Chi-square and F Tests with Models Close to the Normal

A. García-Pérez

[1] Departamento de Estadística (Ciencias), U.N.E.D., Paseo Senda del Rey 9, 28040-Madrid, Spain

Keywords: Robustness in hypotheses testing, von Mises expansions, Saddlepoint approximations.

1 Introduction

Saddlepoint approximations of von Mises expansions (VOM+SAD approximations) have been used in robustness for hypotheses testing by Field and Ronchetti (1985) and recently by García-Pérez (2003).

The basic idea of the VOM+SAD approximations is, firstly, to use the first two terms of the von Mises expansion (VOM approximation) of the considered functional (p-value $p_n^F = P_F\{T_n > t\}$, or critical value k_n^F) at a model F, that will depend on the Tail Area Influence Function (TAIF) of Field and Ronchetti in another model G, integrated with respect to F,

$$p_n^F \simeq p_n^G + \int \mathrm{TAIF}(x;t;T_n,G)\,dF(x) \qquad , \qquad k_n^F \simeq k_n^G + \frac{1}{g_n(k_n^G)} \int \mathrm{TAIF}(x;k_n^G;T_n,G)\,dF(x)$$

where g_n is the density of the test statistic T_n under G. Then, we approximate this TAIF with a saddlepoint approximation to obtain the VOM+SAD approximations.

We have a very interesting situation when distribution G is the normal distribution $\Phi_{\mu,\sigma}$. In this case, we will have χ^2, Student's t- and F-tests and the approximations will be very accurate if F is close to the normal, a common situation in robustness studies. Moreover, the VOM+SAD approximations take a very simple form. This allows us to study, for example, their Robustness of Validity; or that the critical values keep some tail ordering of the underlying models, etc. Although the results and situations are very diverse, we next give some of them. In the presentation, some interesting examples for different models (as contaminated normal models) and some *Robustness of Validity Plots* (diagrams of the nominal level versus the actual level) will be shown.

2 χ^2 tests

With χ^2 tests it is possible to obtain accurate VOM approximations (see García-Pérez, 2004a),

Proposition 2.1 *The VOM approximations for the p-value and the critical value of a χ^2_n test ($n>1$) under a model F for which the integrals exist are, respectively*

$$p_n^F \simeq n \int_{-\infty}^{\infty} P\{\chi^2_{n-1} > t - (\tfrac{x-\mu}{\sigma})^2\}\,dF(x) - (n-1)P\{\chi^2_n > t\}$$

and

$$k_n^F \simeq \chi^2_{n;\alpha} + \frac{n}{g_{\chi^2_n}(\chi^2_{n;\alpha})}\left[\int_{-\infty}^{\infty} P\{\chi^2_{n-1} > \chi^2_{n;\alpha} - (\tfrac{x-\mu}{\sigma})^2\}\,dF(x) - \alpha\right]$$

where $\chi^2_{n;\alpha}$ is the $(1-\alpha)$-quantile and $g_{\chi^2_n}$ the density function of a χ^2_n distribution.

Also VOM+SAD approximations, from which it is possible to obtain analytic expressions,

Proposition 2.2 *The VOM+SAD approximations for the p-value and the critical value of a χ^2_n test under a model F for which the integrals exist are, respectively*

$$p_n^F \simeq P\left\{\chi^2_n > t\right\} - B + B\int_{-\infty}^{\infty} \frac{\sqrt{n}}{\sqrt{t}}\, e^{\frac{(t-n)(x-\mu)^2}{2t\sigma^2}}\,dF(x)$$

and $$k_n^F \simeq \chi^2_{n;\alpha} + \frac{2\,n\,\chi^2_{n;\alpha}}{\chi^2_{n;\alpha} - n}\left[\int_{-\infty}^{\infty} \frac{\sqrt{n}}{\sqrt{\chi^2_{n;\alpha}}}\, e^{\frac{(\chi^2_{n;\alpha}-n)(x-\mu)^2}{2\chi^2_{n;\alpha}\sigma^2}}\, dF(x) - 1\right]$$

where $B = \frac{n\sqrt{n}}{\sqrt{\pi}\,(t-n)}\, e^{-(t-n-n\log(t/n))/2}$.

3 F-tests

In an $F_{(a,b)}$-test we will distinguish if only b depends on n (F-tests with robustness of validity), or both, a and b, depend on n (F-tests with non-robustness of validity).

Proposition 3.1 *In an $F_{(a,b)}$-test in which the first degree of freedom, a, does not depend on n and the second one, b, is a linear function of n and large enough, the VOM+SAD approximations of the p-value p_n^F and the critical value k_n^F under a model F (for which the integrals exist), are, for the p-value*

$$p_n^F \simeq P\{F_{(a,b)} > t\} + A_1 \int_{-\infty}^{\infty} \left\{ \left(A_2\, a - \frac{3\,t+1}{4\,(t-1)} \right) t^{-1/2}\, e^{\frac{t-1}{2t}\frac{(x-\mu)^2}{\sigma^2}} + \frac{3t-1}{2(t-1)}\, t^{-3/2}\, \frac{(x-\mu)^2}{\sigma^2}\, e^{\frac{t-1}{2t}\frac{(x-\mu)^2}{\sigma^2}} \right.$$
$$\left. -\frac{1}{4}\, t^{-5/2}\, \frac{(x-\mu)^4}{\sigma^4}\, e^{\frac{t-1}{2t}\frac{(x-\mu)^2}{\sigma^2}} - \frac{t}{t-1}\, \frac{(x-\mu)^2}{\sigma^2} + \frac{t}{t-1} - A_2\, a \right\} dF(x)$$

where $A_1 = e^{-a(t-1)/2}\, \frac{t^{a/2}}{\sqrt{\pi}\,\sqrt{a}\,(t-1)}$, $A_2 = 1 - \frac{t-1}{a\,\sqrt{2}\,(t-1-\log t)^{3/2}}$. *And for the critical value,*

$$k_n^F \simeq t + \frac{A_1}{f_{a,b}(t)} \int_{-\infty}^{\infty} \left\{ \left(A_2\, a - \frac{3\,t+1}{4\,(t-1)} \right) t^{-1/2}\, e^{\frac{t-1}{2t}\frac{(x-\mu)^2}{\sigma^2}} + \frac{3t-1}{2(t-1)}\, t^{-3/2}\, \frac{(x-\mu)^2}{\sigma^2}\, e^{\frac{t-1}{2t}\frac{(x-\mu)^2}{\sigma^2}} \right.$$
$$\left. -\frac{1}{4}\, t^{-5/2}\, \frac{(x-\mu)^4}{\sigma^4}\, e^{\frac{t-1}{2t}\frac{(x-\mu)^2}{\sigma^2}} - \frac{t}{t-1}\, \frac{(x-\mu)^2}{\sigma^2} + \frac{t}{t-1} - A_2\, a \right\} dF(x)$$

where $t = F_{(a,b);\alpha}$ *is the* $(1-\alpha)$*-quantile of an* $F_{(a,b)}$*-distribution and* $f_{a,b}$ *its density function.*

Proposition 3.2 *In an $F_{(n,n)}$-test the VOM+SAD approximations of the p-value p_n^F and the critical value k_n^F under a model F (for which the integrals exist), are, for the p-value*

$$p_n^F \simeq P\{F_{(n,n)} > t\} + B_1 \left(\sqrt{\frac{t+1}{2}} \left\{ t^{-1/2} \int_{-\infty}^{\infty} e^{\frac{t-1}{4t}\frac{(x-\mu)^2}{\sigma^2}}\, dF(x) + \int_{-\infty}^{\infty} e^{-\frac{t-1}{4}\frac{(x-\mu)^2}{\sigma^2}}\, dF(x) \right\} - 2 \right) + O(n^{-1/2})$$

where $B_1 = \frac{t+1}{(t-1)\sqrt{2\pi}} \left(\frac{2\sqrt{t}}{t+1} \right)^n \sqrt{n}$. *And for the critical value, where* $t = F_{(n,n);\alpha}$

$$k_n^F \simeq t + \frac{B_1}{f_{n,n}(t)} \left(\sqrt{\frac{t+1}{2}} \left\{ t^{-1/2} \int_{-\infty}^{\infty} e^{\frac{t-1}{4t}\frac{(x-\mu)^2}{\sigma^2}}\, dF(x) + \int_{-\infty}^{\infty} e^{-\frac{t-1}{4}\frac{(x-\mu)^2}{\sigma^2}}\, dF(x) \right\} - 2 \right) + O(n^{-1/2})$$

In García-Pérez (2004b), VOM approximations for t-tests are also given.

References

C.A. Field and E. Ronchetti (1985). A tail area influence function and its application to testing. *Communications in Statistics*, 4, 19–41.

A. García-Pérez (2003). von Mises approximation of the critical value of a test. *Test*, 12, 385–411.

A. García-Pérez (2004a). Chi-square tests under models close to the normal distribution. Submitted.

A. García-Pérez (2004b). t-tests with models close to the normal distribution. *Proceedings of the International Conference on Distribution Theory, Order Statistics and Inference in Honor of Barry C. Arnold*, Universidad de Santander.