idea of the proof is to show that the small blocks separate the large blocks by enough to make them asymptotically independent while being small enough that they don't make a substantial contribution to the limiting distribution. The Lindeberg condition can then be checked for the separated large blocks.

To formalize this, we write

$$\begin{aligned} \sqrt{n\,h}\, Q_n &= \frac{1}{\sqrt{n}} \left[\sum_{j=0}^{k-1} U_j + \sum_{j=0}^{k-1} V_j + W_j \right] \\ &:= \frac{1}{\sqrt{n}} \left[Q'_n + Q''_n + Q'''_n \right] . \end{aligned}$$

We will establish the following identities:

$$\frac{1}{n} E\big[(Q''_n)^2\big] \to 0 \ , \tag{7.2}$$

$$\frac{1}{n} E\big[(Q'''_n)^2\big] \to 0 \ , \tag{7.3}$$

$$\left| E\, \exp(it\, Q'_n) - \prod_{j=1}^{k} E\big[\exp(it\, U_j)\big] \right| \to 0 \ , \tag{7.4}$$

$$\frac{1}{n} \sum_{i=1}^{k} E\big[U_i^2\big] \to \theta^2(x) \ , \tag{7.5}$$

and

$$\frac{1}{n} \sum_{j=0}^{k-1} E\Big[U_j^2\, 1_{\left[|U_j| \geq \epsilon\, \theta(x) \sqrt{n}\right]}\Big] \to 0 \ , \tag{7.6}$$

for all $\epsilon > 0$.

Once these have been established, by a Taylor Series expansion, we will have

$$\begin{aligned} &\left| E\Big[\exp\big(it\sqrt{n\,h}\, Q_n\big)\Big] - \exp\big(-t^2\theta^2(x)/2\big) \right| = \\ &\qquad = \left| E\left[\exp\left(it\, \frac{1}{\sqrt{n}} \left[Q'_n + Q''_n + Q'''_n\right]\right)\right] - \exp\big(-t^2\theta^2(x)/2\big) \right| \\ &\qquad \leq \left| E\left[\exp\left(it\, \frac{1}{\sqrt{n}}\, Q'_n\right)\right] - \exp\big(-t^2\theta^2(x)/2\big) \right| \\ &\qquad\quad + E\left[\left|\frac{2t}{\sqrt{n}}\, Q''_n\right|\right] + E\left[\left|\frac{2t}{\sqrt{n}}\, Q'''_n\right|\right] . \end{aligned} \tag{7.7}$$

The final two terms will converge to 0 by the Cauchy–Schwarz inequality and equations (7.2) and (7.3). Equations (7.4), (7.5), and (7.6) are enough to verify the conditions of the Lindeberg–Feller central limit theorem, which will establish the desired result. The proof will be complicated somewhat because (7.6) will be established first for bounded random variables, and then the bound will be allowed to tend to infinity.

We begin by choosing block sizes. By Assumption 8, there exists a sequence q_n such that $q_n \to \infty$ and $q_n s_n = o\bigl(\sqrt{n\,h}\bigr)$, and $q_n \sqrt{n/h}\,\alpha(s_n) \to 0$. Define the large block size r_n by

$$r_n := \bigl\lfloor \sqrt{n\,h}/q_n \bigr\rfloor \ .$$

From this definition, we see

$$\frac{s_n}{r_n} \le \frac{s_n}{\sqrt{n\,h}/q_n - 1} = \frac{q_n s_n/\sqrt{n\,h}}{1 - q_n/\sqrt{n\,h}} \to 0 \ , \tag{7.8}$$

$$\begin{aligned} \frac{n}{r_n}\,\alpha(s_n) &= \frac{n}{\lfloor \sqrt{n\,h}/q_n \rfloor}\,\alpha(s_n) \le \frac{n}{\sqrt{n\,h}/q_n - 1}\,\alpha(s_n) \\ &= \Bigl(\sqrt{n/h}\,q_n + o(1)\Bigr)\,\alpha(s_n) \to 0 \end{aligned} \tag{7.9}$$

and

$$\frac{r_n}{\sqrt{n\,h}} = \frac{\lfloor \sqrt{n\,h}/q_n \rfloor}{\sqrt{n\,h}} \le \frac{\sqrt{n\,h}/q_n + 1}{\sqrt{n\,h}} = \frac{1}{q_n} + \frac{1}{\sqrt{n\,h}} \to 0 \ . \tag{7.10}$$

We begin by verifying equations (7.2) and (7.3):

$$E\bigl[(Q_n'')^2\bigr] = \sum_{i=0}^{k-1} \mathrm{var}[V_j] + \sum_{i \ne j} \mathrm{cov}[V_i, V_j] \ , \tag{7.11}$$

where, by Lemma 5.1,

$$\begin{aligned} \mathrm{var}[V_i] &= s\,h\,\mathrm{var}[Z_1] + 2\,s\,h \sum_{i=1}^{s-1} (1 - j/s)\,\mathrm{cov}[Z_1, Z_{1+i}] \\ &= s\,\bigl[\theta(x) + o(1)\bigr] \ . \end{aligned}$$

By equation (7.8),

$$\begin{aligned} \sum_{i=0}^{k-1} \mathrm{var}[V_j] &\le k_n s_n \bigl[\theta(x) + o(1)\bigr] \\ &\le \frac{n\,s_n}{s_n + r_n} \bigl[\theta(x) + o(1)\bigr] \\ &= o(n) \ . \end{aligned}$$

The second term of (7.11) can be treated as follows:

$$\sum_{i \ne j} \mathrm{cov}[V_i, V_j] = h \sum_{i \ne j}^{k-1} \sum_{l=1}^{s} \sum_{m=1}^{s} \mathrm{cov}\bigl[Z_{i(r+s)+r+m}, Z_{j(r+s)+r+l}\bigr] \ .$$

Since $i \neq j$, the difference between the indices, $\big|i(r+s)+r+m-(j(r+s)+r+l)\big| \geq r$, so

$$\begin{aligned}\left|\sum_{i\neq j} \operatorname{cov}[V_i, V_j]\right| &\leq 2\,h \sum_{l=1}^{n-r} \sum_{m=l+r}^{n} \big|\operatorname{cov}[Z_l, Z_m]\big| \\ &= 2\,h \sum_{l=1}^{n-r} \sum_{j=r}^{n-l} \big|\operatorname{cov}[Z_l, Z_{l+j}]\big| \\ &\leq 2\,n h \sum_{j=r}^{n-1} \big|\operatorname{cov}[Z_1, Z_{j+1}]\big| \;=\; o(n)\;.\end{aligned}$$

This establishes (7.2). We now turn our attention to (7.3):

$$\begin{aligned}\frac{1}{n}\,E\big[(Q_n''')^2\big] &= \frac{h}{n}\Big(n - k(r+s)\Big)\operatorname{var}[Z_1] \;+\; 2\,h \sum_{i=2}^{n-k(r+s)} \operatorname{cov}[Z_1, Z_i] \\ &\leq \frac{r_n + s_n}{n}\,\theta^2(x) + o(1) \;\to\; 0\;.\end{aligned}$$

To prove (7.4), we use a lemma of Volkonskii and Rozanov [23], which is stated in Lemma 7.3, following this proof:

$$\begin{aligned}\left|E\exp(i\,t\,Q_n') - \prod_{j=1}^{k} E\big[\exp(i\,t\,U_j)\big]\right| &\leq 16\,(k-1)\;\alpha(s_n+1) \\ &= 16\,\frac{n}{r_n}\;\alpha(s_n+1) + o(1) \;\to\; 0\;,\end{aligned}$$

by equation (7.9).

We now turn our attention to equation (7.5):

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{k} E[U_i^2] &= \frac{k_n}{n}\operatorname{var}[U_1] \\ &= \frac{k_n\, r_n}{n}\,\big(\theta^2(x) + o(1)\big) \\ &= \frac{r_n}{r_n + s_n}\,\theta^2(x) + o(1) \;\to\; \theta^2(x)\;.\end{aligned}$$

Finally, we verify equation (7.6). We begin establishing the result for truncated random variables, and then subsequently letting the truncation point go to infinity. Define

$$\begin{aligned}Z_i^L &:= (c_1 + c_2 Y_i)\;1_{[|Y_i|\leq L]}\,K_h(X_i - x) \;-\; E\Big[(c_1 + c_2 Y_i)\;1_{[|Y_i|\leq L]}\,K_h(X_i - x)\Big]\;, \\ Q_n^L &:= \frac{1}{n}\sum_{i=1}^{n} Z_i^L\;, \\ \tilde{Q}_n^L &:= \frac{1}{n}\sum_{i=1}^{n}\big(Z_i - Z_i^L\big)\;,\end{aligned}$$

and

$$U_j^L \;:=\; \sqrt{h} \sum_{i=j(r+s)+1}^{j(r+s)+r} Z_i^L .$$

We first need to estimate the asymptotic variance of Z_1^L. Assume conditions strong enough that for all c_1 and c_2, and for all $L > L_0$,

$$E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1 = x\Big]$$

is continuous as a function of x:

$$\Big|E\Big[(c_1 + c_2\, Y_i)\; 1_{[|Y_i| \le L]}\, K_h(X_i - x)\Big]\Big| \;\le\; \Big(|c_1| + |c_2\, L|\Big)\, E\Big[\big|K_h(X_1 - x)\big|\Big] \;\le\; C \;,$$

and

$$\begin{aligned}
h\, E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]}\, K_h^2(X_i - x)\Big] \;=& \\
&= \; h\, E\Big[K_h^2(X_1 - x)\; E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1\Big]\Big] \\
&= \; h \int_{-\infty}^{\infty} K_h^2(u - x)\; E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1 = u\Big]\, f(u)\; du \\
&= \; \int_{-\infty}^{\infty} K^2(v)\; E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1 = x + h\, v\Big]\, f(x + h\, v)\; dv \\
&= \; E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1 = x\Big]\; f(x) \int_{-\infty}^{\infty} K^2(v)\, dv \;+\; o(1) \;.
\end{aligned}$$

Putting these two together,

$$h\, \mathrm{var}[Z_1^L] \;=\; E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1 = x\Big]\; f(x) \int_{-\infty}^{\infty} K^2(v)\, dv \;+\; o(1) \;.$$

For the sake of notational simplicity, we now define

$$(\theta^L)^2(x) \;:=\; E\Big[(c_1 + c_2\, Y_i)^2\; 1_{[|Y_i| \le L]} \,\Big|\, X_1 = x\Big]\; f(x) \int_{-\infty}^{\infty} K^2(v)\, dv \;.$$

Returning now to the proof of (7.6), since K and Y_i^L are bounded, hZ_i^L is also bounded. Equivalently, for some D,

$$\sqrt{h}\; Z_i^L \;\le\; D/\sqrt{h} \;.$$

Therefore, by equation (7.10), $\max_{0 \le j \le k-1} U_j^L/\sqrt{n} \le (D r_n)/\sqrt{n\,h} \to 0$. For large enough n, the set $\big\{|U_j^L| \ge \theta^L(x)\, \epsilon\, \sqrt{n}\big\}$ becomes empty. Therefore, by the same argument as used to establish (7.7),

$$\sqrt{n\,h}\; Q_n^L \;\xrightarrow{\;\mathcal{D}\;}\; N\big(0, (\theta^L)^2(x)\big) \;. \tag{7.12}$$

We are now prepared to put these pieces together to finish the proof of the theorem:

$$\begin{aligned}
&\left| E\Big[\exp\big(it\sqrt{nh}\,Q_n\big)\Big] - \exp\big(-t^2\theta^2(x)/2\big)\right| = \\
&\quad = \left| E\Big[\exp\Big(it\sqrt{nh}\,\big[Q_n^L + \tilde{Q}_n^L\big]\Big)\Big] - \exp\big(-t^2(\theta^L)^2(x)/2\big)\right. \\
&\qquad \left. + \exp\big(-t^2(\theta^L)^2(x)/2\big) - \exp\big(-t^2\theta^2(x)/2\big)\right| \\
&\quad = \left| E\Big[\exp\big(it\sqrt{nh}\,Q_n^L\big) + \exp\big(it\sqrt{nh}\,Q_n^L\big)\Big(\exp\big[it\sqrt{nh}\,\tilde{Q}_n^L\big] - 1\Big)\Big]\right. \\
&\qquad \left. - \exp\big(-t^2(\theta^L)^2(x)/2\big) + \exp\big(-t^2(\theta^L)^2(x)/2\big) - \exp\big(-t^2\theta^2(x)/2\big)\right| \\
&\quad \le \left| E\Big[\exp\big(it\sqrt{nh}\,Q_n^L\big)\Big] - \exp\big(-t^2(\theta^L)^2(x)/2\big)\right| \\
&\qquad + \left| E\Big[\exp\big(it\sqrt{nh}\,Q_n^L\big)\Big(\exp\big[it\sqrt{nh}\,\tilde{Q}_n^L\big] - 1\Big)\Big]\right| \\
&\qquad + \left|\exp\big(-t^2(\theta^L)^2(x)/2\big) - \exp\big(-t^2\theta^2(x)/2\big)\right| \\
&\quad \le \left| E\Big[\exp\big(it\sqrt{nh}\,Q_n^L\big)\Big] - \exp\big(-t^2(\theta^L)^2(x)/2\big)\right| \\
&\qquad + E\Big[\left|\exp\big[it\sqrt{nh}\,\tilde{Q}_n^L\big] - 1\right|\Big] + \left|\exp\big(-t^2(\theta^L)^2(x)/2\big) - \exp\big(-t^2\theta^2(x)/2\big)\right|.
\end{aligned}$$

We analyze each term separately, first letting $n \to \infty$, and then letting $L \to \infty$. For fixed t, the first term goes to zero by equation (7.12). The third term goes to 0 by dominated convergence, since $(\theta^L)^2(x) \to \theta^2(x)$ as $L \to \infty$. Only the second term remains to be analyzed. By a Taylor series expansion,

$$\left|\exp\big[it\sqrt{nh}\,\tilde{Q}_n^L\big] - 1\right| \;\le\; 2\left|t\sqrt{nh}\,\tilde{Q}_n^L\right|.$$

By the Cauchy–Schwarz inequality, the bound will converge to 0 if it can be shown that $nh\,\mathrm{var}\big[(\tilde{Q}_N^L)\big] \to 0$. As $\tilde{Q}_n^L$ satisfies the same dependence assumptions as Q_n, the calculations of Lemma 5.1(c) apply. So, it is sufficient to show that $(\theta^L)^2(x) \to 0$ as $L \to \infty$. This follows immediately by dominated convergence. □

Lemma 7.3 (Volkonskii and Rozanov [23]). *Let $V_1, ..., V_N$ be strong mixing random variables, which are measurable with respect to the σ-algebras $\mathcal{F}_{i_1}^{j_1}, ..., \mathcal{F}_{i_N}^{j_N}$ respectively, with $1 \le i_1 < j_1 < i_2 < ... < j_N \le n$, $i_{l+1} - j_l \ge w \ge 1$, and $|V_l| \le 1$, for $l = 1, ..., N$. Then*

$$\left| E\left[\prod_{j=1}^{N} V_j\right] - \prod_{j=1}^{N} E[V_j]\right| \;\le\; 16\,(L-1)\,\alpha(w)\,,$$

where $\alpha(w)$ is the strong mixing coefficient.

REFERENCES

[1] Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, "Lecture Notes in Statistics", vol. 110, Springer-Verlag, New York, second edition.

[2] Davis, K. (1975). Mean square error properties of density estimates, *Annals of Statistics*, **3**(4), 1025–1030.

[3] Davis, K. (1977). Mean integrated square error properties of density estimates, *Annals of Statistics*, **5**(3), 530–535.

[4] Devroye, L. (1992). A note on the usefulness of superkernels in density estimates, *Annals of Statistics*, **20**(4), 2037–2056.

[5] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L^1 View*, Wiley, New York.

[6] Fan, J. (1992). Design-adaptive nonparametric regression, *Journal of the American Statistical Association*, **87**, 998–1004.

[7] Gasser, Th. and Müller, H.-G. (1979). *Kernel estimation of regression functions*. In "Smoothing Techniques for Curve Estimation" (Th. Gasser and M. Rosenblatt, Eds.), Springer Lecture Notes in Mathematics, number 757, Springer-Verlag, Berlin, 23–68.

[8] Gasser, Th. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scandinavian Journal of Statistics*, **11**, 171–185.

[9] Györfi, L.; Härdle, W.; Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, "Lecture Notes in Statistics", vol. 60, Springer-Verlag, New York.

[10] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Applications*, Academic Press, New York.

[11] Hall, P. and Marron, J.S. (1988). Choice of kernel order in density estimation, *Annals of Statistics*, **16**(1), 161–173.

[12] Ibragimov, I.A. and Khasminksii, R.Z. (1983). Estimation of distribution density belonging to a class of entire functions, *Theor. Probab. Appl.*, **27**(3), 551–562.

[13] Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes, *Scandinavian Journal of Statistics*, **24**, 165–179.

[14] McMurry, T. and Politis, D.N. (2004). Nonparametric regression with infinite order flat-top kernels, *Journal of Nonparametric Statistics*, **16**(3–4), 549–562.

[15] Nadaraya, E.A. (1964). On estimating regression, *Theory Probab. Appl.*, **9**, 141–142.

[16] Politis, D.N. (2001). *On nonparametric function estimation with infinite order flat-top kernels*. In "Probability and Statistical Models with Applications" (Ch.A. Charalambides, Markos V. Koutras, and N. Balakrishnan, Eds.), Chapman & Hall/CRC, Boca Raton, 469–483.

[17] Politis, D.N. (2003). Adaptive bandwidth choice, *Journal of Nonparametric Statistics*, **25**(4–5), 517–533.

[18] Politis, D.N. and Romano, J.P. (1993). On a family of smoothing kernels of infinite order. In "Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface" (M. Tarter and M. Lock, Eds.), The Interface Foundation of North America, 141–145.

[19] Politis, D.N. and Romano, J.P. (1995). Bias-corrected nonparametric spectral estimation, *Journal of Time Series Analysis*, **16**(1), 67–103.

[20] Politis, D.N. and Romano, J.P. (1999). Multivariate density estimation with general flat-top kernels of infinite order, *Journal of Multivariate Analysis*, **68**, 1–25.

[21] Robinson, P.M. (1983). Nonparametric estimators for time series, *Journal of Time Series Analysis*, **4**(3), 185–207.

[22] Ruppert, D.; Sheather, S.J. and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association*, **90**, 1257–1270.

[23] Volkonski, V.A. and Rozanov, Yu.A. (1959). Some limit theorems for random functions, *Theory of Probability and its Applications*, **4**, 178–197.

[24] Wand, M. and Ripley, B. (2005). *KernSmooth: Functions for kernel smoothing for Wand and Jones (1995)*, S original by Matt Wand, R port by Brian Ripley, R package version 2.22-15.

[25] Watson, G.S. (1964). Smooth regression analysis, *Sankhya Ser. A*, **26**, 359–372.

REVSTAT – Statistical Journal
Volume 6, Number 2, June 2008, 151–164

ON THE EXTREMES OF RANDOMLY SUB-SAMPLED TIME SERIES

Authors: Andreia Hall
– Departamento de Matemática, UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal
andreia.hall@ua.pt

Manuel G. Scotto
– Departamento de Matemática, UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal
mscotto@ua.pt

Received: May 2007 Revised: November 2007 Accepted: December 2007

Abstract:

- In this paper, we investigate the extremal properties of randomly sub-sampled stationary sequences. Motivation comes from the need to account for the effect of missing values on the analysis of time series and the comparison of schemes for monitoring systems with breakdowns or systems with automatic replacement of devices in case of failures.

Key-Words:

- *extreme value theory; integer-valued stationary sequences; sub-sampling; failure; extremal index.*

AMS Subject Classification:

- 62-02, 60G70.

1. INTRODUCTION

The extremal properties of sub-sampling stationary sequences is a rapidly developing subject and it has been a topic of active research over the last years, mainly due to its wide applicability to the analysis of environmental and financial processes. Sub-sampling may occur according to some deterministic pattern, or may occur randomly. Much of the early work on this topic paid attention on the effect of deterministic sub-sampling on the extremal properties of stationary sequences; see Scotto [22], Hall *et al.* [14], Martins and Ferreira [16], Ferreira and Matins [9], Scotto *et al.* [23], Hall and Scotto [13], Scotto and Ferreira [24], Scotto and Turkman [25] and Robinson and Tawn [21]. In contrast, the effect of random sub-sampling has not received much attention in the literature. We refer to the work of Weissman and Cohen [28] who considered the case of i.i.d. random sub-sampling as a particular case of some mixture models. More recently, Hall and Hüsler [12] have obtained some generalizations of Weissman and Cohen's results for sequences where the sub-sampling pattern has a weak dependence structure.

One reason for the interest in extremes observed at random sampling rates comes from the need to compare schemes for monitoring systems with breakdowns or systems with automatic replacement of devices in case of failures. Examples are encountered, for instance, in ocean engineering. The probabilistic description of the wave climate in specific sites and ocean areas is an important prerequisite for the design and assessment of coastal and offshore structures. The wave climate is commonly described from time series of sea-state parameters, such as the significant wave height and the mean zero upcrossing period. These, as well as other sea-state parameters, provide information about the sea-state that has occurred and about the way the sea-state evolves with time. Most of the early available data has been collected by *waverider* buoys (at present, however, satellite data is becoming widely available and some climate descriptions are based on this type of data). An important aspect for a correct probabilistic description of the wave climate is to work with complete records of wave measurements. Missing values, however, are frequently encountered in time series analysis of wave measurements, mainly when *waverider* buoys are used for collecting data sets. The main reasons are damage by shipping, *freak* waves which appeared out of a calm sea and a failure on the reading device. Similar problems arise in environmental studies. For example, extreme value analysis is of particular interest in assessing the impact of high air pollution levels, because air quality guidelines are formulated in terms of the high level of permitted emissions. This methodology has been used in the analysis of levels of ozone (Smith, [26], Nui, [18], and Tobias and Scotto, [27]) and nitrogen dioxide (Coles and Pan, [7]). Ozone data is usually collected from sampling stations integrated within a local automatic network for the control of atmospheric pollution in a specific area. In this case, missing observations appear when the equipment is not working properly or it is out of service.

As the title of the paper suggest, the aim of this work is to extend the results known for deterministic sub-sampled processes to random-generated sub-sampling processes. In particular, we investigate the maximum limiting distribution and its corresponding extremal index, when the underlying process is represented as a moving average driven by heavy-tailed innovations and the sub-sampling process is strongly mixing. Our results both exemplify some of the findings of Hall and Hüsler [12] and offer more precise details for this particular class of models.

The examples given in the previous paragraphs illustrate the need to account for non-i.i.d. patterns of missing-values since, in general, when an equipment is out of order its recovery time may be considerably long. In this paper we also pay special attention to discrete-valued sequences. Motivation to include discrete data models comes from the need to account for the discrete nature of certain data sets, often counts of events, objects or individuals. Examples of applications can be found in the analysis of time series of count data that are generated from stock transactions (Quoreshi, [20]), where each transaction refers to a trade between a buyer and a seller in a volume of stocks for a given price, and also in experimental biology (Zhou and Basawa, [29]), social science (McCabe and Martin, [17]), international tourism demand (Nordström, [19], Garcia-Ferrer and Queralt, [10], Brännäs *et al.* [4], and Brännäs and Nordström, [5]), and queueing systems (Ahn *et al.* [1]).

The rest of the paper is organized as follows: Section 2 provides a background description of basic theoretical results related to conventional and non-negative integer-valued moving averages with regularly varying tails. Moreover, a suitable representation for the randomly sub-sampled process is described. In Section 3 we obtain the limiting distribution of the maximum term of the sub-sampled moving average sequence and the expression of its extremal index. Finally, in Section 4 the results are applied to conventional and discrete autoregressive processes.

2. PRELIMINARIES

For the purpose of this work we shall consider stationary sequences $\mathbf{X} = (X_n)_{n\in\mathbb{N}_0}$ of the form

$$X_n = \sum_{j=0}^{\infty} \beta_j * Z_{n-j} \ , \tag{2.1}$$

where $\mathbf{Z} = (Z_n)_{n\in\mathbb{Z}}$ is an i.i.d. sequence of random variables (rv's) with distribution function F_Z belonging to the domain of attraction of the Fréchet distribution

with parameter $\alpha > 0$, (hereafter $F_Z \in D(\Phi_\alpha)$):

$$P(|Z_1| > x) = x^{-\alpha} L(x) \ , \qquad x > 0 \ , \tag{2.2}$$

where L is slowly varying at infinity and

$$\lim_{x\to\infty} \frac{P(Z_1 > x)}{P(|Z_1| > x)} = p \ , \qquad \lim_{x\to\infty} \frac{P(Z_1 < -x)}{P(|Z_1| > x)} = q \ , \tag{2.3}$$

for some $p+q=1$ with $0 \le p \le 1$. We further assume that the coefficients $(\beta_j)_{j\in\mathbb{N}_0}$ are such that

$$\sum_{j=0}^{\infty} |\beta_j|^\delta < \infty \ , \qquad \delta < \min(\alpha, 1) \ . \tag{2.4}$$

Throughout the paper we consider two different cases:

(**a**) The $*$-operator denotes multiplication and $\mathbf{Z}$ is an i.i.d. sequence of continuous rv's. In this case $\mathbf{X}$ represents a conventional (i.e., continuous-valued) moving average model.

(**b**) The $*$-operator denotes *binomial thinning*, say $\circ$, and $\mathbf{Z}$ represents an i.i.d. sequence of non-negative integer-valued rv's; that is

$$\beta \circ Z = \sum_{s=1}^{Z} B_s(\beta) \ , \qquad \beta \in [0,1] \ ,$$

where $(B_s(\beta))$ forms an i.i.d. sequence of Bernoulli rv's satisfying $P[B_s(\beta)=1]=\beta$. In this case $\mathbf{X}$ represents a discrete analogue of case (a). It is important to stress the fact that discreteness of the process $\mathbf{X}$ is ensured by the $\circ$-operator since this operator incorporates the discrete nature of the variates and acts as the analogue of the standard multiplication used in the continuous-valued moving average model. Note that thinning is a random operation which reflects the behavior of many natural phenomenons. For instance, if Z_n represents the number of individuals of a certain specie at time n, $\beta \circ Z_n$ will represents the number of survivors at the next time instant with β representing the probability of surviving. The concept of thinning is well known in classical probability theory and has been in use in the Bienaymé–Galton–Watson branching processes literature as well as in the theory of stopped-sum distributions.

We further consider within the discrete case the general class of models consisting of all stationary sequences defined by (2.1) in which all thinning operations involved are independent, for each n. Nevertheless, dependence is allowed to occur between the thinning operators $\beta_j \circ Z_n$ and $\beta_i \circ Z_n$, $j \neq i$ (which belong to X_{n+j} and X_{n+i} respectively). We therefore obtain a rich class of discrete models which share some properties with the conventional case. For particular examples and estimation procedures see Brännäs and Hall [3].

The tail properties of X_n have been studied by Davis and Resnick [8] for the conventional case and by Hall [11] for the discrete case. The result below summarises the tail behavior of the random variables $W = \beta * Z$ and X_n, when $F_Z \in D(\Phi_\alpha)$.

Theorem 2.1. *Let Z be a random variable with $F_Z \in D(\Phi_\alpha)$, $\alpha > 0$.*

1. *For both meanings of the $*$-operator, $F_W \in D(\Phi_\alpha)$ and*

(**a**) *for the conventional case*

$$\lim_{n\to\infty} \frac{1 - F_W(n)}{1 - F_Z(n)} = p(\beta^+)^\alpha + q(\beta^-)^\alpha \ ,$$

with $\beta^+ = \max(\beta, 0)$ and $\beta^- = \max(-\beta, 0)$;

(**b**) *for the discrete case*

$$\lim_{n\to\infty} \frac{1 - F_W(n)}{1 - F_Z(n)} = \beta^\alpha \ .$$

2. *If $F_Z \in D(\Phi_\alpha)$ then, for both meanings of the $*$-operator, $F_X \in D(\Phi_\alpha)$, and for all $\tau > 0$ and some sequence of constants (u_n)*

$$\lim_{n\to\infty} n\big(1 - F_Z(u_n)\big) = \tau' \implies \lim_{n\to\infty} n\big(1 - F_X(u_n)\big) = \tau \ ,$$

with

$$\tau' = \frac{\tau}{\sum_{j=0}^{\infty} p(\beta_j^+)^\alpha + q(\beta_j^-)^\alpha} \ . \tag{2.5}$$

for the conventional case and

$$\tau' = \frac{\tau}{\sum_{j=0}^{\infty} \beta_j^\alpha} \ , \tag{2.6}$$

for the discrete case.

The result above implies that every random variables Z_n contributes to the tail $P(X > x)$. This contribution depends on the size of the weight β_j for both meanings of the $*$-operator, as well as on the sign of the weight β_j in the conventional case.

Now we define the randomly sub-sampled sequence $\mathbf{Y} = (Y_n)_{n\in\mathbb{N}_0}$ obtained from $\mathbf{X}$ and induced through a strictly increasing function $g(n)\colon \mathbb{N}_0 \to \mathbb{N}_0$ as follows:

$$Y_n = X_{g(n)} \ , \qquad n \geq 0 \ .$$

In addition, let $\mathbf{U} = (U_n)_{n\in\mathbb{N}_0}$ be a Bernoulli stationary sequence independent of $\mathbf{X}$ having marginal distribution with parameter γ $(0 \leq \gamma \leq 1)$. The U_n s are used as indicator variables that signal which observations are sampled whereas the $g(\cdot)$

function gives the sampled time, that is the increasing sequence of ns for which $U_n = 1$. As an example take

$$U_1 = 1, \;\; U_2 = 0, \;\; U_3 = 1, \;\; U_4 = 0, \;\; U_5 = 0, \;\; U_6 = 1, \;\; U_7 = 1, \;\; \ldots ,$$

providing

$$g(1) = 1, \;\; g(2) = 3, \;\; g(3) = 6, \;\; g(4) = 7, \;\; \ldots .$$

The sequences **U** considered in this paper will either be i.i.d. or strongly mixing.

The study of the extremal properties of stationary sequences is frequently based on the verification of appropriate dependence conditions which assure that the limiting distribution of the maximum term is of the same type as the limiting distribution of the maximum of i.i.d. rv's with the same marginal distribution F. For stationary sequences, usual conditions used in the literature are Leadbetter's $D(u_n)$ condition (Leadbetter *et al.* [15]) and condition $D^{(k)}(u_n)$, $k \in \mathbb{N}$, (Chernick *et al.* [6]). For completeness and reader's convenience the definition of conditions $D(u_n)$ and $D^{(k)}(u_n)$ are given below.

Definition 2.1. The condition $D(u_n)$ is said to hold for a stationary sequence $(X_n)_{n\in\mathbb{N}}$ with marginal distribution F, if for any integers $i_1 < \ldots < i_p < j_1 < \ldots < j_q < n$ such that $j_1 - i_p \geq l_n$ we have

$$\Big| F_{i_1,\ldots,i_p,j_1,\ldots,j_q}(u_n, \ldots, u_n) - F_{i_1,\ldots,i_p}(u_n, \ldots, u_n)\, F_{j_1,\ldots,j_q}(u_n, \ldots, u_n) \Big| \leq \alpha_{n,l_n}$$

with $\alpha_{n,l_n} \to 0$ for some sequence (l_n), $l_n = o(n)$.

Definition 2.2. The condition $D^{(k)}(u_n)$, $k \geq 1$, holds for a stationary sequence $(X_n)_{n\in\mathbb{N}}$ if there exist sequences (s_n) and (l_n) of integers, and (u_n) of reals, such that $s_n \to \infty$, $s_n\, \alpha_{n,l_n} \to 0$, $\frac{s_n\, l_n}{n} \to 0$, and

$$\lim_{n\to\infty} n\, P\Big(X_1 > u_n \geq M_{2,k}\,,\; M_{k+1,r_n} > u_n \Big) = 0 \;, \tag{2.7}$$

where $r_n = \left[\frac{n}{s_n} \right]$ and

$$M_{i,j} = \begin{cases} -\infty & i > j\,, \\ \max_{i\leq t\leq j} X_t & i \leq j\;. \end{cases}$$

The main result is due to Chernick *et al.* [6], in which the extremal index is computed by knowledge of the joint distribution of k consecutive terms.

Theorem 2.2 (Chernick *et al.* [6])**.** *Suppose that for some $k \geq 1$ the conditions $D(u_n)$ and $D^{(k)}(u_n)$ hold for $u_n = u_n(\tau)$, $\forall\, \tau > 0$. Then, the extremal index of $(X_n)_{n\in\mathbb{N}}$ exists and is equal to θ iff*

$$P\big(M_{2,k} \leq u_n \,|\, X_1 > u_n \big) \to \theta, \qquad \text{as} \;\; n \to \infty, \;\; \forall\, \tau > 0\;.$$

A convenient way to apply the above result may be through the following:

Theorem 2.3 (Chernick *et al.* [6]). *Suppose* $(X_n)_{n\in\mathbb{N}}$ *and* $(X_n^{(m)})_{n\in\mathbb{N}}, m\geq 1$, *are stationary sequences defined on the same probability space such that for some sequence of constants* $\{u_n\}$

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty} n P\Big((1-\epsilon)u_n < X_1 \leq (1+\epsilon)u_n\Big) = 0 ,$$

$$\lim_{m\to\infty}\limsup_{n\to\infty} n P\Big(|X_1 - X_1^{(m)}| > \epsilon u_n\Big) = 0 , \qquad \epsilon > 0 .$$

Then

1. *If condition* $D(u_n)$ *holds for* $(X_n^{(m)})_{n\in\mathbb{N}}$, *for each* m, *then it holds for* $(X_n)_{n\in\mathbb{N}}$ *as well.*
2. *If* $(X_n^{(m)})_{n\in\mathbb{N}}$ *has extremal index* $\theta^{(m)}$, $(X_k)_{n\in\mathbb{N}}$ *has extremal index* θ *iff*
$$\lim_{m\to\infty}\theta^{(m)} = \theta .$$

3. EXTREMAL BEHAVIOR

The main task of this section is to derive the extremal behavior of the sub-sampled $\mathbf{Y}$ process. The main result is formalized through the following theorem.

Theorem 3.1. *Let* $\mathbf{X}$ *be a moving average process defined as previously. Assume that* $(|\beta_j|)_{j\geq 0}$ *forms a decreasing sequence. Consider the sub-sampled sequence* $\mathbf{Y}$ *obtained by random sub-sampling according to an auxiliary stationary sequence* $\mathbf{U}$. *Furthermore, assume that* $F_Z \in D(\Phi_\alpha)$ *satisfying* $\lim_{n\to\infty} n(1-F_Z(u_n)) = \tau'$ *with* τ' *defined as in* (2.5) *for the conventional case, and defined as in* (2.6) *for the discrete case. Then, the distribution of* Y_k *satisfies*

$$\lim_{n\to\infty} n\big(1-F_Y(u_n)\big) = \tau' ,$$

and it holds that:

1. *The sequence* $\mathbf{Y}$ *has extremal index*

$$\theta_C = \frac{\sum_{j=1}^{\infty} P\Big(g(2)-g(1)=j\Big)\Big(\sum_{i=0}^{j-1} p(\beta_i^+)^\alpha + q(\beta_i^-)^\alpha\Big)}{\sum_{j=0}^{\infty} p(\beta_j^+)^\alpha + q(\beta_j^-)^\alpha} , \tag{3.1}$$

for the conventional case, with β_j^+ *and* β_j^- *defined as in Theorem 2.1, and*

$$\theta_D = \frac{\sum_{j=1}^{\infty} P\Big(g(2)-g(1)=j\Big)\ \sum_{i=0}^{j-1}\beta_i^\alpha}{\sum_{j=0}^{\infty}\beta_j^\alpha} , \tag{3.2}$$

for the discrete case.

2. *Moreover the limiting distribution of the maximum* $M_n(Y) = \max\limits_{1\le g(k)\le n}\{Y_k\}$ *is given by*

$$\lim_{n\to\infty} P\big(M_n(Y)\le u_n\big) \;=\; \exp\{-\theta^* x^{-\alpha}\}\;,$$

where θ^* *equals* θ_C *for the conventional case and* θ_D *for the discrete case.*

Proof: By Theorem 2.3, to prove (3.1) we first obtain the extremal index of the the auxiliary finite-order sub-sampled moving average sequence

$$Y_k^{(m)} = \sum_{j=0}^{m} \beta_j * Z_{g(k)-j}\;,$$

for fixed $m>0$. We also temporarily take $\beta_j = 0$ for $j>m$. Note that the local dependence $D^{(m+1)}(u_n)$ condition trivially holds for $\mathbf{Y}^{(m)} = (Y_k^{(m)})$. For simplicity in notation we define $M_{2,m+1}^{(m)} = \max_{2\le k\le m+1} Y_k^{(m)}$, and

$$\mu_{m+1}^{(m)}(u_n) \;=\; P\Big(Y_1^{(m)} > u_n \ge M_{2,m+1}^{(m)}\Big)\,.$$

By Theorem 2.2 we have that the extremal index of the sequence $\mathbf{Y}^{(m)}$, for both meanings of the $*$-operator, is given by

$$\theta^{(m)} \;=\; \lim_{n\to\infty} \frac{n\,\mu_{m+1}^{(m)}(u_n)}{n\,P\big(Y_1^{(m)}>u_n\big)}\;.$$

Moreover by arguments as in Chernick *et al.* ([6], Prop. 2.1)

$$\begin{aligned}
\lim_{n\to\infty} n\,\mu_{m+1}^{(m)}(u_n) \;&=\; \lim_{n\to\infty} n\sum_{j=0}^{m} P\Big(M_{2,m+1}^{(m)}\le u_n\,,\,\beta_j * Z_{g(1)-j} > u_n\Big)\\
&=\; \lim_{n\to\infty} n\sum_{j=0}^{m}\Big[P\Big(\beta_j * Z_{g(1)-j} > u_n\Big)\\
&\qquad - P\Big(M_{2,m+1}^{(m)} > u_n\,,\,\beta_j * Z_{g(1)-j} > u_n\Big)\Big]\,.
\end{aligned}$$

Now

$$\begin{aligned}
&\lim_{n\to\infty} n\,P\Big(M_{2,m+1}^{(m)} > u_n\,,\,\beta_j * Z_{g(1)-j} > u_n\Big) \;=\\
&\quad=\; \lim_{n\to\infty}\left\{ n\,P\left(M_{2,m+1}^{(m)} > u_n\,,\,\beta_j * Z_{g(1)-j} > u_n\,,\,\bigvee_{\substack{0\le i'\le m\\ 2\le t\le m+1}} \beta_{i'} * Z_{g(t)-i'} > u_n\right)\right.\\
&\qquad\left. +\;\; n\,P\left(M_{2,m+1}^{(m)} > u_n\,,\,\beta_j * Z_{g(1)-j} > u_n\,,\,\bigvee_{\substack{0\le i'\le m\\ 2\le t\le m+1}} \beta_{i'} * Z_{g(t)-i'} \le u_n\right)\right\}\\
&\quad=\; \lim_{n\to\infty} n\,P\left(\beta_j * Z_{g(1)-j} > u_n\,,\,\bigvee_{\substack{0\le i'\le m\\ 2\le t\le m+1}} \beta_{i'} * Z_{g(t)-i'} > u_n\right)\,,
\end{aligned}$$

since as in Chernick *et al.* ([6], p. 842) and with the convention that $\beta_j = 0$ for $j > m$ it follows that

$$\lim_{n\to\infty} n\, P\left(M_{2,m+1}^{(m)} \leq u_n\,,\, \beta_j * Z_{g(1)-j} > u_n\,,\, \bigvee_{\substack{0\leq i'\leq m \\ 2\leq t\leq m+1}} \beta_{i'} * Z_{g(t)-i'} > u_n \right) = 0\ .$$

This makes explicit the precise way in which a single large Z asymptotically dominates the behavior of the maximum of the sequence $\mathbf{Y}^{(m)}$. For the conventional case, it follows that

$$\begin{aligned}
\lim_{n\to\infty} n\,\mu_{m+1}^{(m)}(u_n) &= \lim_{n\to\infty} n \sum_{j=0}^{m} P\left(\beta_j * Z_{g(1)-j} > u_n\,,\, \bigvee_{\substack{0\leq i'\leq m \\ 2\leq t\leq m+1}} \beta_{i'} * Z_{g(t)-i'} \leq u_n \right) \\
&= \lim_{n\to\infty} n \sum_{j=0}^{m} P\left(\beta_j * Z_1 > u_n\,,\, \bigvee_{2\leq t\leq m+1} \beta_{g(t)-g(1)+j} * Z_1 \leq u_n \right) \\
&= \lim_{n\to\infty} n \sum_{j=0}^{m} \Bigg[P\Bigg(\bigvee_{2\leq t\leq m+1} \beta^{+}_{g(t)-g(1)+j} * Z_1 \leq u_n \Bigg) \\
&\quad + P\left(\bigvee_{2\leq t\leq m+1} \beta^{-}_{g(t)-g(1)+j} * Z_1 \leq u_n \right) \\
&\quad - P\left(\bigvee_{1\leq t\leq m+1} \beta^{+}_{g(t)-g(1)+j} * Z_1 \leq u_n \right) \\
&\quad - P\left(\bigvee_{1\leq t\leq m+1} \beta^{-}_{g(t)-g(1)+j} * Z_1 \leq u_n \right) \Bigg] \\
&= \lim_{n\to\infty} n \sum_{j=0}^{m} \Big[P\Big(\beta^{+}_{g(2)-g(1)+j} * Z_1 \leq u_n \Big) \\
&\quad + P\Big(\beta^{-}_{g(2)-g(1)+j} * Z_1 \leq u_n \Big) \\
&\quad - P\Big(\beta_j^{+} * Z_1 \leq u_n \Big) - P\Big(\beta_j^{-} * Z_1 \leq u_n \Big) \Big]\ ,
\end{aligned}$$

since $(|\beta_j|)_{j\in\mathbb{N}_0}$ forms a decreasing sequence with $\beta_j = 0$ for $j \geq m+1$. Conditioning on $V = g(2) - g(1)$ we obtain

$$\lim_{n\to\infty} n\,\mu_{m+1}^{(m)}(u_n) \;=\; \sum_{j=0}^{m} P\Big(g(2) - g(1) = j \Big) \left(\sum_{i=0}^{j-1} p(\beta_i^{+})^{\alpha} + q(\beta_i^{-})^{\alpha} \right).$$

Following Davis and Resnick [8] the tail behavior of $Y_k^{(m)}$ is given as follows:

$$\lim_{n\to\infty} \frac{P\big(Y_k^{(m)} > u_n\big)}{P\big(Z_1 > u_n\big)} \;=\; \sum_{j=0}^{m} p(\beta_j^{+})^{\alpha} + q(\beta_j^{-})^{\alpha}\ ,$$

yielding

$$\theta^{(m)} \;=\; \frac{\sum_{j=1}^{m} P\Big(g(2)-g(1)=j\Big)\,\Big(\sum_{i=0}^{j-1} p(\beta_i^{+})^{\alpha}+q(\beta_i^{-})^{\alpha}\Big)}{\sum_{j=0}^{m} p(\beta_j^{+})^{\alpha}+q(\beta_j^{-})^{\alpha}}\;.$$

Finally as an application of Lemma 3.1 in Hall and Hüsler ([12], p. 547), condition $D(u_n)$ holds for the sub-sampled sequence $\mathbf{Y}$, and hence by Theorem 2.3 the extremal index θ_C is

$$\theta_C \;=\; \lim_{m\to\infty} \theta^{(m)}\;.$$

The discrete case follows as an application of the results given in Hall [11] and Hall *et al.* [14]. □

4. EXAMPLES

We now illustrate the effect of random sub-sampling on the extremal index of an AR(1) process

$$X_k \;=\; \beta * X_{k-1} + Z_k\;,$$

considering two different cases:

(**a**) the conventional case with $\beta \in (-1,0)$ and the sequence of innovations $\mathbf{Z}$ satisfying (2.2) and (2.3);

(**b**) the discrete case with $\mathbf{Z}$ being a sequence of non-negative integer-valued rv's.

This type of autoregressive sequence is known as *INteger-valued AutoRegressive process of order one* (INAR(1) in short) process and has been considered by several authors in the literature; see Aly and Bouzar [2] for details. It is worth noting that in the former case, Hall and Hüsler's results can not be applied since condition $D''(u_n)$ does not hold. In contrast, the AR(1) model with $\beta \in (0,1)$ satisfies $D''(u_n)$ condition.

Furthermore, for the sequence $\mathbf{U}$ two different cases will be considered:

- Independent and identically distributed failure instants: in this case $\mathbf{U}$ forms an i.i.d. sequence with $P(U_k=1)=\gamma=1-P(U_k=0)$, providing

$$P\Big(g(2)-g(1)=j\Big) \;=\; \gamma(1-\gamma)^{j-1}\;, \qquad j=1,2,\ldots\;;$$

- Failures via a Markov Chain: within this framework $\mathbf{U}$ forms a stationary Markov sequence defined by

$$\begin{cases} P\Big(U_k=1 \mid U_{k-1}=1\Big) \;=\; \eta\;, \\ P\Big(U_k=1 \mid U_{k-1}=0\Big) \;=\; \nu\;. \end{cases}$$

This model defines a system where the probability of failure depends only on whether there occurred or not a failure just before. Given any values of $\eta, \nu \in [0,1]$ it is easy to obtain that

$$P(U_1 = 1) = \frac{\nu}{1-\eta+\nu} .$$

Note that for a fixed value of $\kappa = \frac{\nu}{1-\eta+\nu} \in [0,1]$, the parameters ν and η are not entirely arbitrary since if $\kappa > 1/2$ then $\eta \in [2-1/\kappa, 1]$. The sequence $\mathbf{U}$ is regenerative with finite mean duration of renewal epochs and hence it is strongly mixing. Moreover

$$P\Big(g(2)-g(1)=j\Big) = \begin{cases} \eta & j=1 , \\ (1-\nu)^{j-2}(1-\eta)\nu & j \geq 2 . \end{cases}$$

4.1. Conventional case with negative parameter

In this case, the sub-sampled sequence $\mathbf{Y}$ generated through the i.i.d. sequence $\mathbf{U}$ has extremal index

$$\theta_C = \frac{1-\beta^{2\alpha}}{1-(1-\gamma)\beta^{2\alpha}} . \tag{4.1}$$

When $\gamma = 1$, (i.e., no sub-sampling), the extremal index in (4.1) becomes $\theta_C = 1-\beta^{2\alpha}$ which may be derived from the results given in Davis and Resnick [8]. Moreover, if the sub-sampled sequence $\mathbf{Y}$ is generated through the stationary Markov sequence $\mathbf{U}$, the extremal index becomes

$$\theta_C = \frac{1-\beta^{2\alpha}\Big[1-(\nu-\eta)(1-\beta^{2\alpha})\Big]}{1-(1-\nu)\beta^{2\alpha}} .$$

4.2. Discrete case

In the discrete case, the extremal index of the sub-sampled sequence $\mathbf{Y}$ generated through the i.i.d. sequence $\mathbf{U}$, takes the form

$$\theta_D = \frac{1-\beta^{\alpha}}{1-(1-\gamma)\beta^{\alpha}} ;$$

whereas for the stationary Markov sequence, the extremal index is given by

$$\theta_C = \frac{1-\beta^{\alpha}\Big[1-(\nu-\eta)(1-\beta^{\alpha})\Big]}{1-(1-\nu)\beta^{\alpha}} .$$

ACKNOWLEDGMENTS

Research (partially) supported by Unidade de Investigação Matemática e Aplicações of Universidade de Aveiro through Programa Operacional "Ciência, Tecnologia, Inovação" of the Fundação para a Ciência e a Tecnologia (FCT) cofinanced by the European Community fund FEDER. The second author is also supported by PTDC/MAT/64353/2006. The authors would also like to express their gratitude to the Associate Editor and an anonymous referee for all helpful comments.

REFERENCES

[1] Ahn, S.; Gyemin, L. and Jongwoo, J. (2000). Analysis of the M/D/1-type queue based on an integer-valued autoregressive process, *Oper. Res. Lett.*, **27**, 235–241.

[2] Aly, E.-E. and Bouzar, N. (2005). Stationary solutions for integer-valued autoregressive processes, *Int. J. Math. Math. Sci.*, **1**, 1–18.

[3] Brännäs, K. and Hall, A. (2001). Estimation in integer-valued moving average models, *Appl. Stoch. Model Bus.*, **17**, 277–291.

[4] Brännäs, K.; Hellström, and Nordström, J. (2002). A new approach to modelling and forecasting monthly guest nights in hotels, *Int. J. Forecast.*, **18**, 9–30.

[5] Brännäs, K. and Nordström, J. (2006). Tourist accomodation effects of festivals, *Tourism Economics*, **12**, 291–302.

[6] Chernick, M.; Hsing, T. and McCormick, W. (1991). Calculating the extremal index for a class of stationary sequences, *Adv. Appl. Probab.*, **23**, 835–850.

[7] Coles, S.G. and Pan, F. (1996). The analysis of extreme pollution levels: a case study, *J. Appl. Stats.*, **23**, 333–348.

[8] Davis, R.A. and Resnick, S.I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities, *Ann. Probab.*, **13**, 179–195.

[9] Ferreira, H. and Martins, A.P. (2003). The extremal index of sub-sampled periodic sequences with strong local dependence, *REVSTAT*, **1**, 15–24.

[10] Garcia-Ferrer, A. and Queralt, R.A. (1997). A note on forecasting international tourism deman in Spain, *Int. J. Forecast.*, **13**, 539–549.

[11] Hall, A. (2001). Extremes of integer-valued moving averages models with regularly varying tails, *Extremes*, **4**, 219–239.

[12] Hall, A. and Hüsler, J. (2006). Extremes of stationary sequences with failures, *Stoch. Models*, **22**, 537–557.

[13] Hall, A. and Scotto, M.G. (2003). Extremes of sub-sampled integer-valued moving average models with heavy-tailed innovations, *Statist. Probab. Lett.*, **63**, 97–105.

[14] Hall, A.; Scotto, M.G. and Ferreira, H. (2004). On the extremal behaviour of generalised periodic sub-sampled moving average models with regularly varying tails, *Extremes*, **7**, 149–160.

[15] Leadbetter, M.R.; Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

[16] Martins, A.P. and Ferreira, H. (2004). The extremal index of sub-sampled processes, *J. Statist. Plann. Inference*, **1**, 145–152.

[17] McCabe, B.P.M. and Martin, G.M. (2005). Bayesian prediction of low count time series, *Int. J. Forecast.*, **21**, 315–330.

[18] Nui, X.F. (1997). Extreme value for a class of non-stationary time series with applications, *Ann. Appl. Probab.*, **7**, 508–522.

[19] Nordström, J. (1996). Tourism satellite account for Sweden 1992–93, *Tourism Economics*, **2**, 13–42.

[20] Quoreshi, A.M.M.S. (2006). Bivariate time series modelling of financial count data, *Comm. Statist. Theory Methods*, **35**, 1343–1358.

[21] Robinson, M.E. and Tawn, J.A. (2000). Extremal analysis of processes sampled at different frequencies, *J. Roy. Statist. Soc. B*, **62**, 117–135.

[22] Scotto, M. G. (2005). Extremes of a class of deterministic sub-sampled processes with applications to stochastic difference equations, *Stoch. Proc. Appl.*, **115**, 417–434.

[23] Scotto, M.G.; Turkman, K.F. and Anderson, C.W. (2003). Extremes of some sub-sampled time series, *J. Time Ser. Anal.*, **24**, 579–590.

[24] Scotto, M.G. and Ferreira, H. (2003). Extremes of deterministic sub-sampled moving averages with heavy-tailed innovations, *Appl. Stoch. Model Bus.*, **19**, 303–313.

[25] Scotto, M.G. and Turkman, K.F. (2002). On the extremal behavior of sub-sampled solutions of stochastic difference equations, *Portugal. Math.*, **59**, 267–282.

[26] Smith, R.L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, *Statist. Science*, **4**, 367–393.

[27] Tobias, A. and Scotto, M.G. (2005). Prediction of extreme ozone levels in Barcelona, Spain, *Environ. Monit. Assess.*, **100**, 23–32.

[28] Weissman, I. and Cohen, U. (1995). The extremal index and clustering of high values for derived stationary sequences, *J. Appl. Probab.*, **32**, 972–981.

[29] Zhou, J. and Basawa, I.V. (2005). Least-squared estimation for bifurcation autoregressive processes, *Statist. Probab. Lett.*, **74**, 77–88.

REVSTAT – Statistical Journal
Volume 6, Number 2, June 2008, 165–197

INTERVAL ESTIMATORS FOR A BINOMIAL PROPORTION: COMPARISON OF TWENTY METHODS

Authors: Ana M. Pires
– Department of Mathematics and CEMAT, Instituto Superior Técnico, Technical University of Lisbon (TULisbon), Portugal
apires@math.ist.utl.pt

Conceição Amado
– Department of Mathematics and CEMAT, Instituto Superior Técnico, Technical University of Lisbon (TULisbon), Portugal
camado@math.ist.utl.pt

Received: March 2007 Revised: March 2008 Accepted: March 2008

Abstract:

- In applied statistics it is often necessary to obtain an interval estimate for an unknown proportion (p) based on binomial sampling. This topic is considered in almost every introductory course. However, the usual approximation is known to be poor when the true p is close to zero or to one. To identify alternative procedures with better properties twenty non-iterative methods for computing a (central) two-sided interval estimate for p were selected and compared in terms of coverage probability and expected length. From this study a clear classification of those methods has emerged. An important conclusion is that the interval based on asymptotic normality, but after the arcsine transformation and a continuity correction, and the Add 4 method of Agresti and Coull (1998) yield very reliable results, the choice between the two depending on the desired degree of conservativeness.

Key-Words:

- *confidence interval; binomial distribution; proportion test; normal approximation; arcsine transformation; continuity correction; bootstrap; HPD credibility intervals.*

AMS Subject Classification:

- 62F25, 62F40.

1. INTRODUCTION

In many practical situations it is important to compute a two-sided interval estimate for a population proportion (e.g. acceptance sampling by attributes, marketing research, survey sampling). The interval estimate may be either a confidence interval (in the frequentist framework) or a credibility interval (in the Bayesian framework). This is a well known topic considered in almost every introductory course on statistics. However, most of the standard methods rely on asymptotic approximations and the validity of the approximations is not always stated or differs from author to author (Leemis and Trivedi, 1996, give a survey of these "rules of thumb"). Moreover, comparisons between methods are usually based on single cases. We found no recent text book listing the most common methods and making a general comparison, not even in Fleiss, Levin and Paik (2003), dedicated exclusively to rates and proportions. A good discussion but somehow out of date can be found in Santner and Duffy (1989).

Nevertheless, several authors have addressed this subject in the last thirty years: Ghosh (1979); Fujino (1980); Angus and Schafer (1983); Blyth and Still (1983); Blyth (1986); Chen (1990); Copas (1992); Vollset (1993); Cohen and Yang (1994); Newcombe (1994 and 1998); Agresti and Coull (1998); Agresti and Caffo (2000); Brown, Cai and DasGupta (2001 and 2002); Edwardes (1998); Pan (2002); Reiczigel (2003); García-Pérez (2005); Geyer and Meeden (2005), Puza and O'neill (2006) and Lee (2006). Large comparative studies were presented by Vollset (1993), Newcombe (1998), Brown, Cai and DasGupta (2001) and, to a smaller extent, by Agresti and Coull (1998) and Pan (2002).

The present paper considers twenty simple non-iterative methods. Eleven of these have been included in the aforementioned comparisons: eight of the seventeen considered by Vollset (1993); five of the seven considered by Newcombe (1998); the four considered by Agresti and Coull (1998); five of the twelve considered by Brown, Cai and DasGupta (2001) and the four considered by Pan (2002). The nine (also simple) methods considered here and which were not included in previous comparisons are: a Bayesian interval with uniform prior; two simple corrections to the usual interval; four variants of bootstrap intervals not needing Monte Carlo; and two intervals based on the arcsine transformation followed by a correction. As with most of the cited recent studies it was decided not to include methods without explicit solutions and for which there is an explicit almost equivalent method (this is the case namely of the interval obtained by inverting the likelihood ratio test).

The results of this work are important for the applied statistician, who in a particular situation usually wants to use the best method and that this method is available or can easily be implemented in common statistical software, and for

teachers of statistics who have to decide which methods to include in a given course.

Let $[L;U]$ be an interval estimator of a certain parameter θ and attach to it a level (confidence or credibility), $\eta \in (0;1)$. $[L;U]$ is a good interval estimator of θ with level η if the probability of containing the unknown θ (the coverage probability) is in fact η and its length is "small" (usually in a stochastic sense, for instance, on average). Note, however, that in discrete situations, like the one considered here, it is not possible to achieve the target coverage probability for all possible values of the parameter. We will therefore consider two classes of acceptable methods, those strictly conservative (for which the coverage probability is at least η) and those correct on average (for which the mean coverage probability is at least η) and look for small mean expected length within each class. Note also that it is reasonable to apply the same criteria to both confidence and credibility intervals.

Attention will be focused primarily on central intervals, that is, with approximately equal uncertainty associated to each side. This is how practitioners usually interpret two-sided intervals and matches better with one-sided intervals (the two-sided being the intersection of upper and lower one-sided intervals with the appropriate precision). However non-central intervals are also considered for comparison.

The paper is organized as follows: in Section 2 the twenty selected methods are described. In Section 3 results regarding coverage probability and expected length of the different intervals are presented and analyzed, first considering only the central intervals and at second stage including two optimal non-central intervals. Section 4 is devoted to concluding remarks.

2. DESCRIPTION OF THE METHODS

In order to establish the notation suppose that a random sample of size n is observed on a large (possibly infinite) population and that X observations $(0 \leq X \leq n)$ belong to a certain category of interest. Let p be the unknown proportion of the category of interest in the population and suppose that a two-sided central interval estimate for p is wanted. Note that in order to use the methods based on the binomial distribution the total sample size must be fixed a priori and the variable to be observed is the number of successes. If the sampling plan fixes the number of successes and the total sample size is variable (inverse or negative binomial sampling) most of the methods can not be applied directly (see e.g. Lui, 1995, or Cai and Krishnamoorthy, 2005).

For all the intervals the nominal confidence level is fixed in advance as $100\times(1-\alpha)\%$, meaning that the coverage probability of the random interval $[L;U]$ should be $1-\alpha$. The random variables L and U depend on X, number of successes in the random sample, on n and on the method. Twenty methods are described in the following five subsections and the corresponding solutions are represented as $[L_i(X);U_i(X)]$, where $i=I,...,XX$ denotes the method. The final expressions are given in Tables 1 and 2. As in Vollset (1993) it was imposed that, for all i, $0\le L_i(X)$ and $U_i(X)\le 1$, for all X, and that $L_i(0)=0$ and $U_i(n)=1$ (this means that for the boundary cases the centrality property is dropped but it is a natural choice). To be theoretically correct, but otherwise with no practical effect, the confidence intervals do not include the left (right) end point if $L_i(X)=0$ $(U_i(X)=1)$ but $X\neq 0$ $(X\neq n)$.

2.1. Exact results

Under the previous conditions X has a Binomial(n,p) distribution. Because this is a discrete distribution it is not possible to have a confidence interval with exactly the specified confidence level. But an interval with a coverage probability of at least $1-\alpha$ can be obtained by solving

$$\sum_{j=X}^{n}\binom{n}{j}L_I^j(1-L_I)^{n-j}\;=\;\alpha' \tag{2.1}$$

and

$$\sum_{j=0}^{X}\binom{n}{j}U_I^j(1-U_I)^{n-j}\;=\;\alpha''\;, \tag{2.2}$$

where α' and α'', such that $\alpha'+\alpha''=\alpha$, are fixed in advance and do not depend on X. If different values of α' and α'' are chosen for each X, for instance those minimizing the length of the interval, the exactness, meaning a coverage probability of at least $1-\alpha$ can no longer be guaranteed for every p. $\alpha'=\alpha''=\alpha/2$ corresponds to the inversion of the two sided exact binomial test and leads to the central exact interval, usually called Clopper–Pearson interval (Clopper and Pearson, 1934). For $X=0$ and $X=n$ the solutions of (2.1) and (2.2) are explicit:

$$L_I(0)=0,\quad U_I(0)=1-(\alpha/2)^{1/n},\quad L_I(n)=(\alpha/2)^{1/n},\quad U_I(n)=1\;. \tag{2.3}$$

Otherwise, the solution of (2.1) and (2.2) is easy to obtain by using the relation (see for instance Johnson and Kotz 1969, pp. 58–59, or Stevens, 1950)

$$\sum_{j=k}^{n}\binom{n}{j}p^j(1-p)^{n-j}\;=\int_0^p f_B(t)\,dt\;,$$

where f_B denotes the p.d.f. of a Beta$(k, n-k-1)$ random variable. In this way the extremes of the interval are appropriate quantiles of that Beta distribution (see Table 1) and can easily be obtained in most statistical packages. For instance, in S-plus or R the appropriate commands are `LI<-qbeta(alfa/2,X,n-X+1)` and `UI<-qbeta(1-alfa/2,X+1,n-X)`. If the percentiles of the Beta distribution are not available then the relation of those with the percentiles of the F distribution can be used, eventually in tables (this is again mentioned in Johnson and Kotz, 1969, and e.g. in Armitage and Berry, 1987).

It is worth noting that from the papers referred in the Introduction and addressing the same issue only Blyth (1986), Agresti and Coull (1998), Newcombe (1998), Brown, Cai and DasGupta (2001) and García-Pérez (2005) mention the relation with the Beta distribution while some of the others consider approximations to the percentiles of the F distribution (e.g. Fujino, 1980). Vollset (1993) proposes a very sophisticated numerical method (the "Pratt" approximation) which turns out to be completely unnecessary.

Other exact intervals, in the sense of having coverage probability of at least $1-\alpha$, based on the binomial probabilities have been considered in the literature but are not central and do not have explicit solutions (Sterne, 1954, Crow, 1956, Clunies-Ross, 1958, see also Blyth and Still, 1983, and Reiczigel, 2003). As mentioned in the Introduction, an interval of this type will be considered in Subsection 3.2 for comparative purposes.

As a second alternative in the class of exact methods we consider a Bayesian credibility interval. This method is exact in the second sense because it guarantees a mean coverage probability of $1-\alpha$ under the specified prior distribution for p. If this prior is the uniform distribution in $(0,1)$ or Beta$(1,1)$, which is non-informative, we have that *a posteriori* p follows a Beta$(X+1, n-X+1)$ distribution. Brown, Cai and DasGupta (2001) have chosen the Jeffreys prior, Beta$(1/2, 1/2)$, which is also non-informative. The results shall not differ considerably, but the uniform prior seems more intuitive.

In order to obtain a central interval, equal credibility tails $(\alpha/2)$ are considered, except for the boundary cases, $X=0$ and $X=n$. The explicit results are again shown is Table 1 (Method II). The similarity between these results and the results for the Clopper–Pearson interval is a consequence of the chosen *a priori* distribution.

In the Bayesian framework the optimal exact interval (that is of minimal length) is the HPD interval, but this again is non-central and has no explicit solution, and will be considered only in Subsection 3.2.

Table 1: Explicit limits of the intervals for Methods I to X.

i (Method)	$L_i(X)$ (Lower limit)	$U_i(X)$ (Upper limit)
I [(a)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, $B_{X,n-X+1;\alpha/2}$ if $0<X<n$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, $B_{X+1,n-X;1-\alpha/2}$ if $0<X<n$
II [(a)]	0 if $X=0$, $\alpha^{1/(n+1)}$ if $X=n$, $B_{X+1,n-X+1;\alpha/2}$ if $0<X<n$	$1-\alpha^{1/(n+1)}$ if $X=0$, 1 if $X=n$, $B_{X+1,n-X+1;1-\alpha/2}$ if $0<X<n$
III [(b)]	$\dfrac{2X+c^2-c\sqrt{c^2+4X(1-X/n)}}{2(n+c^2)}$	$\dfrac{2X+c^2+c\sqrt{c^2+4X(1-X/n)}}{2(n+c^2)}$
IV [(b)]	0 if $X=0$, otherwise $\dfrac{2X+c^2-1-c\sqrt{c^2-(2+1/n)+4X(1-X/n+1/n)}}{2(n+c^2)}$	1 if $X=n$, otherwise $\dfrac{2X+c^2+1+c\sqrt{c^2+(2-1/n)+4X(1-X/n-1/n)}}{2(n+c^2)}$
V [(b)]	$\max\left\{\frac{X}{n}-c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\ 0\right\}$	$\min\left\{\frac{X}{n}+c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\ 1\right\}$
VI [(b)]	$\max\left\{\frac{X}{n}-c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}-\frac{1}{2n};\ 0\right\}$	$\min\left\{\frac{X}{n}+c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}+\frac{1}{2n};\ 1\right\}$
VII [(b)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\max\left\{\frac{X}{n}-c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\ 0\right\}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\min\left\{\frac{X}{n}+c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\ 1\right\}$
VIII [(b)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\max\left\{\frac{X}{n}-c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}-\frac{1}{2n};\ 0\right\}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\min\left\{\frac{X}{n}+c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}+\frac{1}{2n};\ 1\right\}$
IX [(b)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\max\left\{\frac{X+c^2/2}{n+c^2}-c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\ 0\right\}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\min\left\{\frac{X+c^2/2}{n+c^2}+c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\ 1\right\}$
X [(b)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\max\left\{\frac{X+c^2/2}{n+c^2}-c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}-\frac{1}{2n};\ 0\right\}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\min\left\{\frac{X+c^2/2}{n+c^2}+c\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}+\frac{1}{2n};\ 1\right\}$

[(a)] $B_{\theta_1,\theta_2;\gamma}$ is the γ percentile of the Beta(θ_1,θ_2) distribution.

[(b)] $c=z_{1-\alpha/2}$ where z_γ is the γ percentile of the $\mathcal{N}(0,1)$ distribution.

Table 2: Explicit limits of the intervals for Methods XI to XX.

i (Method)	$L_i(X)$ (Lower limit)	$U_i(X)$ (Upper limit)
XI [(a)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\dfrac{\mathrm{Bin}_{n,X/n;\alpha/2}}{n}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\dfrac{\mathrm{Bin}_{n,X/n;1-\alpha/2}}{n}$
XII [(a)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\max\left\{\frac{\mathrm{Bin}_{n,X/n;\alpha/2}}{n}-\frac{1}{2n};\,0\right\}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\min\left\{\frac{\mathrm{Bin}_{n,X/n;1-\alpha/2}}{n}+\frac{1}{2n};\,1\right\}$
XIII [(a)(b)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\dfrac{\mathrm{Bin}_{n,X/n;\alpha'}}{n}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\dfrac{\mathrm{Bin}_{n,X/n;\alpha''}}{n}$
XIV [(a)(b)]	0 if $X=0$, $(\alpha/2)^{1/n}$ if $X=n$, otherwise $\max\left\{\frac{\mathrm{Bin}_{n,X/n;\alpha'}}{n}-\frac{1}{2n};\,0\right\}$	$1-(\alpha/2)^{1/n}$ if $X=0$, 1 if $X=n$, otherwise $\min\left\{\frac{\mathrm{Bin}_{n,X/n;\alpha''}}{n}+\frac{1}{2n};\,1\right\}$
XV [(c)]	0 if $X=0$, otherwise $\sin^2\left(\arcsin\sqrt{\frac{X}{n}}-\frac{c}{2\sqrt{n}}\right)$	1 if $X=n$, otherwise $\sin^2\left(\arcsin\sqrt{\frac{X}{n}}+\frac{c}{2\sqrt{n}}\right)$
XVI [(c)]	0 if $X=0$, otherwise $\sin^2\left(\arcsin\sqrt{\frac{X-0.5}{n}}-\frac{c}{2\sqrt{n}}\right)$	1 if $X=n$, otherwise $\sin^2\left(\arcsin\sqrt{\frac{X+0.5}{n}}+\frac{c}{2\sqrt{n}}\right)$
XVII [(c)]	0 if $X=0$, otherwise $\sin^2\left(\arcsin\sqrt{\frac{3/8+X-0.5}{n+3/4}}-\frac{c}{2\sqrt{n+1/2}}\right)$	1 if $X=n$, otherwise $\sin^2\left(\arcsin\sqrt{\frac{3/8+X+0.5}{n+3/4}}+\frac{c}{2\sqrt{n+1/2}}\right)$
XVIII [(c)]	$\max\left\{\frac{X+2}{n+4}-c\sqrt{\frac{X+2}{(n+4)^2}\left(1-\frac{X+2}{n+4}\right)};\,0\right\}$	$\min\left\{\frac{X+2}{n+4}+c\sqrt{\frac{X+2}{(n+4)^2}\left(1-\frac{X+2}{n+4}\right)};\,1\right\}$
XIX [(d)]	$\max\left\{\frac{X}{n}-t'\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\,0\right\}$	$\min\left\{\frac{X}{n}+t'\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)};\,1\right\}$
XX [(d)]	$\max\left\{\frac{X+2}{n+4}-t''\sqrt{\frac{X+2}{(n+4)^2}\left(1-\frac{X+2}{n+4}\right)};\,0\right\}$	$\min\left\{\frac{X+2}{n+4}+t''\sqrt{\frac{X+2}{(n+4)^2}\left(1-\frac{X+2}{n+4}\right)};\,1\right\}$

[(a)] $\mathrm{Bin}_{n,\theta;\gamma}$ is the γ percentile of the $\mathrm{Bin}(n,\theta)$ distribution.

[(b)] α' and α'' are given by equations (2.5) and (2.6) in the text, respectively.

[(c)] $c=z_{1-\alpha/2}$ where z_γ is the γ percentile of the $\mathcal{N}(0,1)$ distribution.

[(d)] t' and t'' are percentiles $(1-\alpha/2)$ of t-distributions with degrees of freedom given by equations (2.8) and (2.9) in the text, respectively.

2.2. Normal approximations

The most referred methods are based on the approximation of the Binomial(n,p) by the $\mathcal{N}\big(np, np(1-p)\big)$ distribution, that is on

$$\frac{X-np}{\sqrt{np\,(1-p)}} = \frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} \mathcal{N}(0,1)$$

and

$$P\left(-z_{1-\alpha/2} \le \frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{1-\alpha/2}\right) \simeq 1-\alpha\ , \tag{2.4}$$

where z_γ denotes the γ percentile of the $\mathcal{N}(0,1)$ distribution.

From (2.4) one can obtain, solving a second degree equation, the score or Wilson (1927) interval (Method III in Table 1). Note that $L_{III}(X) > 0$ and $U_{III}(X) < 1$ for every $0 < X < n$ and that $L_{III}(0) = 0$ and $U_{III}(n) = 1$.

A modification of the score method is obtained by introducing a continuity correction (cc) of $\pm 1/(2n)$ in the numerator of the central expression of (2.4), which is expected to improve the approximation (Method IV in Table 1). For the boundary cases one obtains $L_{IV}(0) > 0$ and $U_{IV}(n) < 1$, which are corrected in the obvious way.

Most of the elementary text books do not present the previous methods, considering instead a second approximation in expression (2.4),

$$P\left(-c \le \frac{\frac{X}{n}-p}{\sqrt{\frac{X}{n^2}\left(1-\frac{X}{n}\right)}} \le c\right) \simeq 1-\alpha\ ,$$

leading to the Wald interval (Method V in Table 1). Method VI is similar but with continuity correction.

Noting that, when $X = 0$ or $X = n$, the Wald interval has zero length, something an applied statistician will be reluctant to present, it is wise to consider replacing it, just in these two cases, by the exact Clopper–Pearson expressions (2.3). This is denoted Method VII (without continuity correction). Method VIII is just the same but with continuity correction. Vollset (1993) also considered these two modifications of the Wald interval.

Another possibility for correcting the Wald interval (mentioned for instance by Brown, Cai and DasGupta, 2002) is to recenter it at the center of the score interval, which is given by $(X + c^2/2)/(n + c^2)$, see Table 1. This modification is considered here together with the previous one, both without and with continuity correction (Methods IX and X, respectively).

2.3. Bootstrap methods

It was considered interesting to include in this study some bootstrap methods because, in this particular situation, it is not necessary to use a Monte Carlo approximation, and also because this example is not usually mentioned in the bootstrap literature (the only reference found was Hjorth, 1994, pp. 110–111, and not with the options taken here).

The non-parametric bootstrap method introduced by Efron (1979), consists on making inferences about a population using solely the empirical distribution of the observed sample. In the present context, as the sample consists on X successes and $n-X$ failures, the empirical distribution function is given by

$$F_n(y) = \begin{cases} 0, & y<0 \\ 1-\dfrac{X}{n}, & 0\le y<1 \\ 1, & y\ge 1 \end{cases},$$

that is, the distribution function of a Bernoulli$\left(\frac{X}{n}\right)$ random variable. Considering the estimator of p, $\hat{p}=X/n$, we obtain the bootstrap distribution of this estimator by noting that the bootstrap distribution of $n\hat{p}$ is the distribution of the number of successes on a random sample of size n from a Bernoulli$\left(\frac{X}{n}\right)$ population, that is, Binomial$\left(n, \frac{X}{n}\right)$.

Given this distribution several methods can be used to obtain two-sided confidence intervals for the parameter of interest, p. One of those methods is the Percentile Method, which in this case consists on taking the percentiles $\alpha/2$ and $1-\alpha/2$ of the bootstrap distribution of $\hat{p}$, leading to Method XI in Table 2.

Since the parameter p varies continuously in $[0,1]$ and the quantities given by $L_{XI}(X)$ and $U_{XI}(X)$ vary discontinuously by $1/n$ it makes sense to introduce here a kind of continuity correction. This is called Method XII (see Table 2).

The Percentile Method is usually considered to have some drawbacks, and several corrections have been proposed for it. One is the BCP (*Bias Corrected Percentile*, see e.g. Shao and Tu 1995) which consists on replacing the previous percentiles ($\alpha/2$ and $1-\alpha/2$) by other percentiles accounting for the asymmetry of the bootstrap distribution. Thus, $\alpha/2$ is replaced by

$$\alpha' = \Phi\left(z_{\alpha/2} + 2\times\Phi^{-1}\left(K_B\left(\frac{X}{n}\right)\right)\right) \tag{2.5}$$

and $1-\alpha/2$ by

$$\alpha'' = \Phi\left(z_{1-\alpha/2} + 2\times\Phi^{-1}\left(K_B\left(\frac{X}{n}\right)\right)\right), \tag{2.6}$$

where K_B denotes the bootstrap distribution function. Due to the discrete nature of this distribution a further correction must be applied and the one used was

$$K_B\left(\frac{X}{n}\right) = \left[F_{\mathrm{Bin}\left(n,\frac{X}{n}\right)}(X) + F_{\mathrm{Bin}\left(n,\frac{X}{n}\right)}(X-1)\right]\Big/2 \ .$$

Note that if the bootstrap distribution is symmetric (which happens when $X/n \simeq 0.5$) then $K_B(X/n) \simeq 0.5$, $\alpha' \simeq \alpha/2$ and $\alpha'' \simeq 1-\alpha/2$, and there is practically no correction to the raw percentile method. Method XIII refers to the bootstrap BCP method and Method XIV is similar but with the continuity correction introduced above.

In Methods XI to XIV the zero length intervals for $X=0$ and $X=n$ have been replaced by the exact Clopper–Pearson expressions (as it was done for Methods VII to X).

2.4. Normal approximations after a transformation

The next methods considered are based on the approximate normal distribution after the variance stabilizing transformation, that is on

$$\arcsin\sqrt{\frac{X}{n}} \ \xrightarrow{d} \ \mathcal{N}\left(\arcsin\sqrt{p},\, \frac{1}{4n}\right) \ .$$

Solving for p and correcting for inconsistencies in the extremes one obtains Method XV in Table 2. Introducing a continuity correction leads to Method XVI. A refinement of this method using a correction due to Anscombe (1948) is also considered (Method XVII).

2.5. Other approximations

This subsection introduces the last three methods consisting on very recent suggestions.

Agresti and Coull (1998) noting that the score interval has very good properties (confirmed in the comparative studies of Vollset 1993, Newcombe 1998, and again in the next section of this paper), that it is centered around

$$\tilde{p}(c) = \frac{X + c^2/2}{n + c^2} \ , \tag{2.7}$$

and that for 95% confidence $c^2 \simeq 4$, suggested a simple, yet effective, method:

add 4 observations to the sample, two successes and two failures, and then use the Wald formula (Method V). This method will be referred as Add 4 or Method XVIII. They also propose the use of $\tilde{p}(2)$ as point estimator and call it the Wilson point estimator, since Wilson (1927) was the first statistician recommending it (as a curiosity note that $\tilde{p}(\sqrt{2})$ is also an old estimator, the Laplace estimator).

Pan (2002) proposes a further modification, both on the Wald interval and the Add 4 interval, which consists on using percentiles of a suitable t-distribution instead of the normal. The modification of the Wald interval is denoted Method XIX whereas the one for the Add 4 interval is referred as Method XX. Let $V(p,n) = p(1-p)/n$ be the variance of $\hat{p}$. By introducing a scaled chi-square distribution and matching its first two moments with those of $V(\hat{p},n)$, Pan (2002) concludes that for the Wald-t interval the appropriate degrees of freedom are given by

$$\nu = \frac{2\,V(\hat{p},n)^2}{\Omega(\hat{p},n)}\ , \tag{2.8}$$

and that for the Add 4-t those are given by

$$\nu = \frac{2\,V\big(\tilde{p}(2),n+4\big)^2}{\Omega\big(\tilde{p}(2),n+4\big)}\ , \tag{2.9}$$

where

$$\begin{aligned}\Omega(p,n) \;=\; & \frac{p-p^2}{n^3} + \frac{p+(6n-7)p^2+4(n-1)(n-3)p^3-2(n-1)(2n-3)p^4}{n^5} \\ & -\,2\,\frac{p+(2n-3)p^2-2(n-1)p^3}{n^4}\ .\end{aligned}$$

3. COMPARISON OF THE METHODS

3.1. Central intervals

In order to evaluate and compare the performance of the twenty methods presented in Section 2, the coverage probability and the corresponding expected length have been computed for 5000 values of p, equally spaced in $[0.0001, 0.5]$, for every $10 \le n \le 1000$ and for $\alpha = 0.05,\ 0.01$.

The coverage probability, function of p, n and the method, $i = \mathrm{I}, ..., \mathrm{XX}$, is given by

$$CP\big(p,n,i\big) \;=\; \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j}\, I_{[L_i(j),U_i(j)]}(p)\ , \tag{3.1}$$

where $I_{[a,b]}(p)$ denotes the indicator function of the interval $[a,b]$, i.e. $I_{[a,b]}(p)=1$, if $p\in[a,b]$, and $I_{[a,b]}(p)=0$, if $p\notin[a,b]$. The expected length is

$$EL\big(p,n,i\big) \;=\; \sum_{j=0}^{n}\binom{n}{j}p^j(1-p)^{n-j}\left(L_i(j)-U_i(j)\right)\;. \tag{3.2}$$

The above computations were performed in R (code available upon request). The results are exact, within machine accuracy, and shall therefore not be interpreted as Monte Carlo results (from the papers cited before the only one giving Monte Carlo results is García-Pérez, 2005).

Figures 1 and 2 show plots of the coverage probability for two non extreme cases, $n=50$ and $n=500$, with $\alpha=0.05$. The plots for other values of n and for $\alpha=0.01$ are qualitatively similar to these two. The non-smooth aspect is expected due to the presence of the indicator function in expression (3.1). In fact, the coverage probability, as a function of p, has as many discontinuity points as the number of distinct values of L_i and U_i, about $2n$ in $(0,1)$. Between the discontinuity points $CP(p,n,i)$ is a polynomial of degree n.

Since it is impossible to analyze the several thousands of plots that could be produced, the results for the coverage probability were summarized in terms of observed minimum and mean on p for each n, and then plotted as a function of n. These plots are shown in Figures 3, 4, 5 and 6.

Considering the criteria described in Section 1, a possible classification of the methods is the following:

1st Group – Strictly conservative methods, i.e. methods for which the minimum coverage probability is, for all $n\geq 10$ and for all p, greater or equal to $1-\alpha-0.005$ (nominal coverage probability rounded to two decimal places):

$$i:\quad \min_p CP\big(p,n,i\big)\;\geq\;1-\alpha-0.005\,,\quad \forall_{n\geq 10}\;.$$

2nd Group – Methods not strictly conservative but correct on average, i.e. with mean coverage probability, for all $n\geq 10$, greater or equal to $1-\alpha-0.005$:

$$i:\quad \int_0^1 CP\big(p,n,i\big)\,dp\;\geq\;1-\alpha-0.005,\quad \forall_{n\geq 10}\;.$$

3rd Group – Methods which are neither strictly conservative nor correct on average.

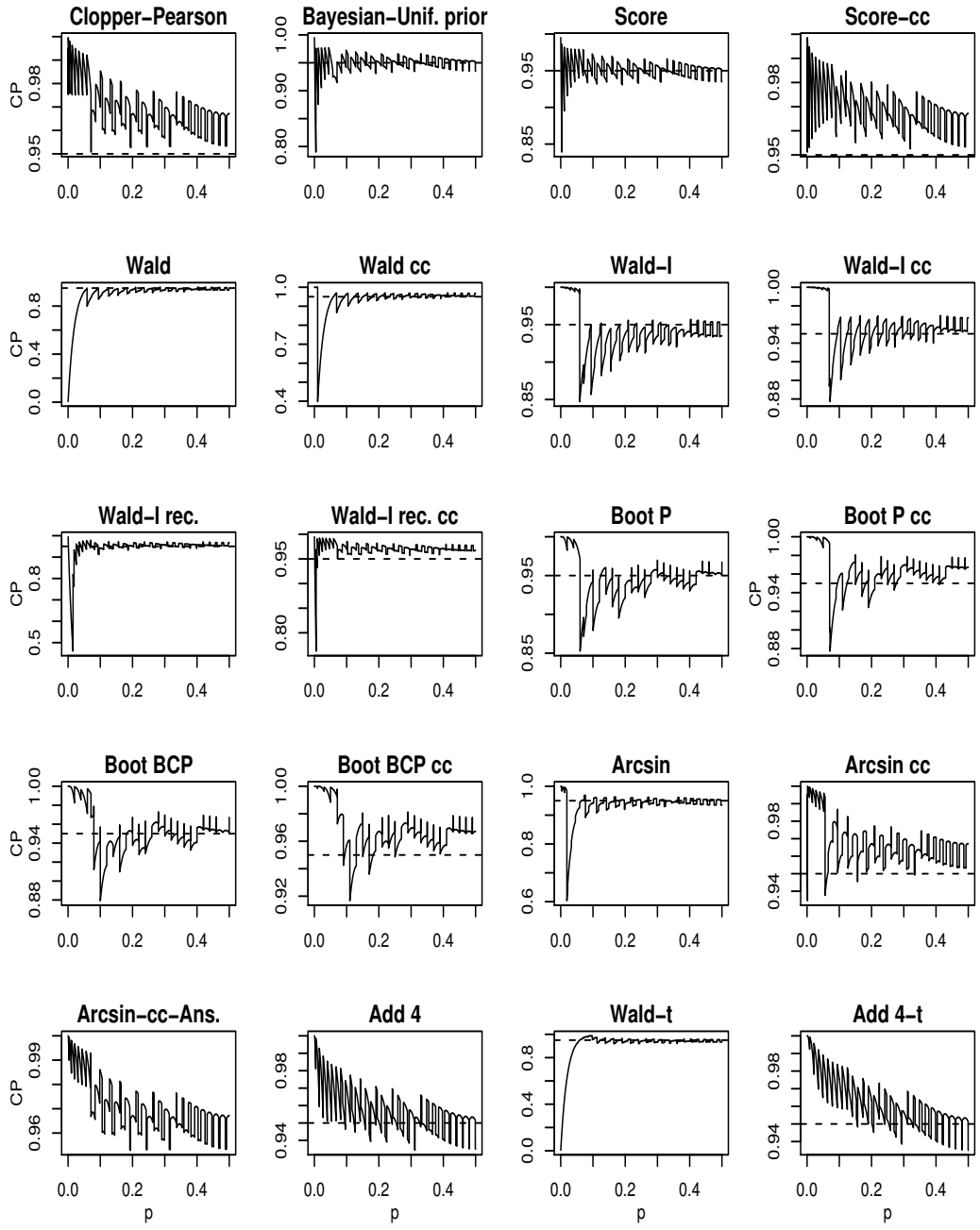

Figure 1: Coverage probability for each method as a function of p for $n = 50$ and $\alpha = 0.05$.

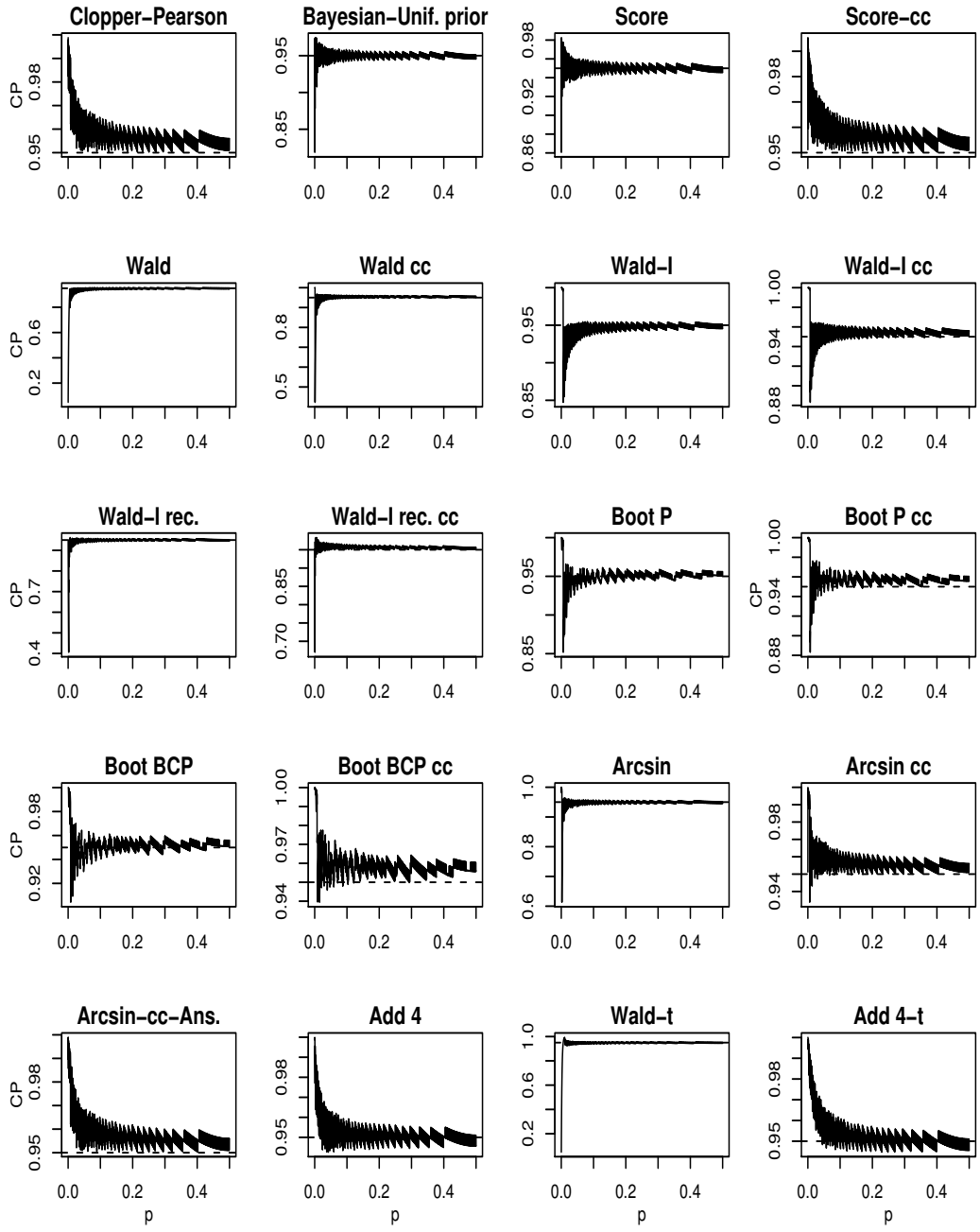

Figure 2: Coverage probability for each method as a function of p for $n = 500$ and $\alpha = 0.05$.

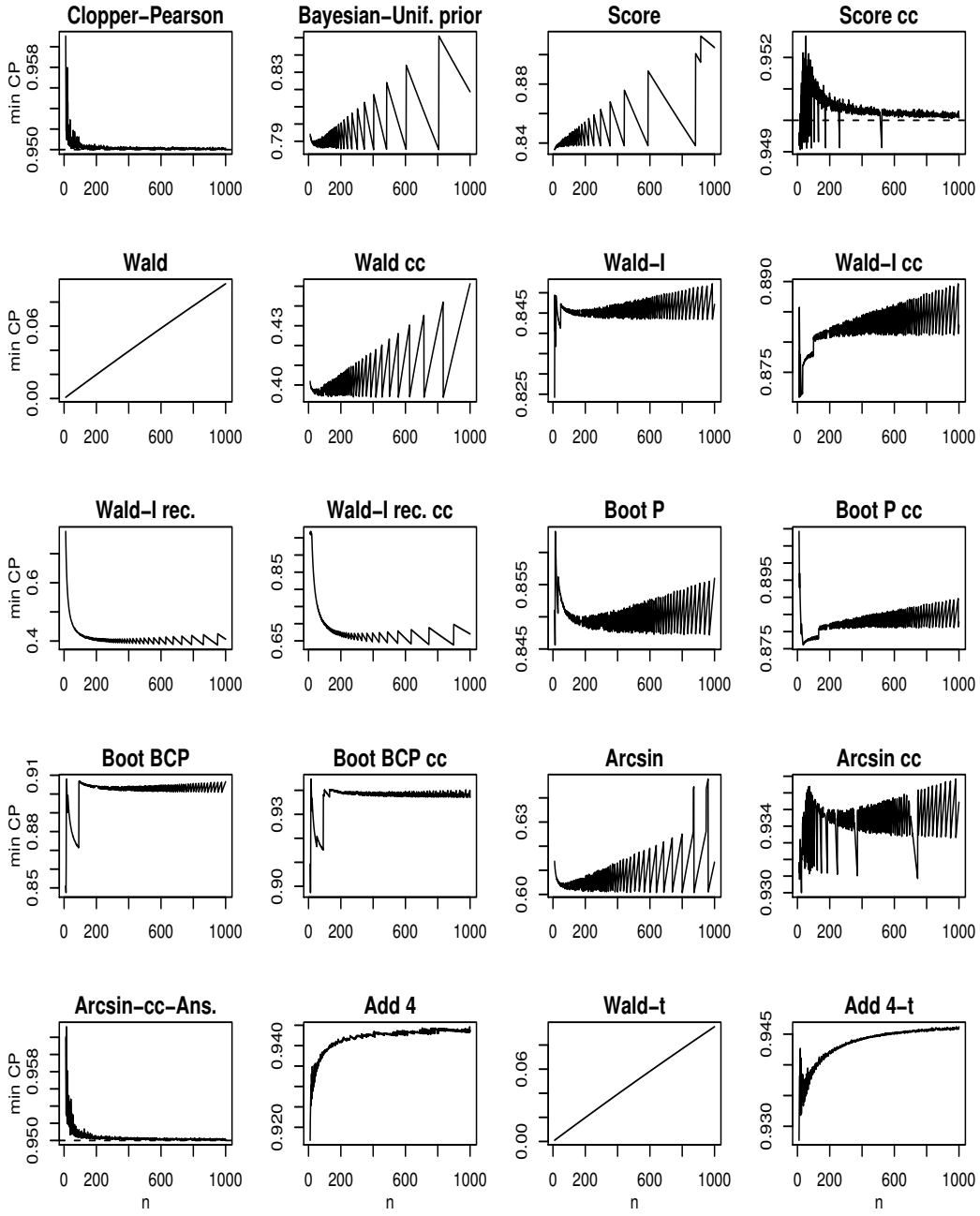

Figure 3: Minimum coverage probability for each method as a function of n $(10 \leq n \leq 1000)$ for 95% confidence.

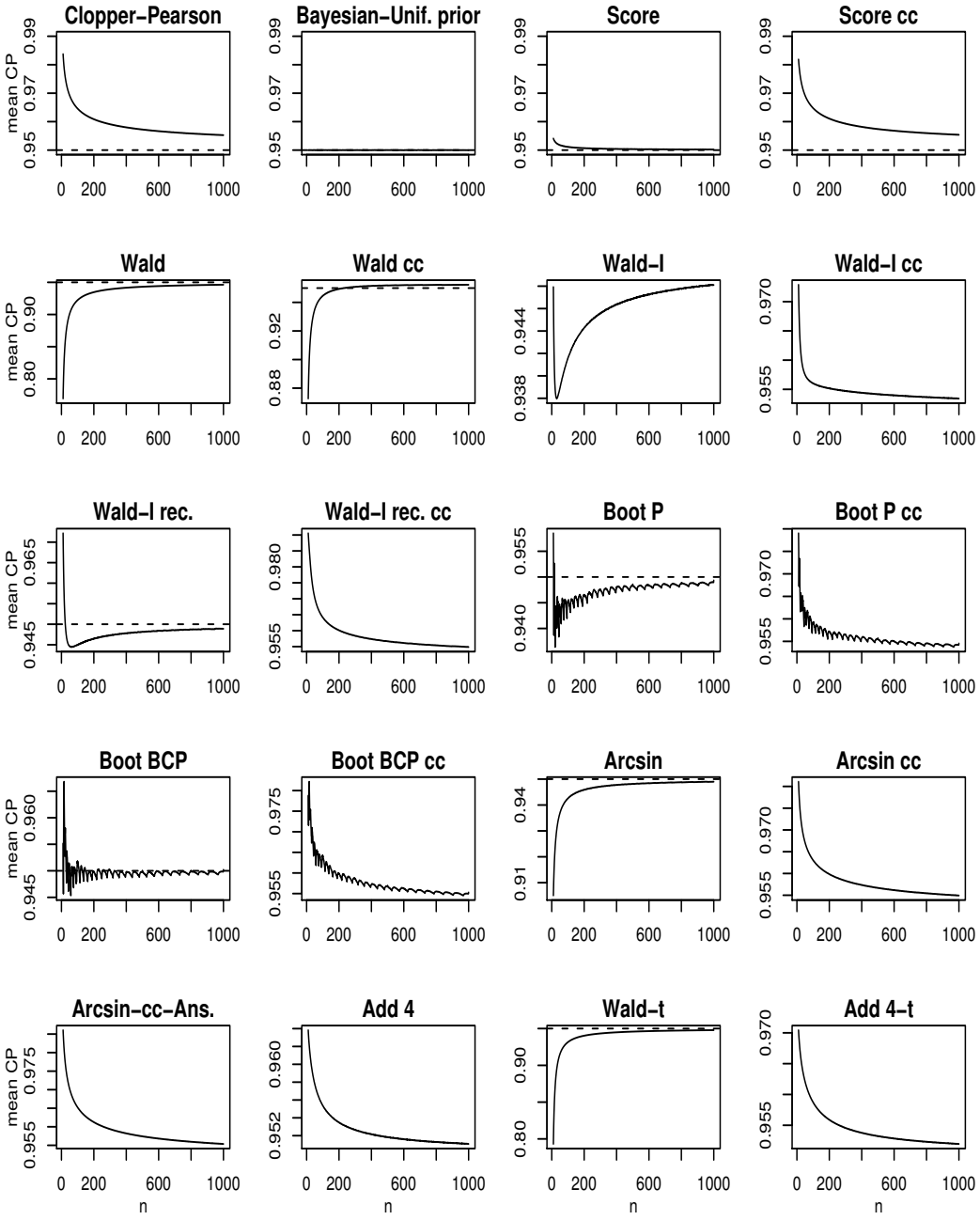

Figure 4: Mean coverage probability for each method as a function of n $(10 \leq n \leq 1000)$ for 95% confidence.

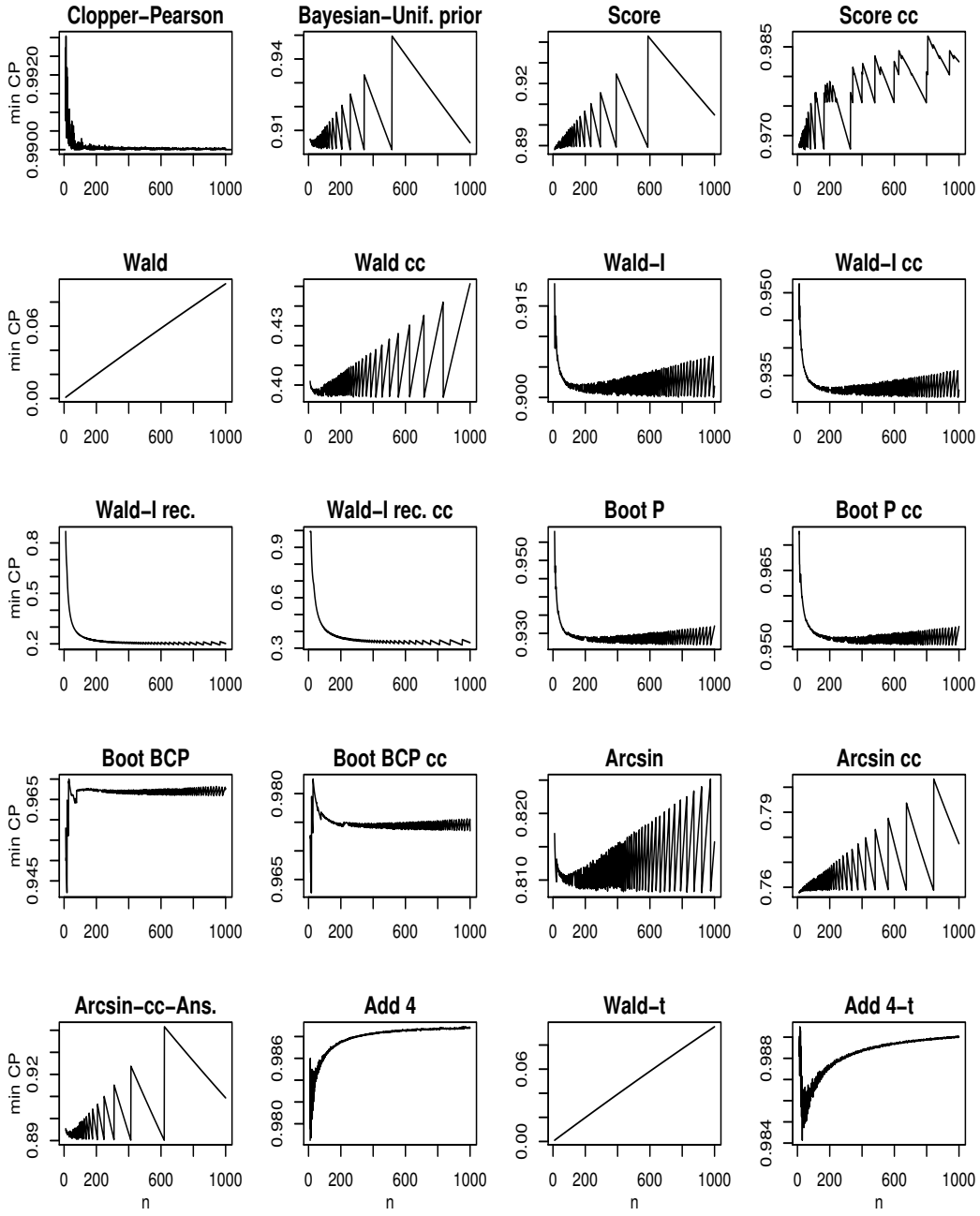

Figure 5: Minimum coverage probability for each method as a function of n $(10 \leq n \leq 1000)$ for 99% confidence.

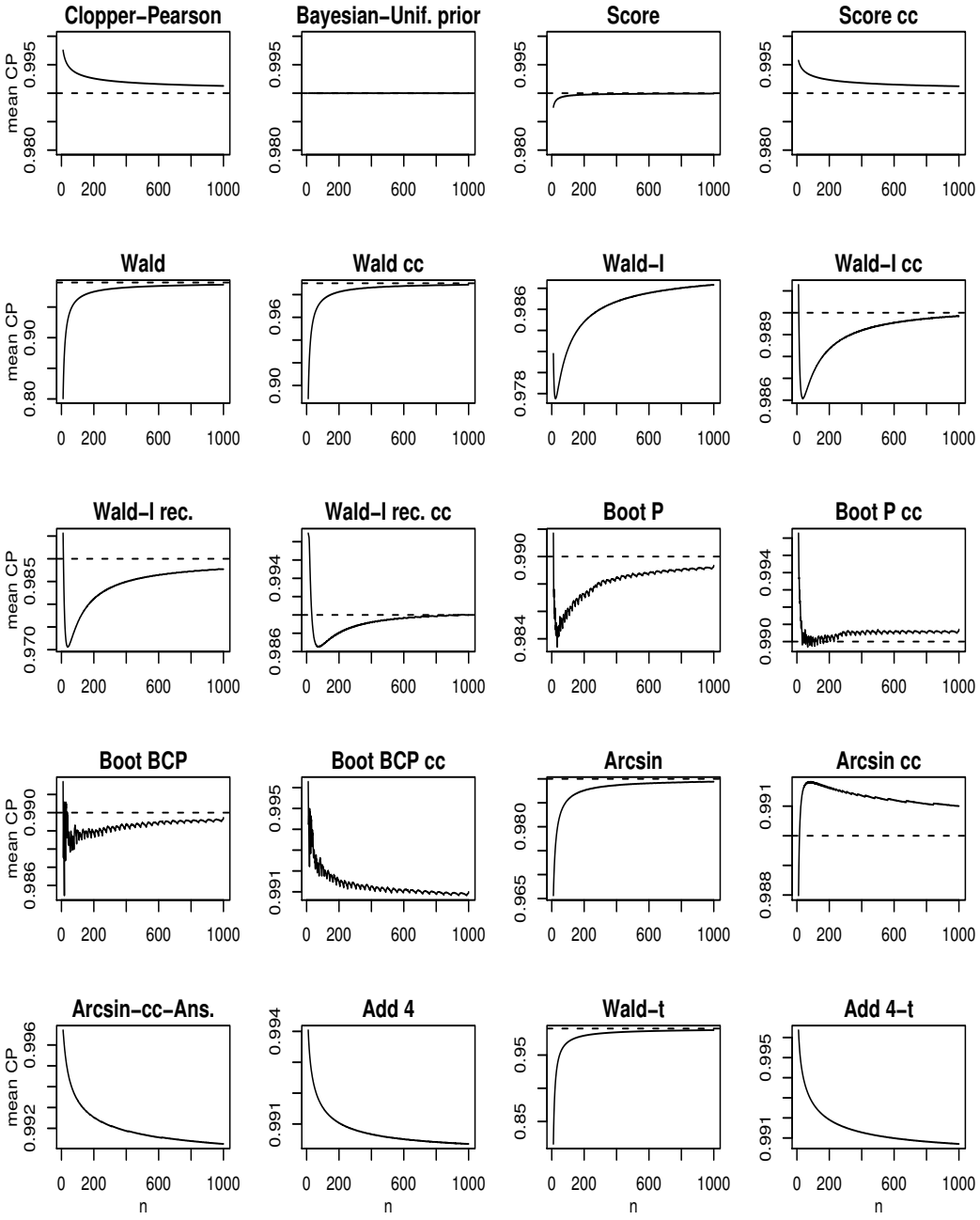

Figure 6: Mean coverage probability for each method as a function of n $(10 \leq n \leq 1000)$ for 99% confidence.

Table 3 summarizes the classification. 2% exceptions were allowed for classifying a method in the first group. The methods in the first group are shown by increasing order of mean coverage probability whereas those in the second group are shown by decreasing order of the minimum of the coverage probabilities overall n and p (shown into brackets). The composition of the groups depends slightly on the minimum value of n considered in this evaluation ($n \geq 10$). If the criterion was much less stringent, for instance $n \geq 200$ then: for 95% confidence Boot BCP cc and Arcsin cc would move to the first group and Wald cc and Boot P would enter the second group; for 99% confidence Add 4 would move to the first group and Boot P would enter the second group.

Table 3: Classification of the methods: strictly conservative (1^{st} group); average correct (2^{nd} group) (overall minimum of the coverage probabilities). Methods shown in **boldface** have not been considered in other comparative studies in the literature.

Group	**95% confidence**	**99% confidence**
1^{st}	I – Clopper Pearson IV – Score cc **XVII – Arcsin cc Anscombe**	XX – Add 4-t **XVII – Arcsin cc Anscombe** I – Clopper Pearson
2^{nd}	**XVI – Arcsin cc** (93%) XX – Add 4-t (93%) XVIII – Add 4 (92%) **XIV – Boot. BCP cc** (90%)	XVIII – Add 4 (98%) IV – Score cc (97%) **XIV – Boot. BCP cc** (96%) **XII – Boot. P cc** (95%)
	XII – Boot. P cc (88%) VIII – Wald-I cc (87%) **XIII – Boot. BCP** (85%) III – Score (84%) **II – Bayesian uniform prior** (79%) **X – Wald-I rec. cc** (64%)	**XIII – Boot. BCP** (94%) VIII – Wald-I cc (93%) **II – Bayesian uniform prior** (90%) III – Score (89%) **XVI – Arcsin cc** (76%) **X – Wald-I rec. cc** (32%)

It is worth noting that seven out of the thirteen methods classified in the first and second groups were not considered in the largest comparative studies in the literature (Vollset 1993, Newcombe 1998, Agresti and Coull 1998, Pan 2002).

Note that Method II is exact on average due to the coincidence between the prior distribution admitted and the distribution of p actually used to compute the mean coverage probability. The performance of this method is not so good due to the rather small value of the minimum coverage probability. If we had chosen the Jeffreys prior instead this aspect would have improved a little: 87% for 95% confidence (placing it between Methods VIII and XIV) and 96% for 99% confidence (between Methods IV and XIV).

It is also remarkable that some of the intervals maintain their good behavior for n as low as 10, in spite of being based on asymptotic results. It is the case

namely of the score intervals (with and without continuity correction) and the ones based on the arcsine transformation.

Another interesting feature revealed by these results is that an apparently small modification may have a great impact in the performance of a method. For instance, the Wald method was completely disqualified in the vast majority of the papers mentioned in the introduction, however a simple modification at the boundary values and the simultaneous use of the continuity correction leads to an acceptable method, with better performance than the score or the Bayesian intervals. Figure 7 shows three plots illustrating this aspect for some values of n and α referred as "unlucky" by Brown, Cai and DasGupta (2001, 2002).

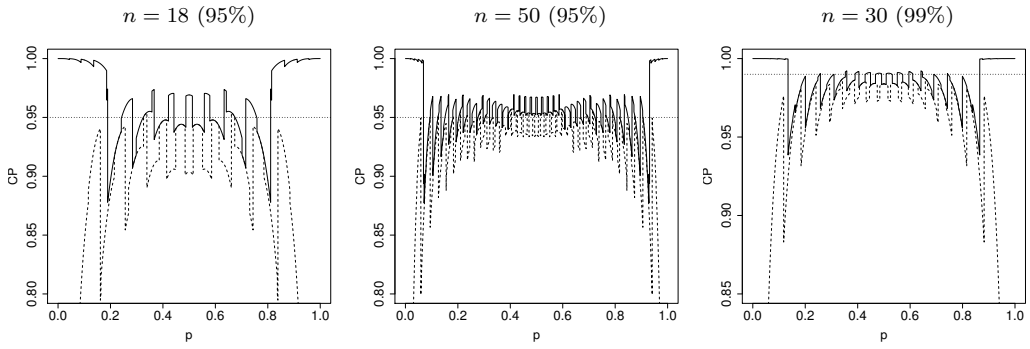

Figure 7: Coverage probability of the Wald interval (Method V, dashed) and the Wald interval with a modification at the boundary and continuity correction (Method VIII, solid) for some "unlucky" values of n and α.

Not surprisingly it is observed that the methods with continuity correction are always better in terms of coverage probability than the corresponding ones without that correction. It is also possible to verify that the bootstrap BCP method is slightly better than the percentile method and that both of these methods outperform the Wald methods. The results also show that the t correction of the Wald method proposed by Pan (2002) does not achieve its aim and it is in fact less effective than the usual continuity correction.

After this analysis of the coverage probabilities it is important to compare the expected lengths of the intervals. This comparison makes sense only within each group and only for the first and the second groups. Taking as reference Method I (Clopper–Pearson or exact) the ratio between the expected length of the intervals obtained using the other methods in the first group and the expected length by Method I was computed. The same was done with the first four methods in the second group (considering that the remaining, in spite of being correct on average, have an undesirable behavior in terms of the minimum coverage

probability). Figure 8 shows the corresponding plots for $n = 100$. The plots for other values of n are qualitatively similar, but with the differences between the methods decreasing with n, especially at medium values of p.

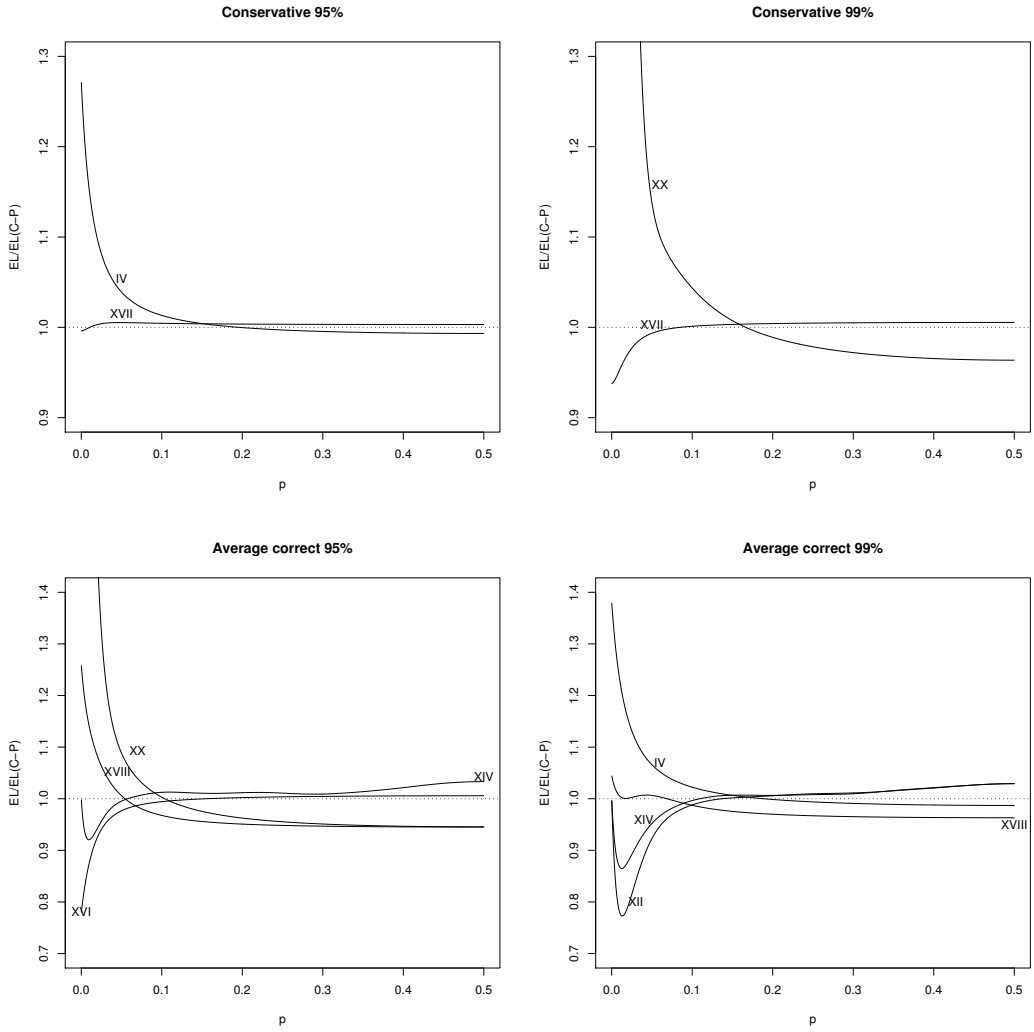

Figure 8: Ratio between the expected length of the conservative intervals and the expected length of the Clopper–Pearson interval (top) for $n = 100$. Ratio between the expected length of the first four top average correct intervals and the expected length of the Clopper–Pearson interval (bottom) for $n = 100$.

For the first group (top plots of Figure 8) the conclusion is that the Arcsine method with Anscombe's continuity correction is almost equivalent to the exact Clopper–Pearson interval in terms of length (and degree of conservativeness) but the Add 4-t is unnecessarily wide (or conservative).

In the second group (bottom plots of Figure 8) the conclusions are not so straightforward because there are more methods involved. However, for 95% confidence, it is possible to conclude that Method XVI (Arcsin cc) dominates Method XIV (Boot. BCP cc), having both better coverage and length. Method XVIII (Add 4) almost dominates Method XX (Add 4-t), it has better length but slightly smaller coverage probability. The choice between the dominating two methods is not so easy and depends on prior knowledge of the true p, if p is not small or large, Method XVIII is better because it leads to smaller length, if, on the contrary, p is either small or large, Method XVI is better. What is "small" or "large" depends on n and may not be easy to choose but after all it is not so important because any of these methods will produce a reasonable and safe interval whatever the value of p. For 99% confidence the conclusion is that Method XVIII (Add 4) dominates Method IV (Score cc) because of better length and coverage probability. The other two may lead to smaller lengths but at the cost of undesirably small minimum coverage probability. The best choice appears, therefore, to be the Add 4 method.

3.2. Non-central intervals

If one feels comfortable with the concept of a non-central interval then there are only two methods to choose from: the exact method according to the two criteria (minimum coverage probability of at least $1-\alpha$ or mean coverage probability equal to $1-\alpha$) and minimizing length.

To meet the first criterion and minimize the length of the interval one has to invert the test $H_0\colon p=p_0$ versus $H_1\colon p\neq p_0$ with size α, choosing for each p_0 the acceptance region, $A_n(p_0)\leq X\leq B_n(p_0)$, with smallest length. Then, given X, the confidence region is the set of those p_0 for which X is in the corresponding acceptance region. This is not an easy task and in fact many authors have addressed it (Sterne, 1954, Crow, 1956, Clunies-Ross, 1958, Blyth and Still, 1983, Casella, 1986, Reiczigel, 2003, see also the discussion in Santner and Duffy, 1989). The method we have implemented is based on Sterne's proposal (the acceptance interval for $p=p_0$ is made by including the most probable value of X, then the next most probable, ..., until the sum of their probabilities is greater than $1-\alpha$) with a slight modification. This modification is needed because although the acceptance region is always an interval the inverted region for p is not. We simply fill in the holes when they appear, which reduces to compute the interval by

$$L_{St}(X) \;=\; \min\Big\{p\colon\; A_n(p)\leq X\leq B_n(p)\Big\}$$

and

$$U_{St}(X) \;=\; \max\Big\{p\colon\; A_n(p)\leq X\leq B_n(p)\Big\}\,.$$

The slight modification has a very small practical effect: when compared with the intervals given in Table 2 of Blyth and Still (1983) for $n \leq 30$ there is only one different.

Figure 9 represents the results corresponding to Figures 1 and 2. It is clear that the Sterne interval is closer to the desired coverage than the Clopper–Pearson interval (it is less conservative but its coverage is still always over $1-\alpha$ as it should). This translates into smaller mean coverage probability (see Figure 10).

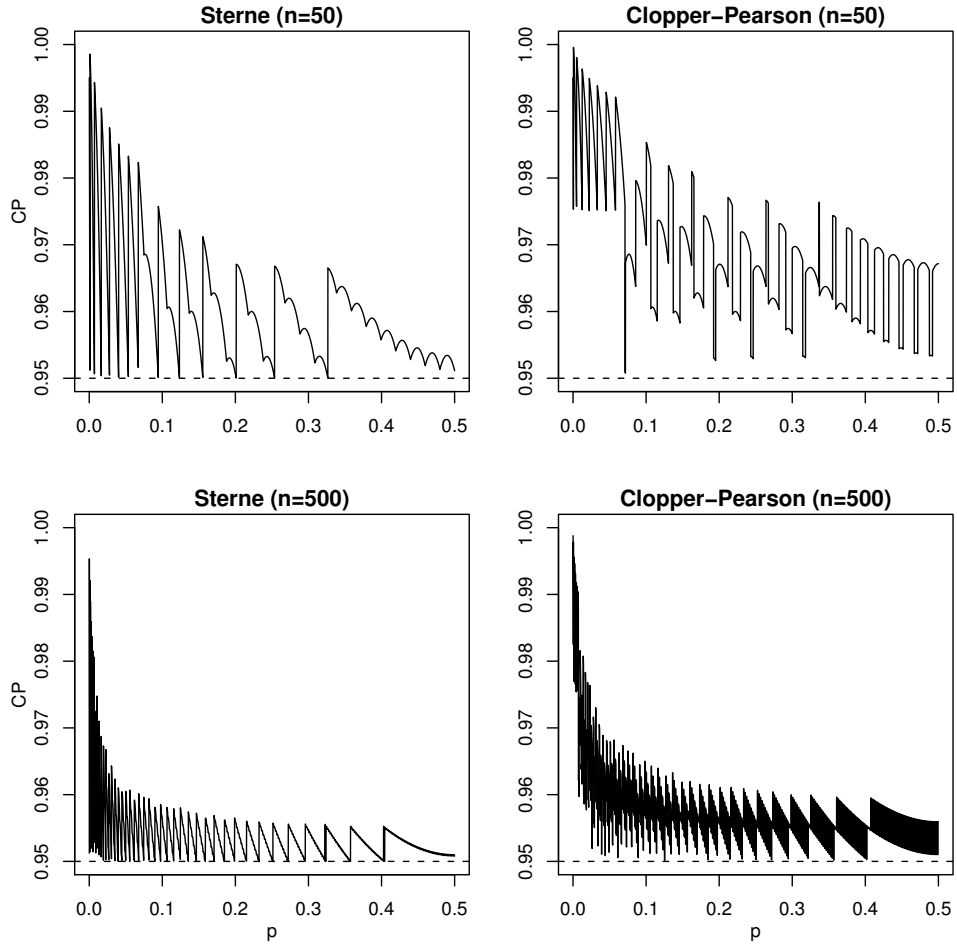

Figure 9: Coverage probability for the two exact frequentist intervals (non-central and central) as a function of p for $n = 50$ and $n = 500$ ($\alpha = 0.05$).

Figure 11 reproduces the top two plots of Figure 8 with the curve corresponding to the Sterne interval. It does not come as a surprise that this interval has smaller expected length than the Clopper–Pearson interval except for values of p very close to the boundary.

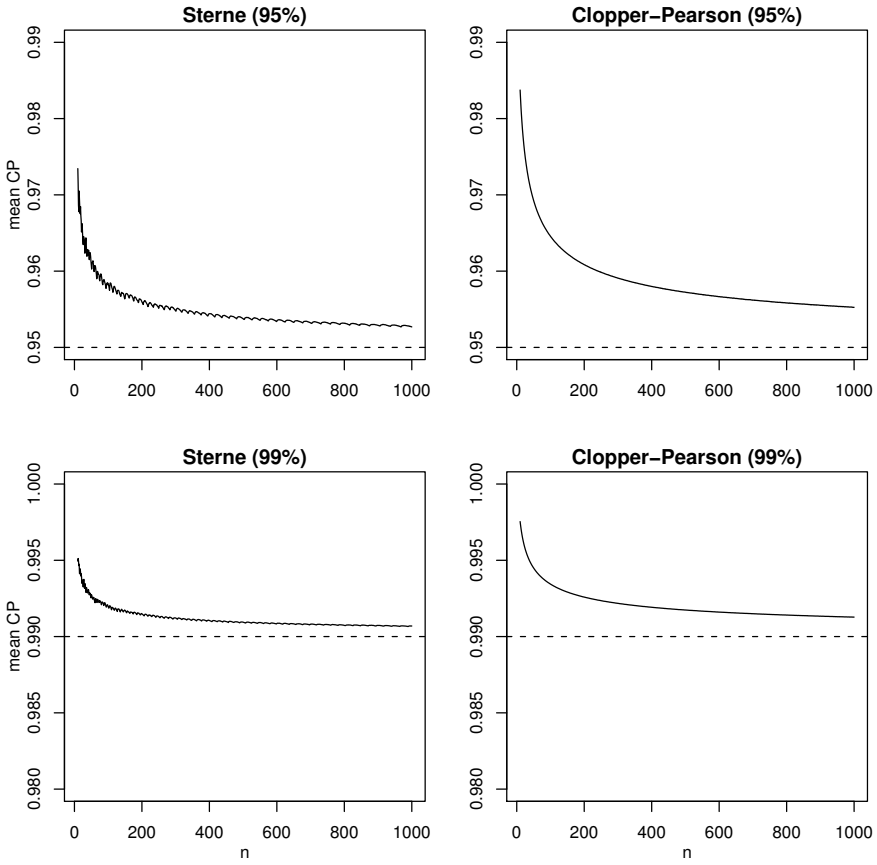

Figure 10: Mean coverage probability for the two exact frequentist intervals (non-central and central) as a function of n ($10 \leq n \leq 1000$) for 95% and 99% confidence.

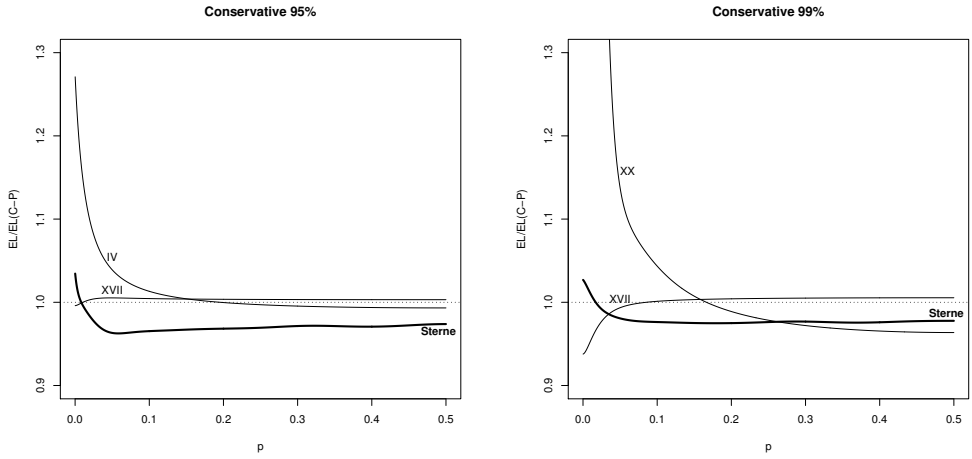

Figure 11: Ratio between the expected length of the conservative central intervals and the expected length of the Clopper–Pearson interval for $n = 100$ (solid thin lines) together with the same ratio for the non-central exact Sterne interval (solid thick lines).

When considering the second criterion (or mean coverage probability equal to $1-\alpha$) minimal length is achieved at ease by computing an HPD credibility interval. Given X, the HPD interval for the *a posteriori* Beta$(X+1, n-X+1)$ distribution can be determined in the following way: for every possible value of the left credibility tail, $0 \leq \alpha' \leq \alpha$, define

$$L(X,\alpha') = B_{X+1,n-X+1;\alpha'} \quad \text{and} \quad U(X,\alpha') = B_{X+1,n-X+1;1-(\alpha-\alpha')}$$

and determine $\alpha'(X)$ such that $U(X,\alpha') - L(X,\alpha')$ is minimum and denote it $\alpha'(X)$. The interval is given by $L_{HPD}(X) = L(X,\alpha'(X))$ and $U_{HPD}(X) = U(X,\alpha'(X))$. Note that, as mentioned in Subsection 2.1, we have $\alpha'(0) = 0$, $L_{HPD}(0) = 0$, $U_{HPD}(0) = 1-\alpha^{1/(n+1)}$; and $\alpha'(n) = \alpha$, $L_{HPD}(n) = \alpha^{1/(n+1)}$, $U_{HPD}(n) = 1$.

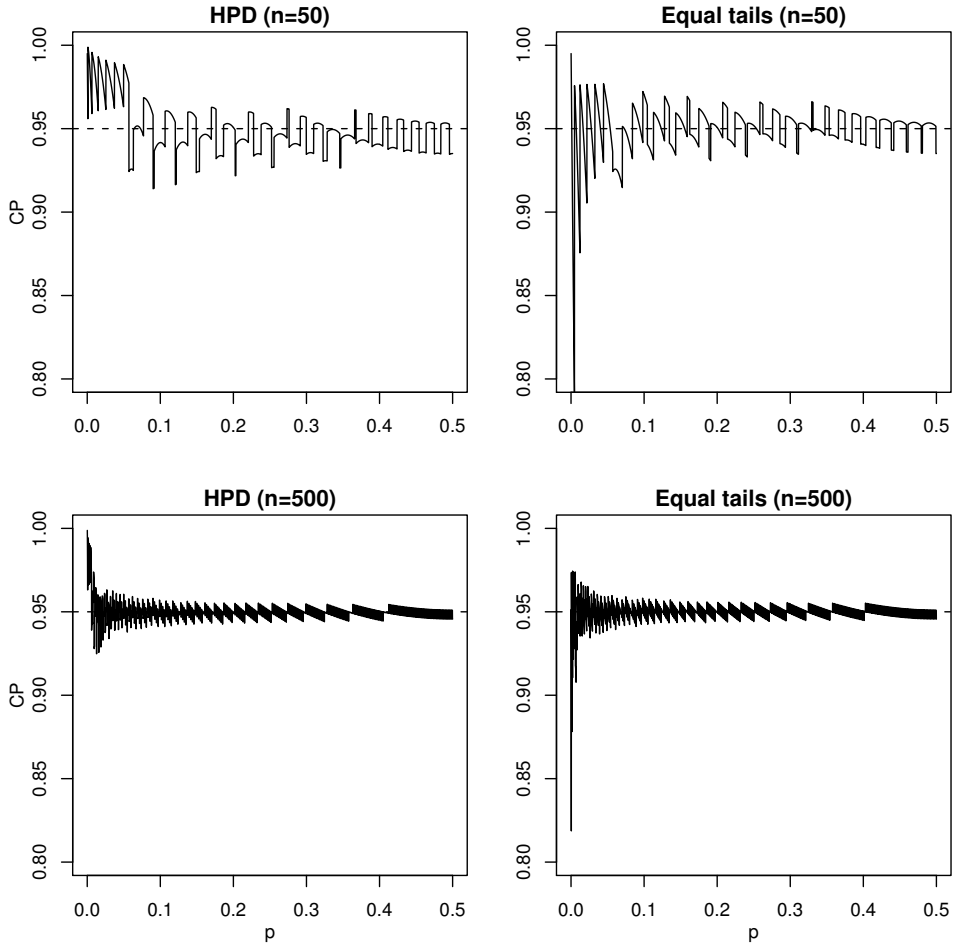

Figure 12: Coverage probability for the two Bayesian intervals with uniform prior (HPD and equal credibility tails) as a function of p for $n = 50$ and $n = 500$ ($\alpha = 0.05$).

Figure 12 represents the results corresponding to Figures 1 and 2. For both intervals the coverage fluctuates around the target value but the HPD does not have the downward spikes lose to the boundaries. This is also evident from the minimum coverage probability plots (see Figure 13).

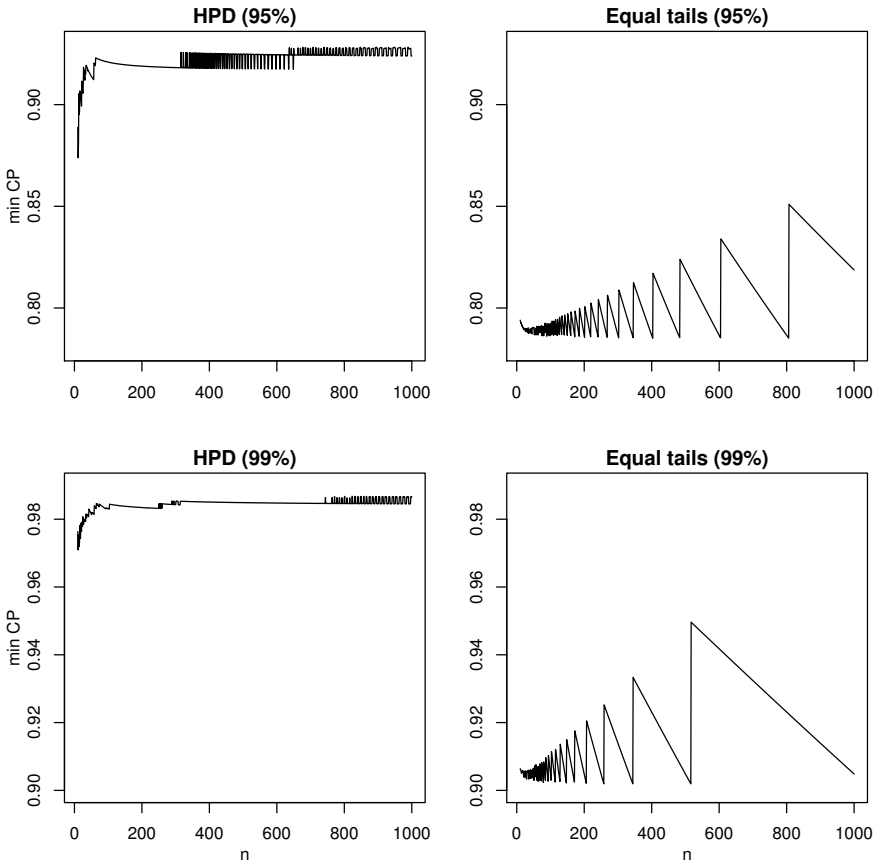

Figure 13: Minimum coverage probability for the two Bayesian intervals with uniform prior (HPD and equal credibility tails) as a function of n ($10 \leq n \leq 1000$) for 95% and 99% confidence.

Figure 14 reproduces the bottom two plots of Figure 8 with the curves corresponding to the two Bayesian intervals. As expected the HPD interval has smaller expected length than the other intervals except for values of p very close to the boundary (only for 99% confidence).

We have thus verified the optimality of the two exact procedures. What the statistician must decide is whether the reduction in expected length (of approximately 2% or 3% in the strictly conservative case and of approximately 5% or 7% in the average correct case) is worth the complications involved in the computations and the somehow different interpretation associated to non-central intervals.

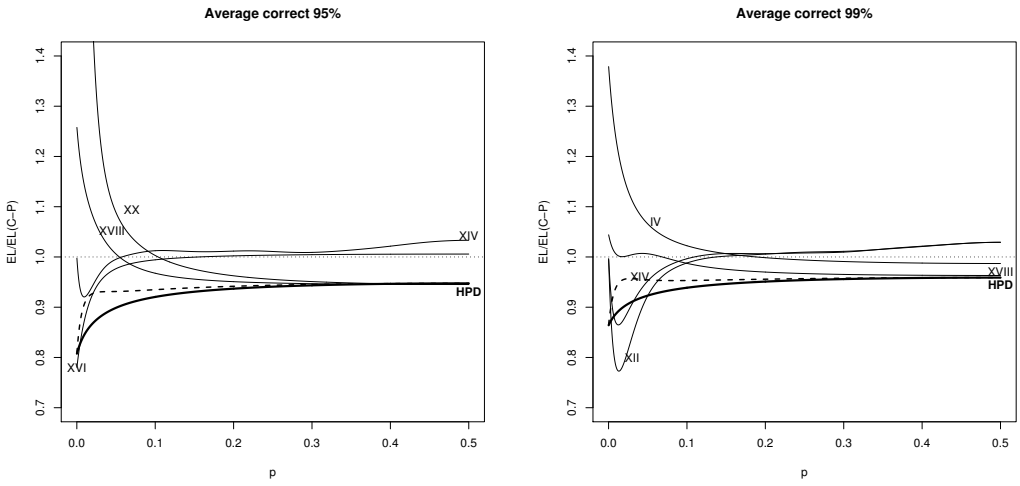

Figure 14: Ratio between the expected length of the first four top average correct central intervals and the expected length of the Clopper–Pearson interval for $n = 100$ (solid thin lines) together with the same ratio for the central Bayesian interval (dashed thick lines) and for the HPD interval (solid thick lines).

4. CONCLUDING REMARKS

The results reported in this paper have brought new insight into the apparently easy problem of determining an interval estimate for a binomial proportion. Although there is not a unique uniformly best choice, it is now easier to answer the two questions posed in the introduction and related, respectively, to applications and teaching.

4.1. Applications

When considering the computation of an interval estimate for a binomial proportion the first decision the applied statistician must take is related to the balance between degree of conservativeness and efficiency (equivalent in this case to the length of the intervals). Let us consider the two extreme options and only the class of central intervals:

(**i**) If strict conservativeness is mandatory than he or she must choose the Clopper–Pearson interval, or, almost equivalently, the arcsine interval with Anscombe's correction.

(**ii**) If strict conservativeness is not a major concern then length must be, subject to being at least correct on average. It is also wise to limit the "damage" measured by the overall minimum of the coverage probabilities and the recommendation is to choose, in the case of 95% confidence, either the Arcsin cc method or the Add 4 method. For 99% confidence the recommended method is the Add 4.

The score interval with continuity correction remains a valid choice, except that it may be too wide if the true p is close to 0 or to 1.

4.2. Teaching

In addition to the concerns of the applied statistician the teacher of statistics must also take into account the nature of the course and this may complicate the decision.

In a course for future statisticians the recommendations given in the previous subsection apply. The various methods should be taught and thoroughly discussed.

For elementary courses, typically less mathematically oriented and often unique, simplicity and lack of time for in depth discussions are a major concern. The Add 4 method of Agresti and Coull (1998) appears as a good choice, its properties are good and it is easy to compute. If, maybe for other reasons, one wants to stick to the Wald method then at least the continuity correction and the boundary modification should always be included.

4.3. Software

Four major statistical packages (SAS 9.1.3, S-Plus 8, SPSS 15 and R 2.6) were analyzed concerning the availability and correct implementation of interval estimates for the binomial parameter.

SAS provides, through PROC FREQ, the Wald interval, with and without continuity correction, and the exact Clopper–Pearson interval obtained with the percentiles of the F distribution.

In S-Plus there are two commands related to binomial proportions. The `prop.test` command gives the normal based hypotheses tests and the score intervals, with and without continuity correction. However, when using the continuity correction and when $X = 0$ or $X = n$, the intervals given by this command

are wrong, leading to $L(0) > 0$ or $L(n) < 1$. Another undesirable feature of this command is that it never applies the continuity correction when $X = n/2$, even when this option is set to TRUE. These apparently small details may have a strong visual impact and determine the classification of the method (see Figure 9). The `binom.test` command gives only the exact hypothesis test but could be easily modified in order to provide the exact Clopper–Pearson interval.

SPSS provides the asymptotic and the exact tests for binomial proportions but no confidence intervals (unlike other situations for which both the test and the confidence interval are provided, e.g. t test for the mean). Separately, there is a document describing how to compute the equal-tailed Jeffreys prior intervals (which are represented in bar charts).

The R software has commands with the same names as those of S-Plus (`prop.test` and `binom.test`), but the first makes the boundary correction and the second also gives the exact Clopper–Pearson interval. However the `prop.test` has the same problem with the continuity correction when $X = n/2$ (see Figure 15 and also Figure 1 of Geyer and Meeden, 2005).

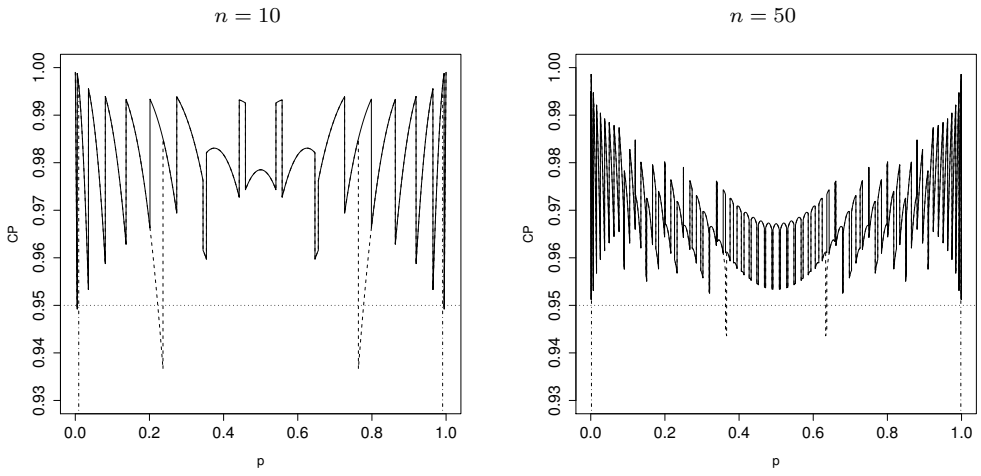

Figure 15: Effect of forcing no continuity correction for $X = n/2$ on the coverage probability of the Score cc interval (95% confidence). Correct implementation of the method (solid line) and implementation with the command `prop.test` of R and S-Plus (dashed line). The dashed-dotted line represents the results for `prop.test` of S-Plus at the extremes (no boundary correction when $X = 0$ or $X = n$).

In summary, the analyzed statistical packages do not treat the subject uniformly. This is perhaps a reflection of the recent spread of publications in the area. We hope that in the near future a consensus is reached and that it will be reflected in the software. This paper aims at contributing in that direction.

ACKNOWLEDGMENTS

We would like to thank an anonymous referee for helpful comments on the paper. This research was partially supported by the Center for Mathematics and its Applications, Lisbon, Portugal, through *Programa Operacional "Ciência, Tecnologia, Inovação"* (POCTI) of the *Fundação para a Ciência e a Tecnologia* (FCT), cofinanced by the European Community fund FEDER.

REFERENCES

[1] Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures, *The American Statistician*, **54**, 280–288.

[2] Agresti, A. and Coull, B. (1998). Approximate is better than "exact" for interval estimation of binomial proportions, *The American Statistician*, **52**, 119–126.

[3] Angus, J.E. and Schafer, R.E. (1983). Improved confidence statements for the binomial parameter, *The American Statistician*, **38**, 189–191.

[4] Anscombe, F.J. (1948). Transformations of Poisson, binomial and negative-binomial data, *Biometrika*, **35**, 246–254.

[5] Armitage, P. and Berry, G. (1987). *Statistical Methods in Medical Research, Second Edition*, Blackwell, Oxford.

[6] Blyth, C.R. (1986). Approximate binomial confidence limits, *Journal of the American Statistical Association*, **81**, 843–855.

[7] Blyth, C.R. and Still, H.A. (1983). Binomial confidence intervals, *Journal of the American Statistical Association*, **78**, 108–116.

[8] Brown, L.D.; Cai, T.T. and DasGupta, A. (2001). Interval estimation for a binomial proportion (with discussion), *Statistical Science*, **16**, 101–133.

[9] Brown, L.D.; Cai, T.T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions, *The Annals of Statistics*, **30**, 160–201.

[10] Cai, Y. and Krishnamoorthy, K. (2005). A simple improved inferential method for some discrete distributions, *Computational Statistics and Data Analysis*, **48**, 605–621.

[11] Casella, G. (1986). Refining binomial confidence intervals, *Canadian Journal of Statistics*, **14**, 113–129.

[12] Chen, H. (1990). The accuracy of approximate intervals for a binomial proportion, *Journal of the American Statistical Association*, **85**, 514–518.

[13] Clopper, C.J. and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, **26**, 404–413.

[14] Clunies-Ross, C.W. (1958). Interval estimation for the parameter of a binomial proportions, *Biometrika*, **45**, 275–279.

[15] Cohen, G.R. and Yang, S.Y. (1994). Mid-p confidence intervals for the Poisson expectation, *Statistics in Medicine*, **13**, 2189–2203.

[16] Copas, J.B. (1992). Exact confidence limits for binomial proportions – Brenner and Quan revisited, *The Statistician*, **41**, 569–572.

[17] Crow, E.L. (1956). Confidence intervals for a proportion, *Biometrika*, **43**, 423–435.

[18] Edwardes, M.D. (1998). The evaluation of confidence sets with application to binomial intervals, *Statistica Sinica*, **8**, 393–409.

[19] Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**, 1–26.

[20] Fleiss, J.L.; Levin, B. and Paik, M.C. (2003). *Statistical Methods for Rates and Proportions, Third Edition*, Wiley, New York.

[21] Fujino, Y. (1980). Approximate binomial confidence limits, *Biometrika*, **67**, 677–681.

[22] García-Pérez, M.A. (2005). On the confidence interval for the binomial parameter, *Quality & Quantity*, **39**, 467–481.

[23] Geyer, C.J. and Meeden, G.D. (2005). Fuzzy and randomized confidence intervals and $p-$values (with discussion), *Statistical Science*, **20**, 358–387.

[24] Ghosh, B.K. (1979). A comparison of some approximate confidence intervals for the binomial parameter, *Journal of the American Statistical Association*, **74**, 884–900.

[25] Hjorth, J.S.U. (1994). *Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap*, Chapman and Hall, London.

[26] Johnson, N.L. and Kotz, S. (1969). *Discrete Distributions*, Wiley, New York.

[27] Leemis, L.M. and Trivedi, K.S. (1996). A comparison of approximate interval estimators for the Bernoulli parameter, *The American Statistician*, **50**, 63–68.

[28] Lee, S.-C (2006). Interval estimation of binomial proportions based on weighted Polya posterior, *Computational Statistics and Data Analysis*, **51**, 1012–1021.

[29] Lui, K.-J. (1995). Confidence limits for the prevalence ratio based on the negative binomial distribution, *Statistics in Medicine*, **14**, 1471–1477.

[30] Newcombe, R.G. (1994). Confidence intervals for a binomial proportion, *Statistics in Medicine*, **13**, 1283–1285.

[31] Newcombe, R.G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods, *Statistics in Medicine*, **17**, 857–872.

[32] Pan, W. (2002). Approximate confidence intervals for one proportion and difference of two proportions, *Computational Statistics and Data Analysis*, **40**, 143–157.

[33] Puza, B. and O'neill, T. (2006). Generalised Clopper–Pearson confidence intervals for the binomial proportion, *Journal of Statistical Computation and Simulation*, **76**, 489–508.

[34] Reiczigel, J. (2003). Confidence intervals for the binomial parameter: some new considerations, *Statistics in Medicine*, **22**, 611–621.

[35] Santner, T.J. and Duffy, D.E. (1989). *The Statistical Analysis of Discrete Data*, Springer, New York.

[36] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer, New York.

[37] Sterne, T.E. (1954). Some remarks on confidence or fiducial limits, *Biometrika*, **41**, 275–278.

[38] Stevens, W.L. (1950). Fiducial limits of the parameter of a discontinuous distribution, *Biometrika*, **37**, 117–129.

[39] Vollset, S.E. (1993). Confidence intervals for a binomial proportion, *Statistics in Medicine*, **12**, 809–827.

[40] Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference, *Journal of the American Statistical Association*, **22**, 209–212.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE), 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.

— Three volumes are scheduled for publication, one in March, one in June and the other in November.

— On average, four 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/indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts*, and *Zentralblatt für Mathematic*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: 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.

Submitted 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 not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, 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), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat.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: liliana.martins@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.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.