Tests for normality in classes of skew-t distributions

Published version:

DOI:10.1016/j.spl.2009.09.005

Tests for normality in classes of skew-t alternatives

Cinzia Carota

Dipartimento di Statistica e Matematica Applicata 'Diego De Castro', Università di Torino, Italy.

Abstract

We construct tests for normality in the Azzalini and Capitanio skew-t and linear skew-t classes of distributions. We also provide an explanation for the presence of the inflection point at zero in the skew-normal log-likelihood when it is obtained from a skew-t log-likelihood with degrees of freedom tending to infinity.

Key words: divergence-based Bayesian tests; score tests; skew-normal alternatives; skew-normal log-likelihood; skew-t alternatives; tests for normality.

1. Introduction

In the vicinity of the normal model we consider a new parameterization of the Azzalini and Capitanio (2003) skew-t distribution and obtain a Bayesian test for normality based on the Jeffreys divergence between prior and posterior distributions of the skewness and kurtosis parameters (Proposition 1). This test depends on the marginal score function, whose two elements prove to be linear functions of the sample skewness and sample kurtosis, respectively. We exploit this key finding in various ways. First of all, it is used to explain the different inferential behavior of the skew-normal log-likelihood in comparison with the skew-t log-likelihood (Proposition 2). Then, it helps determine an upper bound for the Bayesian test which is independent of the prior distribution assigned to the skewness and kurtosis parameters (Proposition 3). Finally, it allows us to derive a 'marginal' score test for normality (Proposition 4). Both the upper bound for the Bayesian test and this score test are closely related to certain traditional composite test statistics. These are therefore retrieved and fully justified as tests for normality within the class of Azzalini and Capitanio skew-t alternatives. We also observe some connections between our results and the results in Jones and Faddy (2003) concerning the underlying logic of the new parameterization under which they are obtained and the way the respective score tests for normality depend on the data. We conclude by showing that all of the above results remain true if the class of alternatives is taken to be the class of linear skew-t distributions

(Proposition 5).

The organization of the paper is as follows. The rest of this section demonstrates how a divergence-based Bayesian method (Carota, Parmigiani and Polson 1996, Carota 2005) can be applied to test for normality in the restricted class of skew-normal alternatives (Azzalini, 1985) and its connection to a previous result is noted. The main results (Propositions 1-5) are presented in section 2 and, finally, section 3 contains some concluding remarks.

Let $\mathbf{y}$=$(y_1, .., y_n)$ be a random sample from the skew-normal density function,

$$f_{SN}(y|\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi(\frac{y-\mu}{\sigma})\Phi(\lambda\frac{y-\mu}{\sigma}),$$

where ϕ and Φ denote the standard normal density function and distribution function, respectively, and $\mu \in (-\infty,\infty)$, $\sigma > 0$, $\lambda \in (-\infty,\infty)$. When $\lambda = 0$, f_{SN} reduces to the normal density; otherwise, the sign of λ gives the sign of skewness. Throughout the paper, the location μ and the scale σ are treated as nuisance parameters and are therefore assigned a default prior, $\pi(\mu,\sigma) \propto 1/\sigma$; conversely, the prior of the parameter of interest, $\pi(\lambda)$, is assumed to be very peaked about $\lambda = 0$ so as to express a high degree of belief in the normal model. By virtue of this assumption, it makes sense to test the null hypothesis of normality, $H_0 : \lambda = 0$ versus $H_1 : \lambda \neq 0$, by using a measure of divergence between $\pi(\lambda)$ and the corresponding posterior, $\pi(\lambda|\mathbf{y})$. The specific measure adopted here is the Jeffreys divergence (Jeffreys, 1948),

$$J = \frac{1}{2}\Big\{ \int_{-\infty}^{\infty} \pi(\lambda|\mathbf{y}) log\Big(\frac{\pi(\lambda|\mathbf{y})}{\pi(\lambda)}\Big) d\lambda + \int_{-\infty}^{\infty} \pi(\lambda) log\Big(\frac{\pi(\lambda)}{\pi(\lambda|\mathbf{y})}\Big) d\lambda \Big\}.$$

Small values of J indicate closeness of the posterior to the prior of the parameter of interest and, given the peakedness of the prior about the null value $\lambda = 0$, they are interpreted as sample evidence in favor of the normal model, which is therefore maintained. (For details see Carota, 2005). In particular, we will consider a meaningful approximation of J obtained as follows. If we denote the likelihood integrated with respect to the nuisance parameters by $\overline{L}(\lambda)$ and set $l(\lambda) = log(\overline{L}(\lambda))$, the Jeffreys divergence can be written as

$$J = \frac{1}{2}\Big\{ \int_{-\infty}^{\infty} l(\lambda)\pi(\lambda|\mathbf{y}) d\lambda - \int_{-\infty}^{\infty} l(\lambda)\pi(\lambda) d\lambda \Big\},$$

and a Taylor expansion of the logarithmic function around the null value $\lambda = 0$ provides

$$J_T = \frac{1}{2}\Big[E(\lambda^3|\mathbf{y}) - E(\lambda^3) \Big] l'''(\lambda)\Big|_{\lambda=0}, \qquad (1)$$

where l''' denotes the third derivative of $l(\lambda)$ evaluated at $\lambda = 0$. Given that $l'(\lambda)\Big|_{\lambda=0} = l''(\lambda)\Big|_{\lambda=0} = 0$ (there is always an inflection point at zero as known since Azzalini 1985), J_T represents the first non-zero term in the Taylor expansion of J.

We recall that $l'''(\lambda)\Big|_{\lambda=0}$ is the locally most powerful invariant test for $H_0 : \lambda = 0$ versus $H_1 : \lambda > 0$ provided by Salvan (1986) (see also Azzalini, 2006). It will be denoted from now on by I_S

$$I_S = l'''(\lambda)\Big|_{\lambda=0} = \frac{2(\frac{4}{\pi} - 1)\sqrt{n}\ \Gamma(\frac{n}{2})}{\sqrt{\pi}\ \Gamma(\frac{n-1}{2})}\ \hat{\gamma}_1,$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^n (y_i - \overline{y})^3}{n(s^2)^{\frac{3}{2}}}, \quad \overline{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad s^2 = \frac{\sum_{i=1}^n (y_i - \overline{y})^2}{n}.$$

2. Tests for normality in classes of skew-t alternatives

Let $\mathbf{y}$=$(y_1, .., y_n)$ be a random sample from the skew-t density function

$$f_{ST}(y|\mu, \sigma, \lambda, \nu) = 2\frac{1}{\sigma}t(z|\nu)T(\lambda\, z\ \sqrt{(\nu+1)/(z^2+\nu)}\Big|\nu+1),$$

where $z = (y - \mu)/\sigma$, t and T denote the univariate standard Student t density function and distribution function with ν and $\nu + 1$ degrees of freedom, respectively, and $\mu \in (-\infty, \infty)$, $\sigma > 0$, $\lambda \in (-\infty, \infty)$ (Azzalini and Capitanio, 2003). As in section 1 the prior of μ and σ is taken to be $\pi(\mu, \sigma) \propto 1/\sigma$, while the skewness λ and kurtosis ν are transformed into a new vector of parameters of interest, $\theta_I = (\lambda_1,\ \eta)'$, where $\lambda_1 = \frac{\lambda}{\nu}$ and $\eta = \frac{1}{\nu}$. This alternative parametrization is useful in constructing tests for normality within this class of skew-t distributions and is adopted only 'near' the normal model, which is the special case of $f_{ST}(y|\mu, \sigma, \lambda_1, \eta)$ occurring when $(\lambda_1,\ \eta) = (0,0)$ with the second zero representing an infinitesimal value of η. Detailed comments on the practical meaning of λ_1 and η and on their 'local' use will be given after Proposition 2.

Assuming that the prior of the parameters of interest, $\pi(\theta_I)$, is peaked about the null value $(0,0)$, denoting the likelihood integrated with respect to μ and σ by $\overline{L}(\theta_I)$, and setting $l(\theta_I) = log(\overline{L}(\theta_I))$, in the Appendix we prove the following result on the divergence-based Bayesian test for normality.

Proposition 1.

The approximation J_T (obtained via Taylor expansion) of the Jeffreys divergence between prior and posterior distributions of θ_I is given by

$$J_T = \frac{1}{2}[E(\theta_I|y) - E(\theta_I)]' \times \frac{\partial}{\partial\theta_I} l(\theta_I)\Big|_{\theta_I=(0,0)}$$

$$\approx \frac{1}{2}\Big\{[E(\lambda_1|y) - E(\lambda_1)] \times \frac{2\sqrt{n}}{\sqrt{\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\ \hat{\gamma}_1 +$$

$$+[E(\eta|y) - E(\eta)] \times \Big[\frac{3(n-1)}{2n} + \frac{(n^2-1)}{4n}\ \hat{\gamma}_2\Big]\Big\}$$

where $\hat{\gamma}_1$ is the sample skewness statistic defined in section 1 and

$$\hat{\gamma}_2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^4}{n(s^2)^2}\ -\ 3.$$

Proposition 1 says that the approximation J_T of J is a linear approximation and that the elements of the marginal score function, $\frac{\partial}{\partial\theta_I} l(\theta_I)$, are both non-zero and given by linear functions of the sample skewness and sample kurtosis, respectively. The interesting structure of J_T leads us to explore three different consequences of this key finding in Propositions 2-4; for details on the use of J_T see Carota (2005).

The next remark provides a possible answer to this problem: 'It would be useful to have some theoretical insight into why the log-likelihood function using the skew-t distribution behaves so differently from the skew-normal model' (Azzalini and Capitanio, 2003, p.384).

<u>Remark</u>.

Given that $f_{SN}(y|\mu, \sigma, \lambda) = lim_{\nu\to\infty} f_{ST}(y|\mu, \sigma, \lambda, \nu)$ and consequently

$$l(\lambda) = lim_{\nu\to\infty} l(\lambda, \nu),$$

where $l(\lambda)$ and $l(\lambda, \nu)$ are the corresponding integrated log-likelihoods with respect to μ and σ, we can explain the fact that $l'(\lambda)\Big|_{\lambda=0} = 0$, due to the inflection point, as follows. Let $\lambda \in U(0)$, where U denotes a neighborhood of 0. Consider the function $\lambda_1 = g(\lambda) = \lambda/\nu$ and write $l(\lambda, \nu) = l(g(\lambda)\,\nu,\, \nu)$, so that l depends on λ only through g. As

$$l'(\lambda)|_{\lambda=0} = \frac{d}{d\lambda}\Big(lim_{\nu\to\infty} l(\lambda, \nu)\Big)|_{\lambda=0} = lim_{\nu\to\infty} \frac{d}{d\lambda} l(\lambda, \nu)|_{\lambda=0} =$$

$$lim_{\nu\to\infty}\frac{d}{d\lambda}l(g(\lambda)\,\nu,\,\nu)|_{\lambda=0} = lim_{\nu\to\infty}\Big(\frac{dl(g(\lambda)\,\nu,\,\nu)}{dg(\lambda)} \times \frac{dg(\lambda)}{d\lambda}\Big)\Big|_{\lambda=0} =$$

$$\Big(\frac{2\sqrt{n}}{\sqrt{\pi}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\;\hat{\gamma}_1 \times lim_{\nu\to\infty}\frac{1}{\nu}\Big) = 0,$$

the statements a) and b) below hold.

Proposition 2.

a) $l'(\lambda)|_{\lambda=0} = 0$ *because of the infinitely large value of* ν*.*

b)

$$lim_{\nu\to\infty}\frac{dl(g(\lambda)\,\nu,\,\nu)}{dg(\lambda)}\Big|_{\lambda=0} \propto I_S.$$

Roughly speaking, the statement b) says that, when the value of ν is sufficiently large, if we regard $l(\lambda,\nu) = l(\lambda_1\,\nu,\,\nu)$ as a function of λ_1 rather than λ, it is as though we are looking at the skew-normal log-likelihood from a sufficiently large distance that the inflection point becomes imperceptible, and, interestingly, the first derivative at 0 becomes proportional to I_S. This 'distancing' effect has been obtained by dividing λ by the degrees of freedom ν, allowing us to determine that $\eta = 1/\nu$ is the right 'distancing' factor to achieve this twofold advantage.

Jones and Faddy (2003) make a similar observation on p.163, where their skewness parameter is judged to be 'not very satisfactorily tied to skewness' and a 'normalization' of it with respect to the degrees of freedom is preannounced to be 'more meaningful'. However, Jones and Faddy adopt their alternative parameterization globally, since they transform the whole original parameter space into a new parameter space. Instead, we use the new vector θ_I only locally, close to the normal model. In particular, θ_I plays a key role in the proof that the marginal score function is given by the two linear functions of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ provided in Proposition 1 (see the Appendix), i.e. when λ_1 and η are used in a very special way: one at a time keeping the other fixed at zero. Thus, in practice, our alternative parameterization is applied only to the axes of the original parameter space which represent the sub-classes of skew-normal densities ($\nu \to \infty$) and Student t densities ($\lambda = 0$) of the $f_{ST}(\cdot|\mu, \sigma, \lambda, \nu)$ family. Nonetheless, there is also an underlying theoretical necessity for this 'partial' use of $\theta_I = (\lambda_1, \eta)'$: a global use would produce a third class of skew-t distributions where the skewing method is similar to the one in Azzalini and Capitanio (2003), but skewness and kurtosis change simultaneously as in the skew-t densities by Jones and Faddy (2003). The introduction of this third family goes beyond the scope of this paper which seeks instead to emphasize commonalities with the previous two families. All the results in the rest of the paper are based on such partial and local use of λ_1 and η.

Proposition 3.

With a further transformation of the vector of parameters of interest, $\theta_I^* = (\delta_1, \zeta)'$, where

$$\delta_1 = \frac{\lambda_1}{\sqrt{1+\lambda_1^2}} \in (-1,1), \qquad \textit{and} \qquad \zeta = \frac{\eta}{\eta+1} \in (0,1),$$

irrespectively of the prior assigned to (δ_1, ζ), we have

$$J_T \leq c_0 + c_1\hat{\gamma}_1 + c_2\hat{\gamma}_2,$$

where $c_0 = 3(n-1)/4n$, $c_1 = 2\sqrt{n}\Gamma(\frac{n}{2})/(\sqrt{\pi}\Gamma(\frac{n-1}{2}))$, $c_2 = (n^2-1)/8n$.

This result follows from $E(\delta_1|y) - E(\delta_1|y) < 2$, $E(\zeta|y) - E(\zeta|y) < 1$ and

$$\frac{\partial l(\theta_I^*)}{\partial \theta_I^*}\Big|_{\theta_I^*=(0,0)} = \frac{\partial l(\theta_I)}{\partial \theta_I}\Big|_{\theta_I=(0,0)}.$$

Proposition 3 provides an upper bound for J_T, independent of the prior, which has the structure of a composite test statistic for normality (see e.g. Cox and Hinkley, 1974, p.71).

Notice also that, if $(y_1, .., y_n)$ is assumed to be a random sample from $f_{SN}(y|\mu, \sigma, \lambda)$, with the usual transformation $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ from (1) we obtain $J_T < I_S$, no matter what prior is assigned to δ.

Moreover, for a large value of n, we can derive the Fisher information matrix at $\theta_I = (0,0)$ corresponding to the marginal score function $\frac{\partial}{\partial\theta_I} l(\theta_I)$,

$$I(\theta_I)\Big|_{\theta_I=(0,0)} = I(0,0) = E\Big(\frac{\partial}{\partial\theta_I} l(\theta_I)\Big(\frac{\partial}{\partial\theta_I} l(\theta_I)\Big)'\Big)\Big|_{\theta_I=(0,0)} =$$

$$\begin{pmatrix} \frac{6}{n}\Big(\frac{2\sqrt{n}}{\sqrt{\pi}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\Big)^2 & 0 \\ 0 & \Big(\frac{3(n-1)}{2n}\Big)^2 + \frac{24}{n}\Big(\frac{(n^2-1)}{4n}\Big)^2 \end{pmatrix},$$

and the expression of the corresponding score test for normality.

<u>Proposition 4.</u>

$$S = \frac{\partial}{\partial\theta_I} l(\theta_I)' I^{-1}(0,0) \frac{\partial}{\partial\theta_I} l(\theta_I)\Big|_{\theta_I=(0,0)} = \frac{n}{6}\hat{\gamma}_1^2 + \frac{\Big(6+(n+1)\hat{\gamma}_2\Big)^2}{36+\frac{24}{n}(n+1)^2}$$

$$\approx \frac{n}{6}\hat{\gamma}_1^2 + \frac{n}{24}\hat{\gamma}_2^2.$$

The last term is the Fisher-Pearson composite test statistic for normality. It is worth noting that S and the score test derived by Jones and Faddy within their class of skew-t densities, originating from completely different variate transformations, are both ‘based on the usual sample skewness and kurtosis’ (Jones and Faddy, 2003, p. 164, (8)).

Our last point concerns the case of linear skew-t alternatives.

<u>Proposition 5.</u>

If $\mathbf{y}=(y_1,..,y_n)$ *is assumed to be a random sample from the linear skew-t density function,*

$$f(y|\mu,\sigma,\lambda,\nu) = 2\frac{1}{\sigma}t(z|\nu)T(\lambda\, z\,|\nu),$$

the meaning of the symbols being equal, Propositions 1-4 still hold.

3. Final comments

Starting from a Bayesian test for normality, two traditional composite test statistics have been retrieved and fully justified as tests for normality within the class of skew-t alternatives by Azzalini and Capitanio (2003) or linear skew-t. The first is an upper bound, independent of the prior, for the approximate Jeffreys divergence between prior and posterior distributions of the skewness and kurtosis parameters. The second is a score test based on the marginal score function. We have also found that, close to the normal model, the skew-t by Azzalini and Capitanio (2003), the linear skew-t and the skew-t by Jones and Faddy (2003) have two aspects in common: 1) the appropriateness of a 'normalization' of the skewness parameter with respect to the degrees of freedom and 2) the fact that the score test for normality depends on the data only via sample skewness and kurtosis.

In addition, we have provided an explanation of the inferential behaviour of the skew-normal log-likelihood given from within the skew-t log-likelihood.

Acknowledgements

I sincerely thank Adelchi Azzalini for his constructive and valuable comments. I also thank the Associate Editor and the Referee for many suggestions that greatly improved the presentation of this paper. This work was partially supported by PRIN 2006, grant No.2006132978, from MIUR, Italy.

References

Azzalini, A. , 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12, pp. 171-178.

Azzalini, A. , 2006. Normality, Salvan test for, in Encyclopedia of Statistical Sciences, John Wiley and Sons,Inc.

Azzalini, A. and Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distributions. Journal of the Royal Statistical Society B 65, pp. 367-389.

Box, G. E. P. and Tiao, G. C., 1973. Bayesian Inference in Statistical Analysis, New York: Wiley.

Carota, C., 2005. Symmetric diagnostics for the analysis of residuals in regression models. Biometrika 92, pp. 787-799.

Carota, C., Parmigiani, G. and Polson, N. G., 1996. Diagnostic measures for model criticism. Journal of the American Statistical Association 91, pp. 753-762.

Cox, D. R. and Hinkley, D. V., 1974. Theoretical Statistics, London: Chapman and Hall.

Jeffreys, H., 1948. Theory of Probability, Oxford: Clarendon Press.

Fisher, R. A. , 1925. Expansion of "Student's" integral in powers of n^{-1}. Metron 5, pp.109-20.

Jones, M. C., Faddy, M. J., 2003. A skew extension of the t-distribution, with applications. Journal of the Royal Statistical Society B 65, pp. 159-174.

Johnson, N.L., Kotz, S. and Balakrishnan, N., 1995. Continuous Univariate Distributions, vol.2 2nd edn. New York: Wiley.

Salvan, A., 1986. Test localmente piu' potenti tra gli invarianti per la verifica dell'ipotesi di normalita', in Societa Italiana di Statistica ed., Atti della XXXIII Riunione della Societa' Italiana di Statistica, 2, pp.173-9.

APPENDIX.

Proof of Proposition 1.

Part I. We show that

$$lim_{\epsilon\to 0}\frac{\partial}{\partial\lambda_1}log(\bar{L}(\lambda_1,\eta))\Big|_{(0,\epsilon)} = \frac{lim_{\epsilon\to 0}\frac{\partial}{\partial\lambda_1}\bar{L}(\lambda_1,\eta)\Big|_{(0,\epsilon)}}{lim_{\epsilon\to 0}\bar{L}(\lambda_1,\eta)\Big|_{(0,\epsilon)}} = \frac{2\sqrt{n}}{\sqrt{\pi}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\ \hat{\gamma}_1.$$

Consider the likelihood function,

$$L(\mu,\sigma,\lambda_1,\eta)=2^n\Big(\frac{1}{\sigma}\Big)^n\Big(\frac{\eta}{\pi}\Big)^{\frac{n}{2}}\Big[\frac{\Gamma(\frac{\eta^{-1}+1}{2})}{\Gamma(\frac{\eta^{-1}}{2})}\Big]^n\times\prod_{i=1}^n\Big[1+z_i^2\eta\Big]^{-\frac{\eta^{-1}+1}{2}}\times\prod_{i=1}^n T\Big(\lambda_1\eta^{-1}z_i\sqrt{\frac{\eta^{-1}+1}{z_i^2+\eta^{-1}}}\Big|\eta^{-1}+1\Big),$$

and observe that

$$\begin{aligned}\frac{\partial}{\partial\lambda_1}\bar{L}(\lambda_1,\eta)&=\frac{\partial}{\partial\lambda_1}\Big\{\int_0^\infty\int_{-\infty}^{+\infty}L(\mu,\sigma,\lambda_1,\eta)\frac{1}{\sigma}d\mu d\sigma\Big\}\Big|_{(\lambda_1,\eta)=(0,\epsilon)}=\\&=\int_0^\infty\int_{-\infty}^{+\infty}\Big\{\frac{\partial}{\partial\lambda_1}L(\mu,\sigma,\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)}\Big\}\frac{1}{\sigma}d\mu d\sigma.\qquad(2)\end{aligned}$$

From

$$\frac{\partial}{\partial\lambda_1}L(\mu,\sigma,\lambda_1,\eta)=2^n\Big(\frac{1}{\sigma}\Big)^n\Big(\frac{\eta}{\pi}\Big)^{\frac{n}{2}}\Big[\frac{\Gamma(\frac{\eta^{-1}+1}{2})}{\Gamma(\frac{\eta^{-1}}{2})}\Big]^n\times\prod_{i=1}^n\Big[1+z_i^2\eta\Big]^{-\frac{\eta^{-1}+1}{2}}$$

$$\times\sum_{i=1}^n t\Big(\lambda_1\eta^{-1}z_i\sqrt{\frac{\eta^{-1}+1}{z_i^2+\eta^{-1}}}\Big|\eta^{-1}+1\Big)\eta^{-1}z_i\sqrt{\frac{\eta^{-1}+1}{z_i^2+\eta^{-1}}}\times\prod_{j\neq i}^n T\Big(\lambda_1\eta^{-1}z_j\sqrt{\frac{\eta^{-1}+1}{z_j^2+\eta^{-1}}}\Big|\eta^{-1}+1\Big),$$

we obtain

$$\frac{\partial}{\partial\lambda_1}L(\mu,\sigma,\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)}=2\Big(\frac{1}{\sigma}\Big)^n\Big(\frac{\epsilon}{\pi}\Big)^{\frac{n}{2}}\Big[\frac{\Gamma(\frac{\epsilon^{-1}+1}{2})}{\Gamma(\frac{\epsilon^{-1}}{2})}\Big]^n\times\prod_{i=1}^n\Big[1+z_i^2\epsilon\Big]^{-\frac{\epsilon^{-1}+1}{2}}$$

$$\times\sum_{i=1}^n\frac{1}{\sqrt{(\epsilon^{-1}+1)\pi}}\frac{\Gamma(\frac{\epsilon^{-1}+2}{2})}{\Gamma(\frac{\epsilon^{-1}+1}{2})}\epsilon^{-1}z_i\sqrt{\frac{\epsilon^{-1}+1}{z_i^2+\epsilon^{-1}}},$$

which, considering that $\sqrt{(\epsilon^{-1}+1)/(z_i^2+\epsilon^{-1})}=\sqrt{1+\epsilon}\ \ [1+z_i^2\epsilon]^{-1/2}$

and that

$$[1+z_i^2\epsilon]^{-1/2}\times\Big[1+z_i^2\epsilon\Big]^{-\frac{\epsilon^{-1}+1}{2}}=\Big[1+z_i^2\frac{1+\epsilon}{\epsilon^{-1}+1}\Big]^{-\frac{(\epsilon^{-1}+1)+1}{2}},$$

can be rewritten as

$$2\Big(\frac{1}{\sigma}\Big)^n\Big(\frac{\epsilon}{\pi}\Big)^{\frac{1}{2}}\frac{\Gamma(\frac{\epsilon^{-1}+1}{2})}{\Gamma(\frac{\epsilon^{-1}}{2})}\sqrt{1+\epsilon}\times\sum_{i=1}^n\epsilon^{-1}z_i\times t\Big(z_i\sqrt{1+\epsilon}\Big|\epsilon^{-1}+1\Big)\times\prod_{j\neq i}^n\times t(z_j|\epsilon^{-1}).$$

We set $w_i = z_i\sqrt{1+\epsilon}$ and apply the Fisher expansion (Fisher 1925, Johnson et al. 1995 p. 375) to the t densities,

$$\frac{\partial}{\partial\lambda_1}L(\mu,\sigma,\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)} = 2\Big(\frac{1}{\sigma}\Big)^n\Big(\frac{\epsilon}{\pi}\Big)^{\frac{1}{2}}\frac{\Gamma(\frac{\epsilon^{-1}+1}{2})}{\Gamma(\frac{\epsilon^{-1}}{2})}\sqrt{1+\epsilon}\ \sum_{i=1}^{n}\epsilon^{-1}z_i\times\phi(w_i)\times$$

$$\Big\{1+\frac{1}{4(\epsilon^{-1}+1)}\Big(w_i^4-2w_i^2-1\Big)+\frac{1}{96(\epsilon^{-1}+1)^2}\Big(3w_i^8-28w_i^6+30w_i^4+12w_i^2+3\Big)+\cdots\Big\}$$

$$\times\prod_{j\neq i}^{n}\phi(z_j)\Big\{1+\frac{\epsilon}{4}\Big(z_j^4-2z_j^2-1\Big)+\frac{\epsilon^2}{96}\Big(3z_j^8-28z_j^6+30z_j^4+12z_j^2+3\Big)+\cdots\Big\}.$$

Now we consider

$$\phi(w_i)\times\prod_{j\neq i}^{n}\phi(z_j) = \Big(\frac{1}{2\pi}\Big)^{\frac{n}{2}}exp\{-\frac{1}{2}w_i^2-\frac{1}{2}\sum_{j\neq i}^{n}z_j^2\}$$

and standard algebra allows us to write

$$exp\{-\frac{1}{2}w_i^2-\frac{1}{2}\sum_{j\neq i}^{n}z_j^2\} = exp\{-\frac{1}{2\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]-\frac{n+\epsilon}{2\sigma^2}\big(\mu-\overline{m}_i\big)^2\},$$

where $\overline{m}_i = (n\overline{y}+\epsilon y_i)/(n+\epsilon)$.

Then, we consider the expression

$$\Big\{1+\frac{1}{4(\epsilon^{-1}+1)}\Big(w_i^4-2w_i^2-1\Big)+\cdots\Big\}\times\prod_{j\neq i}^{n}\Big\{1+\frac{\epsilon}{4}\Big(z_j^4-2z_j^2-1\Big)+\cdots\Big\} =$$

$$\Big\{1+\frac{\epsilon}{4(1+\epsilon)}\Big(z_i^4(1+\epsilon)^2-2z_i^2(1+\epsilon)-1-\epsilon+\epsilon\Big)+\cdots\Big\}\times\prod_{j\neq i}^{n}\Big\{1+\frac{\epsilon}{4}\Big(z_j^4-2z_j^2-1\Big)+\cdots\Big\}$$

and, by ignoring all terms depending on $\epsilon^k,\ k\geq 2$, we approximate it as follows

$$\Big\{1+\sum_{j=1}^{n}\frac{\epsilon}{4}\Big(z_j^4-2z_j^2-1\Big)\Big\}.$$

Consequently, we have

$$\frac{\partial}{\partial\lambda_1}L(\mu,\sigma,\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)} \approx 2\Big(\frac{1}{\sigma}\Big)^n\Big(\frac{\epsilon}{\pi}\Big)^{\frac{1}{2}}\frac{\Gamma(\frac{\epsilon^{-1}+1}{2})}{\Gamma(\frac{\epsilon^{-1}}{2})}\sqrt{1+\epsilon}\ \sum_{i=1}^{n}\epsilon^{-1}z_i$$
$$\times\Big(\frac{1}{2\pi}\Big)^{\frac{n}{2}}exp\{-\frac{1}{2\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]-\frac{n+\epsilon}{2\sigma^2}(\mu-\overline{m}_i)]^2\}$$
$$\times\Big\{1+\sum_{j=1}^{n}\frac{\epsilon}{4}\Big(z_j^4-2z_j^2-1\Big)\Big\}$$

and, setting

$$C(\epsilon)=2\Big(\frac{\epsilon}{\pi}\Big)^{\frac{1}{2}}\frac{\Gamma(\frac{\epsilon^{-1}+1}{2})}{\Gamma(\frac{\epsilon^{-1}}{2})}\sqrt{1+\epsilon}\ \times\Big(\frac{1}{2\pi}\Big)^{\frac{n}{2}},$$

such an approximation is substituted for $\frac{\partial}{\partial\lambda_1}L(\mu,\sigma,\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)}$ in (2).

Now we are going to obtain $lim_{\epsilon\to 0}(2)$ as sum of the terms provided in the next two steps i) and ii).

Considering that

$$z_i=\frac{n}{\sigma}\Big(\frac{y_i-\overline{y}}{n+\epsilon}\Big)-\frac{1}{\sigma}(\mu-\overline{m}_i)$$

and, successively, that

$$\int_0^{\infty}\Big(\frac{1}{\sigma}\Big)^h exp\Big\{-\frac{1}{2\sigma^2}\,D\Big\}d\sigma=2^{\frac{h-3}{2}}\Gamma(\frac{h-1}{2})\,D^{-\frac{h-1}{2}} \qquad (3)$$

(Box and Tiao 1973, p. 145, A 2.1.4), where $D=\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]$, we have:

i)

$$\int_0^{\infty}C(\epsilon)\Big(\frac{1}{\sigma}\Big)^{n+1}\sum_{i=1}^{n}exp\Big\{-\frac{1}{2\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]\Big\}\times$$

$$\int_{-\infty}^{+\infty} \epsilon^{-1}\ z_i\ exp\Big\{-\frac{n+\epsilon}{2\sigma^2}(\mu-\overline{m}_i)^2\Big\}\ d\mu\ d\sigma =$$

$$\int_0^{\infty} C(\epsilon)\Big(\frac{1}{\sigma}\Big)^{n+1}\sum_{i=1}^{n} exp\Big\{-\frac{1}{2\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]\Big\}\times\frac{n}{\epsilon}\ \Big(\frac{y_i-\overline{y}}{n+\epsilon}\Big)\ \Big(\frac{2\pi}{n+\epsilon}\Big)^{\frac{1}{2}}\ d\sigma =$$

$$C(\epsilon)\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{3}{2}}}\ \frac{n}{\epsilon}\ 2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})\sum_{i=1}^{n}\ (y_i-\overline{y})\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]^{-\frac{n}{2}},$$

whose limit, as $\epsilon \to 0$, is zero.

ii)

$$\int_0^{\infty} C(\epsilon)\Big(\frac{1}{\sigma}\Big)^{n+1}\ \sum_{i=1}^{n} exp\Big\{-\frac{1}{\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]\Big\}\times$$

$$\int_{-\infty}^{+\infty} \epsilon^{-1} z_i \times \sum_{j=1}^{n}\frac{\epsilon}{4}\Big(z_j^4-2z_j^2-1\Big)\ exp\Big\{-\frac{n+\epsilon}{2\sigma^2}(\mu-\overline{m}_i)^2\Big\}d\mu d\sigma =$$

$$= \int_0^{\infty} C(\epsilon)\times\Big(\frac{1}{\sigma}\Big)^{n+1}\ \frac{1}{4}\ \sum_{i=1}^{n} exp\Big\{-\frac{1}{2\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]\Big\}\times$$

$$\int_{-\infty}^{+\infty} z_i\Big(\sum_{j=1}^{n} z_j^4 - 2\sum_{j=1}^{n} z_j^2 - n\Big) exp\Big\{-\frac{n+\epsilon}{2\sigma^2}(\mu-\overline{m}_i)^2\Big\}d\mu d\sigma$$

where, applying $(y_j-\mu)^r = [(y_j-\overline{m}_i)-(\mu-\overline{m}_i)]^r = \sum_{k=0}^{r} r!/\{k!(r-k)!\}(y_j-\overline{m}_i)^k(-1)^{r-k}(\mu-\overline{m}_i)^{r-k}$, the value of the integral with respect to μ is obtained from the central moments of the normal distribution $N(\overline{m}_i, \frac{\sigma}{\sqrt{n+\epsilon}})$:

$$\frac{1}{\sigma^4}(y_i-\overline{m}_i)\sum_{j=1}^{n}(y_j-\overline{m}_i)^4\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{1}{2}}}\ +\ \frac{6}{\sigma^2}(y_i-\overline{m}_i)\sum_{j=1}^{n}(y_j-\overline{m}_i)^2\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{3}{2}}}$$

$$+3n(y_i-\overline{m}_i)\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{5}{2}}}\ -\ \frac{2}{\sigma^2}(y_i-\overline{m}_i)\sum_{j=1}^{n}(y_j-\overline{m}_i)^2\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{1}{2}}}$$

$$-2n(y_i-\overline{m}_i)\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{3}{2}}}\ -\ n(y_i-\overline{m}_i)\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{1}{2}}}\ +\ \frac{4}{\sigma^2}\sum_{j=1}^{n}(y_j-\overline{m}_i)^3\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{3}{2}}}$$

$$+12\sum_{j=1}^{n}(y_j-\overline{m}_i)\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{5}{2}}}\ -\ 4\sum_{j=1}^{n}(y_j-\overline{m}_i)\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{3}{2}}},$$

and, for each of these terms, $\tau(\sigma)$ say, the value of the integral with respect to σ,

$$\int_0^\infty C(\epsilon)\Big(\frac{1}{\sigma}\Big)^{n+1}\ \frac{1}{4}\ \sum_{i=1}^{n}exp\Big\{-\frac{1}{2\sigma^2}\Big[ns^2+\frac{n}{1+n\epsilon^{-1}}(y_i-\overline{y})^2\Big]\Big\}\tau(\sigma)d\sigma, \qquad (4)$$

is derived from (3). Since, as $\epsilon\to 0$, (4) takes a non-zero value only when

$$\tau(\sigma)=\frac{4}{\sigma^2}\sum_{j=1}^{n}(y_j-\overline{m}_i)^3\frac{(2\pi)^{\frac{1}{2}}}{(n+\epsilon)^{\frac{3}{2}}}$$

and this value is given by

$$2\pi^{-\frac{n}{2}}\Gamma(\frac{n}{2}+1)\frac{1}{\sqrt{n}}\frac{\sum_{j=1}^{n}(y_j-\overline{y})^3}{(ns^2)^{\frac{n+2}{2}}}, \qquad (5)$$

from i) and ii) we conclude that

$$lim_{\epsilon\to 0}(2)=(5).$$

Similarly, applying the Fisher expansion to $L(\mu,\sigma,\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)}$, we easily obtain

$$lim_{\epsilon\to 0}\overline{L}(\lambda_1,\eta)\Big|_{(\lambda_1,\eta)=(0,\epsilon)}\simeq\Big(\frac{1}{2\pi}\Big)^{(n-1)/2}\Big(\frac{1}{n}\Big)^{\frac{1}{2}}2^{\frac{n-3}{2}}\Gamma\Big(\frac{n-1}{2}\Big)(ns^2)^{-\frac{n-1}{2}}$$

and this completes the proof of Part I.

Part II. With similar techniques it can be shown that

$$lim_{\epsilon\to 0}\frac{\partial}{\partial\eta}log(\bar{L}(\lambda_1,\eta))\Big|_{(0,\epsilon)}=\frac{lim_{\epsilon\to 0}\frac{\partial}{\partial\eta}\bar{L}(\lambda_1,\eta)\Big|_{(0,\epsilon)}}{lim_{\epsilon\to 0}\bar{L}(\lambda_1,\eta)\Big|_{(0,\epsilon)}}=\frac{3(n-1)}{2n}+\frac{(n^2-1)}{4n}\ \hat{\gamma}_2.$$