Asymptotic linearity and limit distributions, approximations

João T. Mexia [a], Manuela M. Oliveira [b],*

[a] FCT Nova University of Lisbon, Department of Mathematics, Quinta da Torre, 2825 Monte da Caparica, Portugal
[b] University of Èvora, Department of Mathematics and CIMA – Center for Research on Mathematics and its Applications, Colégio Luís António Verney, Rua Romão Ramalho 59, 7000-671 Évora, Portugal

ARTICLE INFO

ABSTRACT

Linear and quadratic forms as well as other low degree polynomials play an important role in statistical inference. Asymptotic results and limit distributions are obtained for a class of statistics depending on $\mu + \mathbf{X}$, with $\mathbf{X}$ any random vector and μ non-random vector with $\|\mu\| \to +\infty$. This class contain the polynomials in $\mu + \mathbf{X}$. An application to the case of normal $\mathbf{X}$ is presented. This application includes a new central limit theorem which is connected with the increase of non-centrality for samples of fixed size. Moreover upper bounds for the suprema of the differences between exact and approximate distributions and their quantiles are obtained.

Contents

1. Introduction

Linear and quadratic forms as well as other low degree polynomials play an important role in statistics (e.g. Ito, 1980; Lazraq and Cleroux, 2001; Oliveira and Mexia, 2007a, 2007b; Sabatier and Traissac, 1994; Scheffé, 1959). In Areia et al. (2008) it is pointed out that low degree polynomials in normal independent variables, with sufficiently small variation coefficients, are approximately normal. Moreover, also in Areia et al. (2008), are presented simulations suggesting that, when $\|\mu\| \to +\infty$, a polynomial $P(\mu + \mathbf{X})$ has a leading component linear in $\mathbf{X}$. This linear component will be normal.

We will obtain asymptotic approximations and limit distributions for a class of statistics containing $P(\mu + \mathbf{X})$.

In the next section we will present the concept of asymptotic linear functions which will play a central role in our study. We will also introduce the relevant notations and point out that polynomials are asymptotically linear functions. Next we

---

use a variant of the δ method (e.g. Lehmann and Casella, 2001) to obtain asymptotic and limit results. In the last section we will consider the case when $\mu + \mathbf{X}$ will be normal with mean vector μ. This variant will display a good control of the approximation errors both for distribution functions and their quantiles.

If $\mathbf{V}^+$ is the Moore–Penrose matrix inverse of the variance–covariance matrix $\mathbf{V}$ of $\mu + \mathbf{X}$, the quadratic form $(\mu + \mathbf{X})^t \mathbf{V}^+ (\mu + \mathbf{X})$ will have a chi-square distribution with $r = rank(\mathbf{V})$ degrees of freedom and non-centrality parameter $\delta = \mu^t \mathbf{V}^+ \mu$, (e.g. Mexia, 1990). Thus we may use δ to measure the non-centrality of the sample. In our study of the normal case we will obtain a central limit theorem, for fixed size samples whose non-centrality parameter diverges to $+\infty$. Moreover δ may be used to measure sample non-centrality whenever $\mathbf{V}$, the variance–covariance matrix of $\mathbf{X}$ is defined, and $\delta \longrightarrow +\infty$ will imply $\|\mu\| \to +\infty$. Since our asymptotic and limit results are derived assuming that $\|\mu\| \to +\infty$, they are connected with the increase of non-centrality in samples, normal or not, with fixed size. When the components $X_1, \ldots, X_m$ of $\mathbf{X}$ are i.i.d. with mean value μ and variance σ^2 we have $\mu = \mu \mathbf{1}_m$ and $\mathbf{V} = \sigma^2 \mathbf{I}_m$ so $\mathbf{V}^+ = (1/\sigma^2) \mathbf{I}_m$ and $\delta = m(\mu^2/\sigma^2) = m/VC^2$, with $VC = \sigma/\mu$ the variation coefficient. Thus when the $X_1, \ldots, X_m$ are observations with low variation coefficients we have a sample with large δ. The variation coefficient with high precision observations are used.

2. Notations and concepts

Given a sufficiently regular function $g : R^k \to R$, let $\underline{g}$ and $\underline{\underline{g}}$ be the gradient and the Hessian matrix of g. Then, with $r_d(\mathbf{x})$ the supremum of the spectral radius $r(\mathbf{y})$ of $\underline{\underline{g}}(\mathbf{y})$, when $\|\mathbf{y} - \overline{\overline{\mathbf{x}}}\| \leq d$, the function g is asymptotically linear, if whatever $d > 0$, we have

$$k_d(u) \underset{u \to \infty}{\to} 0,$$

with

$$k_d(u) = sup \left\{ \frac{r_d(\mathbf{x})}{\|\underline{g}(\mathbf{x})\|}; \|\mathbf{x}\| \geq u \right\}.$$

Since the first and second order partial derivatives of polynomials are themselves polynomials with degree decreasing with successive derivations, polynomials will be asymptotically linear.

Given a random vector $\mathbf{X}$ and an asymptotically linear function $g(\cdot)$, we will show that, when $\|\mu\| \to \infty$, the distribution F_Y of

$$Y = g(\mu + \mathbf{X})$$

approaches the distribution of

$$Z = g(\mu) + \underline{g}(\mu)^t \mathbf{X}.$$

With l_p the p th quantile of the random variable L, given

$$Z^\circ = \frac{(Z - g(\mu))}{\|\underline{g}(\mu)\|}$$

and

$$Y^\circ = \frac{(Y - g(\mu))}{\|\underline{g}(\mu)\|}$$

we will obtain upper bounds for $|y_p^\circ - z_p^\circ|$. Moreover, when

$$\frac{1}{\|\underline{g}(\mu)\|} \underline{g}(\mu) \underset{\|\mu\| \to \infty}{\to} \mathbf{b}$$

we will establish that F_{Z° converges to $F_{Z_\circ^\circ}$, with $Z_\circ^\circ = \mathbf{b}^t \mathbf{X}$.

As we shall see these results are easy to apply when $\mathbf{X}$ is normal. Then $Z_\circ^\circ$ will also be normal and a central limit theorem will apply.

3. Asymptotic results

If $g : R^k \to R$ is sufficiently regular we have, see Khuri (2003)

$$Y = g(\mu + \mathbf{X}) = g(\mu) + \underline{g}(\mu)^t \mathbf{X} + \int_0^1 (1 - h)(\mathbf{X}^t \underline{\underline{g}}(\mu + h\mathbf{X})\mathbf{X}) \, dh.$$

Thus, when $\|\mathbf{X}\| \leq d$,

$$|Y - Z| = \left| \int_0^1 (1 - h)(\mathbf{X}^t \underline{\underline{g}}(\mu + h\mathbf{X})\mathbf{X}) \, dh \right| \leq r_d(\mu)d^2$$

as well as

$$|Y^\circ - Z^\circ| \leq k_d(\|\boldsymbol{\mu}\|)d^2.$$

We now establish

Proposition 3.1. *If $g(\cdot)$ is asymptotically linear, then $|Y^\circ - Z^\circ| \underset{\|\mu\| \to \infty}{\overset{a.s}{\to}} 0$, where a.s. indicates almost sure convergence.*

Proof. If $g(\cdot)$ is asymptotically linear, whatever $\varepsilon > 0$ and $d > 0$, there is $u(\varepsilon)$ such that, for $u > u(\varepsilon)$, $k_d(u) < \varepsilon$. Thus, when $\|\boldsymbol{\mu}\| \geq u(\varepsilon)$, $\|\mathbf{X}\| < d$ implies $|Y^\circ - Z^\circ| \leq \varepsilon$. Whatever $\delta > 0$, we may choose $d > 0$ to ensure that $Pr(\|\mathbf{X}\| < d) \geq 1 - \delta$, thus when $\|\boldsymbol{\mu}\| \geq u(\varepsilon)$, we will also have $Pr(|Y^\circ - Z^\circ| \leq \varepsilon) \geq 1 - \delta$ which establishes the thesis. $\square$

Since $k_d(u)$ decreases with u we may define

$$u(d, \varepsilon) = Min\{u; k_d(u)d^2 \leq \varepsilon\}$$

thus, with l_q the p th quantile of the distribution of $\mathbf{L} = \|\mathbf{X}\|^2$, whenever $\mathbf{L} \leq l_q$, $\|\boldsymbol{\mu}\| \geq u_q(\varepsilon) = u(l_q, \varepsilon)$ implies

$$|Y^\circ - Z^\circ| \leq \varepsilon.$$

So, if $q(\varepsilon)$ is the probability of the event $A(\varepsilon) = \{|Y^\circ - Z^\circ| \leq \varepsilon\}$ we see that $q(\varepsilon) \geq q$, when $\|\boldsymbol{\mu}\| > u_q(\varepsilon)$.

Let $A^c(\varepsilon)$ be the complement of $A(\varepsilon)$ and $q^c(\varepsilon) = 1 - q(\varepsilon)$. We now have

Proposition 3.2. *Whatever z and $\varepsilon > 0$, $F_{Z^\circ}(z - \varepsilon) - q^c(\varepsilon) \leq F_{Y^\circ}(z) \leq F_{Z^\circ}(z + \varepsilon) + q^c(\varepsilon)$.*

Proof. The thesis follows from

$$F_{Y^\circ}(z) = Pr(Y^\circ \leq z) = q(\varepsilon)Pr(Y^\circ \leq z|A(\varepsilon)) + q^c(\varepsilon)Pr(Y^\circ \leq z|A^c(\varepsilon)) \leq q(\varepsilon)Pr(Z^\circ \leq z + \varepsilon|A(\varepsilon)) + q^c(\varepsilon)$$

and from

$$F_{Z^\circ}(z - \varepsilon) - q^c(\varepsilon) \leq q(\varepsilon)Pr(Z^\circ \leq z - \varepsilon|A(\varepsilon)) \leq q(\varepsilon)Pr(Y^\circ \leq z|A(\varepsilon)) \leq Pr(Y^\circ \leq z) = F_{Y^\circ}(z). \quad \square$$

Corollary 3.1. *When $F_{Z^\circ}(z)$ has density $f_{Z^\circ}(z)$ bounded by C, we have $Sup\{|F_Y - F_Z|\} = Sup\{|F_{Y^\circ} - F_{Z^\circ}|\} \leq d(\varepsilon)$ with $d(\varepsilon) = 2(C\varepsilon + q(\varepsilon))$.*

Proof. Since

$$F_Y(x) = F_{Y^\circ}\left(\frac{x - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|}\right)$$

and

$$F_Z(x) = F_{Z^\circ}\left(\frac{x - g(\boldsymbol{\mu})}{\|\underline{g}(\boldsymbol{\mu})\|}\right)$$

we have $Sup\{|F_Y - F_Z|\} = Sup\{|F_{Y^\circ} - F_{Z^\circ}|\}$. To complete the proof we have only to apply Proposition 3.1 remembering that, since f_{Z° is bounded by C, $F_{Z^\circ}(z + \varepsilon) - F_{Z^\circ}(z - \varepsilon) \leq 2C\varepsilon$. $\square$

Corollary 3.2. *When f_{Z° is bounded by C and $g(\cdot)$ is asymptotically linear, $Sup\{|F_Y - F_Z|\} \underset{\|\mu\| \to +\infty}{\to} 0$.*

This last corollary gives conditions for F_Z to approach F_Y uniformly when $\|\boldsymbol{\mu}\| \longrightarrow \infty$.

Next we have

Proposition 3.3. *If f_{Z° is bounded by C and if $f_{Z^\circ}(z) \geq m(\delta) > 0$, whenever $z^\circ_{\delta - d(\varepsilon)} \leq z \leq z^\circ_{1 - \delta + d(\varepsilon)}$, we have*

$$Sup\{|y^\circ_p - z^\circ_p|; \delta \leq p \leq 1 - \delta\} \leq 2\left(\varepsilon + \frac{d(\varepsilon)}{m(\delta)}\right).$$

Proof. According to Corollary 3.1 of Proposition 3.2, we have

$$F_{Z^\circ}(y^\circ_p - \varepsilon) - d(\varepsilon) \leq F_{Y^\circ}(y^\circ_p) = p \leq F_{Z^\circ}(y^\circ_p + \varepsilon) + d(\varepsilon)$$

so $F_{Z^\circ}(y^\circ_p - \varepsilon) \leq p + d(\varepsilon)$ and $p - d(\varepsilon) \leq F_{Z^\circ}(y^\circ_p + \varepsilon)$. Thus $y^\circ_p - \varepsilon \leq z_{p+d(\varepsilon)}$ and $z^\circ_{p-d(\varepsilon)} \leq y^\circ_p + \varepsilon$, so $z^\circ_{p-d(\varepsilon)} - \varepsilon \leq y^\circ_p \leq z^\circ_{p+d(\varepsilon)} + \varepsilon$. To complete the proof we have only to point out that

$$z^\circ_{p+d(\varepsilon)} - z^\circ_{p-d(\varepsilon)} \leq \frac{2d(\varepsilon)}{m(\delta)}. \quad \square$$

Corollary 3.3. *If $\|\boldsymbol{\mu}\| \geq u_q(\varepsilon)$, f_{Z° is bounded by C and $f_{Z^\circ}(z) \geq m(\delta) > 0$, then whenever $z^\circ_{\delta - d(\varepsilon)} \leq z \leq z^\circ_{1 - \delta + d(\varepsilon)}$, we have*

$$Sup\{|y^\circ_p - z^\circ_p|, \delta \leq p \leq 1 - \delta\} \leq 2(\varepsilon + d_q(\varepsilon))m(\delta)),$$

with $d_q(\varepsilon) = 2(C\varepsilon + q^c)$ where $q^c = 1 - q$.

Proof. The thesis follows from Proposition 3.3 since $u_q(\varepsilon) > q$, when $\|\boldsymbol{\mu}\| > u_q(\varepsilon)$. □

We also have

Proposition 3.4. *If*

$$\frac{1}{\|\underline{g}(\boldsymbol{\mu})\|} g(\boldsymbol{\mu}) \underset{\|\mu\| \to \infty}{\to} \mathbf{b}, \quad F_{Z^\circ} \underset{\|\mu\| \to \infty}{\to} F_{Z_\circ}$$

this convergence being uniform when f_{Z° is bounded by C and $g(.)$ is asymptotically linear.

Proof. The first part of the thesis follows from

$$\left(\frac{1}{\|\underline{g}(\boldsymbol{\mu})\|} g(\boldsymbol{\mu}) - \mathbf{b} \right)^t \quad \mathbf{X} \underset{\|\mu\| \to \infty}{\overset{a.s.}{\to}} 0$$

whenever

$$\frac{1}{\|\underline{g}(\boldsymbol{\mu})\|} g(\boldsymbol{\mu}) \underset{\|\mu\| \to \infty}{\to} \mathbf{b}.$$

The second part of the thesis follows from Proposition 3.3 since $u_q(\varepsilon) > q$, when $\|\boldsymbol{\mu}\| > u_q(\varepsilon)$. □

4. Application: the normal case

Let $\mathbf{X}$ be normal with null mean vector and regular variance–covariance matrix $\mathbf{M}$. Then

$$f_{Z^\circ}(z) = \frac{e^{-\|\underline{g}(\boldsymbol{\mu})\|^2 (z - g(\boldsymbol{\mu}))^2 / \underline{g}(\boldsymbol{\mu})^t \mathbf{M} \underline{g}(\boldsymbol{\mu})}}{\sqrt{2\pi \underline{g}(\boldsymbol{\mu})^t \mathbf{M} \underline{g}(\boldsymbol{\mu})}} \|\underline{g}(\boldsymbol{\mu})\|$$

is bounded by

$$C(\boldsymbol{\mu}) = \frac{\|\underline{g}(\boldsymbol{\mu})\|}{\sqrt{2\pi \underline{g}(\boldsymbol{\mu})^t \mathbf{M} \underline{g}(\boldsymbol{\mu})}}.$$

If θ_k is the smallest eigenvalue of $\mathbf{M}$ we have

$$C(\boldsymbol{\mu}) \leq C = \frac{1}{\sqrt{2\pi\theta_k}}.$$

With x_p the quantile for probability p of the standardized normal density, the quantiles for f_{Z° will be

$$z_p^\circ(\boldsymbol{\mu}) = g(\boldsymbol{\mu}) + x_p \frac{\sqrt{g(\boldsymbol{\mu})^t \mathbf{M} \underline{g}(\boldsymbol{\mu})}}{\|\underline{g}(\boldsymbol{\mu})\|},$$

while the minimum of f_{Z° in $[z_{\delta - d(\varepsilon)}^\circ(\boldsymbol{\mu}); z_{1-\delta+d(\varepsilon)}^\circ(\boldsymbol{\mu})]$ will be

$$m(\delta|\boldsymbol{\mu}) = \frac{e^{-x_{\delta - d(\varepsilon)}^2 / 2}}{\sqrt{2\pi \underline{g}(\boldsymbol{\mu})^t \mathbf{M} \underline{g}(\boldsymbol{\mu})}} \|\underline{g}(\boldsymbol{\mu})\|$$

so that

$$\frac{C(\boldsymbol{\mu})}{m(\delta|\boldsymbol{\mu})} = e^{x_{\delta - d(\varepsilon)}^2 / 2}$$

and

$$\frac{d(\varepsilon|\boldsymbol{\mu})}{m(\delta|\boldsymbol{\mu})} = \frac{2(C(\boldsymbol{\mu})\varepsilon + q^c(\varepsilon))}{m(\delta|\boldsymbol{\mu})} = 2e^{x_{\delta - d(\varepsilon)}^2 / 2} \left(\varepsilon + \frac{q^c(\varepsilon)\|\underline{g}(\boldsymbol{\mu})\|}{\sqrt{2\pi \underline{g}(\boldsymbol{\mu})^t \mathbf{M} \underline{g}(\boldsymbol{\mu})}} \right) \leq 2e^{x_{\delta - d(\varepsilon)}^2 / 2} \left(\varepsilon + \frac{q^c(\varepsilon)}{\sqrt{2\pi\theta_k}} \right).$$

We thus obtain bounds that do not depend on $\boldsymbol{\mu}$, both for $f_{Z^\circ}(z)$ and for $d(\varepsilon|\boldsymbol{\mu})/m(\delta|\boldsymbol{\mu})$. It is now easy to apply the previous results to this case.

Moreover if $\mathbf{X}$ is normal with null mean vector and variance–covariance matrix $\mathbf{M}$ and if $\theta_1 \geq, \ldots, \geq \theta_h$ are the eigenvalues of $\mathbf{M}$ with multiplicities $g_1, \ldots, g_h$,

$$\mathbf{L} = \|\mathbf{X}\|^2$$

may be written (see Imhof, 1961) as a linear combination $\sum_{j=1}^h \theta_j \chi_{g_j}^2$, of independent chi-square variates. When $\mathbf{M}$ is known it is easy to compute the quantiles l_p (Fonseca et al., 2007).

5. Final remarks

In this work we derived asymptotic and limit results for samples, with fixed size, whose non-centrality diverges to $+\infty$. In the case of normal samples a central limit theorem of a new type is established. Our approach is a variant of the δ method. This variant will apply to asymptotically linear statistics $g(\boldsymbol{\mu} + \mathbf{X})$ such as the polynomials $P(\boldsymbol{\mu} + \mathbf{X})$ and leads to a good control of the approximation involved (see Corollary 3.1, Proposition 3.2 and Corollary 3.3).

References

Areia, A., Oliveira, M.M., Mexia, J.T., 2008. Models for series of studies based on geometrical representation. Statistical Methodology 1, 277–288.

Fonseca, M., Mexia, J.T., Zmyślony, R., 2007. Jordan Algebras. Generating pivot variables and orthogonal normal models. Journal of Interdisciplinary Mathematics 1, 305–326.

Imhof, J.P., 1961. Computing the distribution of quadratic forms in normal variables. Biometrics 48 (3–4), 419–426.

Ito, K., 1980. Robustness of anova and manocova test procedures. In: Krishnaiah, P.R. (Ed.), Handbook of Statistics, vol. I. North Holland, Amsterdam.

Khuri, A.I., 2003. Advanced Calculus with Applications in Statistics, second ed. Wiley, New York.

Lazraq, A., Cleroux, R., 2001. The PLS multivariate regression model: testing the significance of successive PLS components. Journal of Chemometrics 15, 523–536.

Lehmann, E.L., Casella, G., 2001. Theory of Point Estimation. Springer, Berlin.

Mexia, J.T., 1990. Best linear unbiased estimates, duality of F tests and the Scheffé multiple comparison method in presence of controlled heteroscedasticity. Computational Statistics & Data Analysis 10, 3.

Oliveira, M.M., Mexia, T.J., 2007a. ANOVA like analysis of matched series of studies with a common structure. Journal of Statistical Planning and Inference 137, 1862–1870.

Oliveira, M.M., Mexia, J.T., 2007b. Modelling series of studies with a common structure. Computational Statistics and Data Analysis 51, 5876–5885.

Sabatier, R., Traissac, P., 1994. The ACT (STATIS method). Computational Statistics & Data Analysis 18, 97–119.

Scheffé, H., 1959. The Analysis of Variance. Wiley, New York.