



A RESULT CONCERNING QUADRATIC FORMS IN FUNCTIONS OF ORDER STATISTICS

H. F. Coronel-Brizio and A. R. Hernandez-Montoya

Faculty of Physics and Artificial Intelligence

Universidad Veracruzana

Mexico

e-mail: hcoronel@uv.mx

Abstract

A theoretical result concerning the limiting values of quadratic forms in functions of the vector \mathbf{m} , of expected values and V_X , the covariance matrix of the order statistics in a sample of size n from an absolutely continuous distribution F , which satisfy certain regularity conditions, is presented.

1. Introduction

In goodness-of-fit work based on probability plots, it is required to have results concerning limiting values which could be also used as approximations, for large sample sizes, to quadratic forms involving functions of \mathbf{m} and V_X , where \mathbf{m} is the vector of expected values of order statistics $X_{(i)}$ from an absolutely continuous distribution $F(\cdot)$, where V_X denotes the covariance matrix of the $X_{(i)}$.

Received: April 19, 2014; Accepted: May 29, 2014

2010 Mathematics Subject Classification: 62-07, 62F10, 62F12, 62N01, 62N02.

Keywords and phrases: order statistics, normal distribution, covariance matrix, quadratic forms, asymptotics.

Suppose that the random variable Y has an absolutely continuous distribution $F(y)$ in the location-scale family, that is, $F(y)$ is of the form $F_0(x)$ with $x = (y - \alpha)/\beta$, with $F_0(x)$ completely specified. Here, α and β are the location and scale parameters, respectively. Let $X_{(1)}, \dots, X_{(n)}$ be the order statistics of a random sample of X of size n . Let $m_i = E(X_{(i)})$ and let V_X be the covariance matrix of the vector $\mathbf{X}^\top = (X_{(1)}, \dots, X_{(n)})$.

If a sample of n Y -values is given, the order statistics $Y_{(1)}, \dots, Y_{(n)}$ can be represented as $Y_{(i)} = \alpha + \beta X_{(i)}$, and then

$$E(Y_{(i)}) = \alpha + \beta m_i. \quad (1)$$

The covariance matrix of the vector $\mathbf{Y}^\top = (Y_{(1)}, \dots, Y_{(n)})$ is $\beta^2 V_X$.

2. Covariances of Linear Combinations

Consider two linear combinations of the order statistics in a sample of size n from an absolutely continuous distribution F , defined by

$$T_g = \frac{1}{n} \sum_{i=1}^n g(m_i) Y_{(i)} \quad \text{and} \quad T_h = \frac{1}{n} \sum_{i=1}^n h(m_i) Y_{(i)}.$$

Let \mathbf{g} and \mathbf{h} be the vectors whose i th components are, respectively, $g(m_i)$ and $h(m_i)$. Then, $\text{Cov}(T_g, T_h)$, the covariance between T_g and T_h is given by

$$n\text{Cov}(T_g, T_h) = \beta^2 \frac{\mathbf{g}^\top V_X \mathbf{h}}{n}.$$

Since β^2 is only a multiplying factor, it will be assumed that $\beta = 1$.

We shall examine asymptotic limits of the form $\frac{\mathbf{g}^\top V_X \mathbf{h}}{n}$ and present a theorem whose proof is given in the Appendix. Unpublished prior versions of the proof can be found in Coronel-Brizio [5] and Coronel-Brizio and Stephens [6].

Theorem 2.1. Let \mathbf{g} and \mathbf{h} be two n -dimensional vectors with components $g_i = g(m_i)$ and $h_i = h(m_i)$, respectively, for $i = 1, \dots, n$ and let $G(x) = \int g(x)dx$, $H(x) = \int h(x)dx$ denote the antiderivatives of the functions g and h , respectively. If an absolutely continuous distribution function F satisfies:

$$\begin{aligned} \lim_{x \rightarrow -\infty} G(x)F(x) &= \lim_{x \rightarrow \infty} G(x)[1 - F(x)] = 0, \\ \lim_{x \rightarrow -\infty} H(x)F(x) &= \lim_{x \rightarrow \infty} H(x)[1 - F(x)] = 0, \end{aligned}$$

then

$$\lim_{n \rightarrow \infty} \frac{\mathbf{u}^\top V_X \mathbf{v}}{n} = \text{Cov}[G(X), H(X)]. \quad (2)$$

Table 1. Values of expression (2.2) for selected distributions and values of r and s

Distribution	r	s	$\lim_{n \rightarrow \infty} n^{-1}(\mathbf{m}^r)^\top V_X \mathbf{m}^s$
Normal	1	1	1/2
	2	2	5/3
	3	3	6
	4	4	189/5
	0	2	1
	0	4	3
	0	6	15
	0	8	105
Logistic	0	1	$\pi^2/3$
	0	2	$7\pi^4/45$
	1	1	$4\pi^4/45$

Extreme value	1	3	$52\pi^6/315$
	2	2	$31\pi^6/189$
	3	3	$116\pi^8/225$
	0	0	$\pi^2/6$
	0	1	$-\pi^2\gamma/6 - \zeta(3)$
	1	2	$\pi^2\gamma^2/6 + 2\zeta(3)\gamma + \pi^4/20$

An important case which has applications in estimation and tests of fit occurs when $g(t) = t^r$ and $h(t) = t^s$, where r and s are non-negative integers. In this situation, we have the following:

Corollary 2.2.

$$\lim_{n \rightarrow \infty} \frac{(\mathbf{m}^r)^\top V_X \mathbf{m}^s}{n} = \frac{\mu'_{r+s+2} - \mu'_{r+1}\mu'_{s+1}}{(r+1)(s+1)}, \quad (3)$$

where μ'_j denotes the j th central moment about the origin of the random variable X , and $\mathbf{m}^r = (m_1^r, \dots, m_n^r)^\top$.

The above result is an extension to second order moments of the results given in Burr [3].

3. Examples

In order to illustrate the use of the above result, some useful limits in parameter estimation from probability plots for some standard distributions are presented.

The normal density is the well-known function $f(x) = 1/2 \frac{e^{-1/2x^2} \sqrt{2}}{\sqrt{\pi}}$, $-\infty < x < \infty$; the functional form of the logistic distribution is $F(x) =$

$e^x/(1 + e^x)$, $-\infty < x < \infty$; due to the symmetry of these two distributions, $\lim_{n \rightarrow \infty} \frac{(\mathbf{m}^r)^\top V_X \mathbf{m}^s}{n} = 0$ whenever $r + s$ is odd. The type I extreme value distribution has density $F(x) = 1 - \exp(-e^x)$, $-\infty < x < \infty$.

Some particular results are presented in Table 1, where $\gamma \approx 0.5772$ is Euler's constant and $\zeta(n)$ denotes the Riemann's Zeta function.

It can be seen that application of the (2.2) reproduces some of the relations between some operations over the columns of the stochastic matrix for the case of the normal distribution; e.g., $\frac{(\mathbf{1})^\top V_X \mathbf{m}^{2k}}{n} = E(X^k)$.

4. Applications

Goodness of fit tests for a distribution in the location-scale family, based on the use of polynomial regression models, have been developed for the case of the normal distribution, (LaBrecque [9], see also Stephens [13], Puri and Rao [10]) and for the logistic and type I extreme value distributions (Coronel-Brizio [5]).

The tests are based on fitting, alternatively to the model (1), an extended model of the form

$$E(Y_{(i)}) = \alpha_0 \psi_0(m_i) + \alpha_1 \psi_1(m_i) + \alpha_2 \psi_2(m_i) + \cdots + \alpha_p \psi_p(m_i), \quad (4)$$

where $\alpha_0, \alpha_1, \dots, \alpha_p$ are constants and $\psi_j(m_i)$ is a polynomial of degree j and the coefficients of the polynomials are chosen so that

$$\psi_i^\top V_X^{-1} \psi_j = 0 \text{ for } i \neq j.$$

Without loss of generality, we can take $\psi_0(m_i) = 1$ and $\psi_1(m_i) = m_i$; then in (1) $\alpha = \alpha_0$ and $\beta = \alpha_1$, and the null hypothesis that the model is (1) can be made by testing

$$H_0 : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0.$$

Usually the test is carried out by fitting a quadratic or cubic model so that $p \leq 3$.

Expression (2.2) is of practical and theoretical interest in the calculation of the asymptotic variances of the estimated parameters, in particular when this is done by ordinary least squares.

5. Conclusions

The result presented here, can be considered as an extension to second order moments of those given by Burr [3] and closely connected with the work by Chernoff et al. [4] however, it does not seem to have been discussed before, at least in an explicit form.

6. Appendix: Proof of Theorem 2.1

Let us consider the quantity

$$\frac{\mathbf{g}'V_X\mathbf{h}}{n} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n g(m_i)h(m_j)v_{ij}.$$

Using Cramer's asymptotic approximation for covariances of order statistics, it is not difficult to obtain the limiting expression:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\mathbf{u}'V_X\mathbf{v}}{n} &= \int_{-\infty}^{\infty} \int_x^{\infty} g(x)h(y)F(x)[1 - F(y)]dydx \\ &+ \int_{-\infty}^{\infty} \int_y^{\infty} g(x)h(y)F(y)[1 - F(x)]dydx. \end{aligned} \quad (5)$$

We now must show that the above sum of integrals equals the right hand side of (2). For simplicity, in the following the arguments of the functions will be omitted.

Let us now find the value of

$$\int_x^{\infty} H dF.$$

Using two different expressions for the antiderivative of f , namely F and $-[1 - F]$, integration by parts gives:

$$\int_x^\infty HdF = H[1 - F]|_x^\infty + \int_x^\infty [1 - F]dH. \quad (6)$$

Similarly,

$$\int_{-\infty}^x HdF = HF|_{-\infty}^x - \int_{-\infty}^x FdH. \quad (7)$$

Multiplying (6) and (7) by the integral expression of $E[G]$, we obtain:

$$\begin{aligned} \int_{-\infty}^\infty GdF \int_x^\infty HdF &= \int_{-\infty}^\infty GH[1 - F]dF \\ &\quad + \int_{-\infty}^\infty GdF \int_x^\infty [1 - F]dH \\ \int_{-\infty}^\infty GdF \int_{-\infty}^x HdF &= \int_{-\infty}^\infty GHFdF \\ &\quad - \int_{-\infty}^\infty GdF \int_{-\infty}^x FdH. \end{aligned}$$

Adding the above two equations we have that

$$\begin{aligned} E[G]E[H] &= E[GH] + \int_{-\infty}^\infty GdF \int_x^\infty [1 - F]dH \\ &\quad - \int_{-\infty}^\infty GdF \int_{-\infty}^x FdH. \end{aligned} \quad (8)$$

Changing the order of integration above, the two integrals are, respectively, equal to

$$\int_{-\infty}^\infty GF[1 - F]dH - \int_{-\infty}^\infty \int_x^\infty F[1 - F]dGdH \quad (9)$$

and

$$\int_{-\infty}^{\infty} GF[1-F]dH + \int_{-\infty}^{\infty} \int_y^{\infty} F[1-F]dGdH. \quad (10)$$

Finally, substitution of (9) and (10) into (8) gives the right hand side of (2). \square

Acknowledgements

This work has been supported by National Council for Science and Technology of the United Mexican States under project grant numbers 155492 and 135297. The authors also thank PROMEP and the National Researcher's System for continued support.

References

- [1] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, M. Abramowitz and Irene A. Stegun, eds., National Bureau of Standards, 1964.
- [2] M. M. Ali and L. K. Chan, On Gupta's estimates of the parameters of the normal distribution, *Biometrika* 51 (1964), 498-501.
- [3] I. W. Burr, Cumulative frequency functions, *Ann. Math. Statist.* 13 (1942), 215-232.
- [4] H. Chernoff, J. L. Gastwirth and M. V. Johns, Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation, *Ann. Math. Statist.* 38 (1967), 52-73.
- [5] H. F. Coronel-Brizio, Regression tests of fit and some comparisons, Ph.D. Thesis, Department of Mathematics and Statistics, Simon Fraser University, 1994.
- [6] H. F. Coronel-Brizio and M. A. Stephens, Asymptotic limits for covariances of linear combinations of order statistics, Tech. Report, Department of Mathematics and Statistics, Simon Fraser University, 1994.
- [7] A. K. Gupta, Estimation of the mean and standard deviation of a normal population from a censored sample, *Biometrika* 39 (1952), 260-273.
- [8] S. S. Gupta, A. S. Querishi and B. K. Shah, Best linear unbiased estimators of the parameters of the logistic distribution using order statistics, *Technometrics* 9 (1967), 43-56.

- [9] J. LaBrecque, Goodness-of-fit tests based on nonlinearity in probability plots, *Technometrics* 19(3) (1977), 293-306.
- [10] M. L. Puri and C. R. Rao, Augmenting Shapiro-Wilk Test for Normality, *Contributions to Applied Statistics*, Birkhauser (Grossenhaus), Berlin, 1976, pp. 129-139.
- [11] S. S. Shapiro and R. S. Francia, Approximate analysis of variance test for normality, *J. Amer. Statist. Assoc.* 67 (1972), 215-216.
- [12] S. S. Shapiro and M. B. Wilk, An analysis of variance test for normality, *Biometrika* 52 (1965), 591-611.
- [13] M. A. Stephens, Asymptotic properties for covariance matrices of order statistics, *Biometrika* 62 (1975), 23-28.
- [14] M. A. Stephens, Asymptotic calculations of functions of expected values and covariances of order statistics, *Canad. J. Statist.* 18(3) (1990), 265-270.
- [15] S. M. Stigler, Linear functions of order statistics, *Ann. Math. Statist.* 40 (1969), 770-788.
- [16] S. M. Stigler, Linear functions of order statistics with smooth weight functions, *Ann. Statist.* 2 (1974), 676-693. Correction 7, 466.