Statistics and Probability Letters 95 (2014) 48–56

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

The distribution of the maximum of the multivariate AR(p) and multivariate MA(p) processes

Christopher S. Withers[a], Saralees Nadarajah[b,*]

[a] Industrial Research Limited, Lower Hutt, New Zealand
[b] University of Manchester, Manchester M13 9PL, UK

ARTICLE INFO

Article history:
Received 13 June 2014
Received in revised form 4 August 2014
Accepted 8 August 2014
Available online 18 August 2014

Keywords:
Autoregressive
Fredholm kernel
Moving average
Multivariate maximum

ABSTRACT

We give the cumulative distribution function (cdf) of $\mathbf{M}_n$, the (element-wise) maximum of a sequence of n observations from a multivariate AR(p) process. We do the same for a multivariate MA(p) process. Solutions are first given in terms of repeated integrals and then for the case, where the marginal cdf of the observations is absolutely continuous. The cdf of the multivariate maximum $\mathbf{M}_n$ is then given as a weighted sum of the nth powers of the eigenvalues of a non-symmetric Fredholm kernel. The weights are given in terms of the left and right eigenfunctions of the kernel.

1. Introduction

Suppose $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent and identical s-variate observations. The need for the distribution of the following arises in many applied areas:

$$\mathbf{M}_n = \max\left(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\right)$$

for $n \geq 1$, where this is defined as the s-vector with jth element

$$M_{n,j} = \max\left(X_{1,j}, X_{2,j}, \ldots, X_{n,j}\right)$$

and $X_{i,j}$ is the jth element of $\mathbf{X}_i$. Examples include: if $\mathbf{X}_i$ contains the levels of s different rivers on the ith day then $\mathbf{M}_n$ will contain the maximum river levels over n days; if $\mathbf{X}_i$ contains the sea levels at s different ports on the ith day then $\mathbf{M}_n$ will contain the maximum sea levels over n days; if $\mathbf{X}_i$ contains the temperatures at s different cities on the ith day then $\mathbf{M}_n$ will contain the maximum temperatures over n days; if $\mathbf{X}_i$ contains the risks at s different markets on the ith day then $\mathbf{M}_n$ will contain the maximum risks over n days; and so on.

There is considerable theoretical development on the distribution of $\mathbf{M}_n$ (Galambos, 1987; de Haan and Ferreira, 2006; Resnick, 2008). Much of the theory gives the limiting distribution of $\mathbf{M}_n$ as $n \to \infty$. In practice, n is finite, so such theory is not very useful. Furthermore, multivariate data are very limited, so n can never be considered large enough for the limiting distribution to be used as an approximation. So, what is needed are theoretical developments on the *exact* distribution of $\mathbf{M}_n$ for a finite n.

In this note, we shall focus on the distribution of $\mathbf{M}_n$ for multivariate ARMA (autoregressive-moving average) processes. There has been little work on this distribution. We are aware only of the work of Davis et al. (1985), Martins and Ferreira

* Corresponding author.
E-mail addresses: kit.withers@gmail.com (C.S. Withers), mbbsssn2@manchester.ac.uk (S. Nadarajah).

http://dx.doi.org/10.1016/j.spl.2014.08.009

(2005), Klüppelberg and Pergamenchtchikov (2007) and Ferreira and Ferreira (2013). Sen and Tan (2012) provide an excellent review of multivariate extreme value theory for time series. All these papers provide the limiting distributions of $\mathbf{M}_n$ or assume that $\mathbf{X}_i$ come from a specific class. We are aware of no work giving the exact distribution of $\mathbf{M}_n$ for a finite n for multivariate ARMA processes.

This note continues the application of a powerful new method for obtaining the *exact* distribution of extremes of n correlated observations as weighted sums of nth powers of certain eigenvalues. The method was first illustrated for a univariate moving average of order 1 in Withers and Nadarajah (2014a) and a univariate autoregressive process of order 1 in Withers and Nadarajah (2011).

Let $\{\mathbf{e}_i\}$ be independent and identically distributed random variables from some cumulative distribution function (cdf) F on $\mathbb{R}^s$. Let $\left\{\boldsymbol{\rho}_i\right\}$ be $s \times s$ real matrices. For convenience we assume that $\boldsymbol{\rho}_0 = \mathbf{I}$, the identity matrix. In Sections 2 and 4, we consider the MMA(p) process (multivariate moving average process of order p)

$$\mathbf{X}_i = \sum_{j=0}^{p} \boldsymbol{\rho}_j \mathbf{e}_{i-j}. \tag{1}$$

In Sections 3 and 5, we consider the MAR(p) process (multivariate autoregressive process of order p)

$$\mathbf{X}_i - \sum_{j=1}^{p} \boldsymbol{\rho}_j \mathbf{X}_{i-j} = \mathbf{e}_i. \tag{2}$$

We give expressions for the cdf of the multivariate maximum $\mathbf{M}_n$ under the models (1) and (2). The expressions that we give hold for any finite n, even if the limiting cdf of $\mathbf{M}_n$ as $n \to \infty$ does not exist. Many multivariate distributions like the multivariate Poisson and multivariate geometric distributions do not yield non-degenerate forms for the limiting cdf of $\mathbf{M}_n$. Hence, our results are more general and better applicable than the results of classical extreme value theory.

In Sections 2 and 3, the cdf is given in terms of n repeated integrals via recurrence relationships. In Sections 4 and 5, we consider the case when F is absolutely continuous with probability density function (pdf) f with respect to Lebesgue measure on $\mathbb{R}^s$. We show that the integral operators on $\mathbb{R}^{ps}$ used in the previous sections are Fredholm operators. This allows us to express the cdf of $\mathbf{M}_n$ in each case as a weighted sum of nth powers of its eigenvalues, with the weights being given in terms of the left and right eigenfunctions.

The main results of this note are: Theorem 2.1, Remark 2.1 following easily from Theorem 2.1, Theorem 3.1, Remark 3.1 following easily from Theorem 3.1, Theorem 4.1, Theorem 4.2, Remark 4.1 following easily from Theorem 4.2 and Remark 5.1. Example 2.1, Figs. 1–3 and the three paragraphs following Example 2.1 illustrate the results of Theorem 2.1, Remark 2.1, Theorem 4.1 and Theorem 4.2. Example 3.1 and Fig. 4 illustrate the results of Theorem 3.1, Remark 3.1, Remark 5.1.

Set $I(A) = 1$ if A is true, $I(A) = 0$ if A is false, and $\mathbf{M}_0 = -\infty \in \mathbb{R}^s$. For $\mathbf{y}_1, \ldots, \mathbf{y}_p \in \mathbb{R}^s$, set $\mathbf{y}' = \left(\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_p'\right) \in \mathbb{R}^{ps}$ and similarly for $\mathbf{z}'$. For a, b functions on $\mathbb{R}^p$, set $\int a = \int a\left(\mathbf{y}_1\right) d\mathbf{y}_1 = \int_{\mathbb{R}^p} a\left(\mathbf{y}_1\right) d\mathbf{y}_1$ and similarly for functions on $\mathbb{R}^{ps}$ and similarly for $\int ab$. Generally, we shall suppress dependency on $\mathbf{x}$.

2. Solutions for the MMA(p) using repeated integrals

Here, we consider the MMA(p) process (1). Set

$$G_n(\mathbf{y}) = P\left(\mathbf{M}_n \le \mathbf{x},\ \mathbf{e}_{n-j} \le \mathbf{y}_{j+1},\ j = 0, 1, \ldots, p-1\right) \tag{3}$$

for $n \ge 0$. Our goal is to determine $u_n = P\left(\mathbf{M}_n \le \mathbf{x}\right) = G_n(\infty)$, where $\infty \in \mathbb{R}^{ps}$. Theorem 2.1 gives an expression for $G_n(\mathbf{y})$. Remark 2.1 deduces an expression for $G_n(\infty)$ and thus for u_n.

Theorem 2.1. *We have G_n of (3) satisfying the recurrence relation*

$$G_n(\mathbf{y}) = \mathcal{K} G_{n-1}(\mathbf{y})$$

for $n \ge 1$, where

$$\mathcal{K} h(\mathbf{y}) = \int_{\mathbf{z}_1 \le \mathbf{y}_2,\ \ldots,\ \mathbf{z}_{p-1} \le \mathbf{y}_p} F\left(\min\left(\mathbf{y}_1, \mathbf{x} - \sum_{j=1}^{p} \boldsymbol{\rho}_j \mathbf{z}_j\right)\right) dh(\mathbf{z}) \tag{4}$$

and the restriction over the range of integration is removed if $p = 1$. So,

$$G_n(\mathbf{y}) = \mathcal{K}^n G_0(\mathbf{y}) \tag{5}$$

for $n \ge 0$.