TESTING GAUSSIANITY WITH THE CHARACTERISTIC FUNCTION

M. Jay Arnold, D. Robert Iskander and Abdelhak M. Zoubir

Signal Processing Research Centre
Queensland University of Technology
Brisbane, Queensland 4000, Australia
email: m.arnold@qut.edu.au

ABSTRACT

We wish to formulate a test for the hypothesis $X_i \sim N(\mu, \sigma^2)$ for $i = 0, 1, ..., N-1$ against unspecified alternatives. We assume independence of the components of $\mathbf{X} = [X_0, X_1, ..., X_{N-1}]$. This is a problem of universal importance as the assumption of Gaussianity is prevalent and fundamental to many statistical theories and engineering applications. Many such tests exist, the most well-known being the χ^2 goodness-of-fit test with its variants and the Kolmogorov-Smirnov one-sample cumulative probability function test. More powerful modern tests for the hypothesis of Gaussianity include the D'Agostino K^2 and Shapiro-Wilk W tests. Recently, tests for Gaussianity have been proposed which use the characteristic function. It is the purpose of this paper to highlight and resolve problems with these tests and to improve performance so that the test is competitive with, and in some cases better than, the most powerful known tests for Gaussianity.

1. CHARACTERISTIC FUNCTION BASED TESTS OF DISTRIBUTION

We assume X_i, for all i, has the probability density function (pdf) $f_X(x)$ and characteristic function (cf) $\phi_X(t) = E[e^{jXt}]$. If $\mathbf{X}$ is Gaussian then it is uniquely characterised by the cf $\phi_X(t) = \exp(j\mu t - \sigma^2 t^2/2)$. Koutrouvelis [3] introduced a test of distribution based on the *empirical* cf (ECF),

$$\hat{\phi}^e_X(t) = \frac{1}{N}\sum_{i=0}^{N-1} e^{jtX_i} \qquad (1)$$

We produce a realisation graph (Figure 1) consisting of overlayed realisations of the (absolute value) ECF with $X_i \sim N(0,1)$ and $N = 64$. The fact that the variance of the ECF approaches a constant $(1/2N)$ with t is pointing to a systematic non-optimality in its use. It turns out that the estimator, while being unbiased, has unacceptable variance over all values of t and must be replaced. Epps [2] formulated a general estimator based on Koutrouvelis' proposal, using a base function $\hat{\phi}^E_X(X_i; t_k)$, $k = 0, 1, ..., T-1$, and a sum-of-squares statistic $Q^E_X(\mathbf{t}) \sim \chi^2_{T-2}$ where $\mathbf{t} = [t_0, t_1, ..., t_{T-1}]$. The performance of this test, which we will call the Koutrouvelis-Epps, or K-E, test, is controlled by the choice of $\mathbf{t}$ and $Q^E_X(X_i; t_k)$. We highlight the following problems encountered with using the test in practice as implemented with choice of $\hat{\phi}^E_X(X_i; t_k)$ corresponding to the ECF.

1. For large values of T numerical and computational problems exist in the computation of $Q^E_X(\mathbf{t})$. Also the distributional assumptions of the test statistic fail - indicating that the more information we try to use, the less valid our test becomes.
2. Given that T must be small the performance of the test is critically dependent upon the choice of $\mathbf{t}$ [2, 5, 4], which cannot be optimised in any even reasonably general way.
3. Estimates for μ and σ are allowed to be unbounded and are chosen so as to minimise $Q^E_X(\mathbf{t})$. When T is small this decreases the ability of the method to resolve many non-Gaussian processes as their ECF's merely have to correspond to that of an *arbitrary* Gaussian cf at a small number of points.
4. The ECF is an extremely poor characteristic function estimator.
5. The inference that the test does not assume independence of the data is falsified by simple experimentation with synethetic data.

In this contribution we resolve problems (1)-(4) by using an alternative cf estimator and defining a new test statistic.

2. THE KERNEL CHARACTERISTIC FUNCTION ESTIMATOR

We formulate the *kernel* cf estimator (KCFE) in analogy with the theory of *kernel density estimation* [8]. The most general formulation for the KCFE $\hat{\phi}_X(t)$ of the cf of an independent, indentically distributed random sample $\mathbf{X} = [X_0, X_1, ..., X_{N-1}]$ is [1],

$$\hat{\phi}_X(t; \varphi_{\mathbf{X},i}) = \frac{1}{N}\sum_{i=0}^{N-1} \varphi_{\mathbf{X},i}(t) e^{jX_i t} \qquad (2)$$

where, $\varphi_{\mathbf{X},i}$ is the *cf domain kernel function* (ckf) which in general may depend on the data and the sample number. We can note that the convolutive smoothing operation of kernel density estimation has been replaced by a multiplicative smoothing operation in the cf domain. While the result

0-7803-2431-5/95 $4.00 © 1995 IEEE

of the operation in both domains is a bias-variance tradeoff the optimal form of smoothing in each domain will necessarily be different. The *data dependent KCFE* (DKCFE) can be simplified further,

$$\hat{\phi}_X(t;\varphi_X) = \varphi_X(t)\hat{\phi}_X(t;1) \quad (3)$$

where, $\hat{\phi}_X(t;1) = \hat{\phi}_X^e(t)$ is just the ECF. We can note that the ECF corresponds to a "zero-width histogram" in the density domain, which is not generally highly regarded as a density estimator. Likewise the performance of the ECF as an estimator of the cf is poor. Figure 2 shows multiple realisations of an adaptive KCFE (magnitude). Comparison with Figure 1 clearly displays the improved variance performance. More elaborate studies are given in [1].

3. THE FIXED KERNEL CHARACTERISTIC FUNCTION ESTIMATOR

The most general form of KCFE which is amenable to statistical analysis is the *fixed KCFE* (FKCFE) whose ckf depends only upon t. Often an optimal FKCFE can then be approximated by a DKCFE. To examine performance in more detail we first seperate real and imaginary components of the FKCFE,

$$\Re\hat{\phi}_X(t;\varphi) = \Re\varphi(t)\Re\hat{\phi}_X(t;1) - \Im\varphi(t)\Im\hat{\phi}_X(t;1) \quad (4)$$

$$\Im\hat{\phi}_X(t;\varphi) = \Re\varphi(t)\Im\hat{\phi}_X(t;1) + \Im\varphi(t)\Re\hat{\phi}_X(t;1) \quad (5)$$

where $\Re$ and $\Im$ are the real and imaginary component operators respectively. We can now obtain the expectations of these components,

$$E[\Re\hat{\phi}_X(t;\varphi)] = \Re\varphi(t)\Re\phi_X(t) - \Im\varphi(t)\Im\phi_X(t) \quad (6)$$

$$E[\Im\hat{\phi}_X(t;\varphi)] = \Re\varphi(t)\Im\phi_X(t) + \Im\varphi(t)\Re\phi_X(t) \quad (7)$$

We also obtain the estimator variances,

$$V[\Re\hat{\phi}_X(t;\varphi)] = \left[\Re^2\varphi(t)(1+\Re\phi_X(2t)) - 2E^2[\Re\hat{\phi}_X(t;\varphi)] + \Im^2\varphi(t)(1-\Re\phi_X(2t)) - 2\Re\varphi(t)\Im\varphi(t)\Im\phi_X(2t)\right]\frac{1}{2N} \quad (8)$$

$$V[\Im\hat{\phi}_X(t;\varphi)] = \left[\Re^2\varphi(t)(1-\Re\phi_X(2t)) - 2E^2[\Im\hat{\phi}_X(t;\varphi)] + \Im^2\varphi(t)(1+\Re\phi_X(2t)) + 2\Re\varphi(t)\Im\varphi(t)\Im\phi_X(2t)\right]\frac{1}{2N} \quad (9)$$

From these statistics we can formulate a *mean square error* (MSE) criterion for obtaining the best ckf for the estimator of both the real and imaginary components of the cf. If a particular function of the cf was required, for example its magnitude or phase, then alternate statistics could be developed. Note that as N increases the optimum ckf should, for *consistency*, approach that of the ECF,

$$\lim_{N\to\infty} \varphi^{\Re}(t) = \lim_{N\to\infty} \varphi^{\Im}(t) = 1, \quad (10)$$

This does not correspond to the case that we are most interested in, which is that of testing distribution with a small number of samples when the power of the test is most important.

Minimum Integrated Mean Square Error FKCFE. We formulate the FKCFE in the sense of the minimum integrated MSE for the real and imaginary components individually. For estimating $\Re\phi_X(t)$ we obtain the ckf $\varphi^{\Re}(t)$ minimising,

$$IM[\Re\hat{\phi}_X(t;\varphi^{\Re})] = \int_t V[\Re\hat{\phi}_X(t;\varphi^{\Re})] + (E[\Re\hat{\phi}_X(t;^{\Re})] - \Re\phi_X(t))^2 dt \quad (11)$$

where IM represents the integrated MSE. A similar result is obtained for estimating $\Im\phi_X(t)$ with a ckf $\varphi^{\Im}(t)$,

$$IM[\Im\hat{\phi}_X(t;\varphi^{\Im})] = \int_t V[\Im\hat{\phi}_X(t;\varphi^{\Im})] + (E[\Im\hat{\phi}_X(t;^{\Im})] - \Im\phi_X(t))^2 dt \quad (12)$$

4. MINIMUM INTEGRATED MSE GAUSSIAN FKCFE FOR THE GAUSSIAN CASE

For zero-mean symmetrical processes it is easy to see that optimum solutions are given by $\Im\varphi^{\Re}(t) = \Re\varphi^{\Im}(t) = \Im\varphi^{\Im}(t) = 0$, leaving us to select $\Re\varphi^{\Re}(t)$. A natural choice of ckf for estimation of a Gaussian cf would be itself Gaussian,

$$\varphi^{\Re}(t) = e^{-\sigma_{\Re}^2 t^2/2} \quad (13)$$

where $\sigma_{\Re}$ is the ckf (inverse) width. This is also expected to have a wide range of applicability amongst non-Gaussian processes. Substituting the Gaussian ckf and Gaussian cf for X into (11) and minimising with respect to σ_φ gives the optimum width for a Gaussian ckf for a Gaussian process, $\sigma_{\Re'}$,

$$\frac{1}{2\sigma_{\Re'}^3} + \frac{1}{2(2\sigma^2+\sigma_{\Re'}^2)^{3/2}} + \frac{N-1}{(\sigma^2+\sigma_{\Re'}^2)^{3/2}} - \frac{N}{(\sigma^2+\sigma_{\Re'}^2/2)^{3/2}} = 0 \quad (14)$$

An approximation to the solution (which can be obtained numerically) is given by $\sigma_{\Re'} = \sigma(2/3N)^{1/5}$. As before, numerical analysis gives a more accurate approximation as $\sigma_{\Re'} = .97\sigma N^{-.2}$.

5. CHARACTERISTIC FUNCTION BASED TEST STATISTICS

We choose to standardize our data in order to avoid having to test against all possible values of (μ, σ^2). This is approximately equivalent to simply testing the null hypothesis with $(\hat{\mu}, \hat{\sigma}^2)$ where we use the usual sample estimates, optimum under the Gaussian hypothesis. However, we shall not have to formulate our test for all possible values of $(\hat{\mu}, \hat{\sigma}^2)$).

From our data $\mathbf{X}$ we form the standardised data $\mathbf{Y} = (\mathbf{X}-\hat{\mu})/\hat{\sigma}$ and the null hypothesis $H_0 : \phi_Y(t) = \phi_o(t)$ where $\phi_o(t)$ is the cf of Y under the null hypothesis. We can note that the components of $\mathbf{Y}$ are independent and, for $N > 30$ they are approximately distributed as Gaussian.

We have experimented with many different forms of KCFE for the problem of testing Gaussianity. We note that the issue of *estimation* is quite different from that of *detection*. That is, we require our KCFE to perform quite well under the null hypothesis, but we recognise that it must also we able to adapt for the alternate hypotheses. Clearly, for example, we cannot maintain $\varphi^{\Im}(t) = 0$ if we hope to detect statistical assymetries. DKCFE's formulated for the exact purpose of adapting to the data proved to display too great a variance for the Gaussian case. Whereas AKCFE's formulated for the Gaussian case failed to perform under some alternatives (which were essentially "Gaussianified", usually as a result of the filtering of high t components in the cf domain). Generally then, the additional analytic and, usually, computational, complexity that occurs with more sophisticated KCFE techniques is not justified by any improvement in the results of their *omnibus* use. Thus we revert to the simplicitly and convenience of the Gaussian FKCFE wish we have found to have performance virtually equal to more intractable methods. We do note however that potentially huge improvements in *specific* cases (against specified alternatives) is almost always possible.

General Test Statistics. The real and imaginary parts of the cf contain quite different information and we intend to form two test statistics based on the Gaussian ckf:

$$Q_X^{\Re} = \sup_t \left| \Re \left\{ \hat{\phi}_Y(t; \varphi^{\Re}) - \phi_o^{\Re}(t) \right\} \right| \quad (15)$$

$$Q_X^{\Im} = \sup_t \left| \Im \left\{ \hat{\phi}_Y(t; \varphi^{\Im}) - \phi_o^{\Im}(t) \right\} \right| \quad (16)$$

where $\phi_o^{\Re}(t)$ and $\phi_o^{\Im}(t)$ are expectations, under the null hypothesis, of $\hat{\phi}_Y(t;\varphi^{\Re})$ and $\hat{\phi}_Y(t;\varphi^{\Im})$ respectively. Note that the absolute difference statistic was found to provide the best detection performance even when derivations based on integrated MSE were used to form the estimators. Since, in general, these two statistics are not independent (making joint tests unwieldy) it is desirable to combine them in some sensible manner in order to obtain a test using information present in both. Thus, we have formed the magnitude KCFE test as,

$$Q_X = \sup_t \left| (\Re\hat{\phi}_Y(t;\varphi^{\Re}) + j\Im\hat{\phi}_Y(t;\varphi^{\Im})) - (\Re\phi_o^{\Re}(t) + j\Im\phi_o^{\Im}(t)) \right|^2 \quad (17)$$

Other obvious forms of combination were experimented with, not altering the results significantly. $Q_X^{\Re}$ will be, theoretically, an omnibus test which is expected to be, in general, more powerful than Q_X for symmetrical alternatives and less powerful for asymmetrical alternatives. $Q_X^{\Im}$, on the other hand, is only expected to be useful for detecting asymmetrical alternatives.

Gaussian FKCFE Test Statistics. Based on the ideas expounded above we form our KCFE from a fixed Gaussian ckf, seperately estimating the real and imaginary cf components.

$$\hat{\phi}_Y(t) = e^{-\sigma_{\Re}^2 t^2/2}\Re\hat{\phi}_Y(t;1) + je^{-\sigma_{\Im}^2 t^2/2}\Im\hat{\phi}_Y(t;1) \quad (18)$$

with $\phi_o^{\Re}(t) \approx e^{-(1+\sigma_{\Re}^2)t^2/2}$ and $\phi_o^{\Im}(t) = 0$. We use the optimum value of $\sigma_{\Re}$ under the null hypothesis $\sigma_{\Re} \approx .97\sigma N^{-.2}$. Unfortunately $\sigma_{\Im} = \infty$ is optimum for the estimation problem under the null hypothesis, so this lends us no clue as to how to optimise the detection problem for the imaginary cf component. We have found empirically that $\sigma_{\Im} \approx 1.25\sigma_{\Re}$ provides good power (at least against the alternatives we chose). Finally, substituting these estimators into the test statistics (15)-(17) we calculate significance level thresholds empirically with multiple realisations of $X \sim N(\mu, \sigma^2)$. Thresholds for $N = 64$ at the 5% level of significance are found for $(Q_X^{\Re}, Q_X^{\Im}, Q_X)$ to be $(.1053, .0973, .0151)$.

6. RESULTS AND DISCUSSION

We have implemented the K-E test with three values of $\mathbf{t}$ recomended in the literature, $\mathbf{t}_1 = [.5, 1.5, 2.5, 3.0]/\hat{\sigma}$ [5], $\mathbf{t}_2 = [-3, -2, -1, 1, 2, 3]/\hat{\sigma}$ [4] and $\mathbf{t}_3 = [1, 1, 2, 2]/\hat{\sigma}$ [2]. Also, we have implemented the classical Pearson χ^2 goodness-of-fit test and two powerful modern techniques: the Shapiro-Wilk W [7] and the D'Agostino K^2 [6] tests. We have calculated *all* percentage points empirically. Generally it was found that all theoretical values were significantly in error for N not large, and tended to reject more often than they should. Clearly this would tend to indicate a test of greater power than is warranted unless one tests also under the null hypothesis and realises that the distributional assumptions are breaking down. The significance level estimates were performed with (at least) 50000 realisations, while the results for the non-Gaussian random variables were obtained with 10000 realisations. Results are shown in Table 1, with $N = 64$. We believe the results indicate the futility of attempting to construct a general test with $Q^E(\mathbf{t})$ based on a low value of T. Note that, for each choice of $\mathbf{t}$, the K-E test performs well for some alternatives (when the choice *happens* to correspond to values of the alternative's cf that are far from that of a Gaussian) but quite poorly for others (when the converse is true). We do not believe it is possible to specify a small number of t values at which *all* alternatives will be detected with large power. If the alternative hypothesis is specified this approach may be warranted. The results for the Q_X methods show a much greater *stability* as well as overall performance improvement. We note that $Q_X^{\Re}$ and $Q_X^{\Im}$ tests have provided improved performance for symmetrical and asymmetrical alternatives respectively. An interesting result is the performance of the K^2 statistic for the $T(5)$ data. We believe the reason for the complete breakdown observed here may be that the test is based only on third and fourth cumulant information and therefore cannot be truly called an omnibus test. Results for other values of N have been obtained and show the Q_X tests to be (at least) as consistent with respect to N as the other tests, but often better.

7. CONCLUSIONS

We conclude that we have devised and implemented a test for Gaussianity based on the cf which resolves the problems of implementation and performance that were associated with previously proposed methods. Additionally, we have shown the test to be a very powerful all-purpose test for Gaussianity as compared to other available methods. Further, we note that adaptions to the method to test for other densities is a simple matter, and that enhancement of the test for specified alternatives is also possible. Finally, the extension of the methodology to the cases of non-identically distributed and non-independent data remain as the goal of future research.

8. REFERENCES

[1] M. J. Arnold, A. M. Zoubir, and B. Boashash. Estimation of the characteristic function. Submitted to the IEEE Signal Processing/Athos Workshop on Higher-Order Statistics.

[2] T. W. Epps. Testing that a stationary time series is Gaussian. *The Annals of Statistics*, 15(4):1683–98, 1987.

[3] I. A. Koutrouvelis. A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, 67(1):238–40, 1980.

[4] E. Moulines, K. Choukri, and M. Charbit. Testing that a multivariate stationary time series is Gaussian. In *Proceedings of the Sixth IEEE SSAP Workshop on Statistical Signal and Array Processing*, pages 185–88, Victoria, Canada, 1992.

[5] E. Moulines, J. W. Dalle Molle, K. Choukri, and M. Charbit. Testing that a stationary time series is Gaussian: time domain vs. frequency domain approaches. In G. B. Giannakis and A. Swami, editors, *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, pages 336–40, South Lake Tahoe, California, June 1993.

[6] E. S. Pearson, R. B. D'Agostino, and K. O. Bowman. Tests for departure from normality: comparison of powers. *Biometrika*, 64(2):231–46, 1977.

[7] S. S. Shapiro, M. B. Wilk, and H. J. Chen. Comparative study of various tests of normality. *Journal of the American Statistical Association*, 63:1343–72, 1968.

[8] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*.

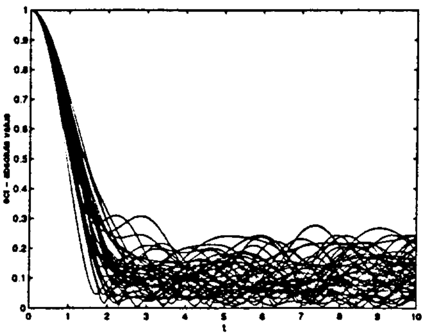

Figure 1: Realisation graph of the ecf.

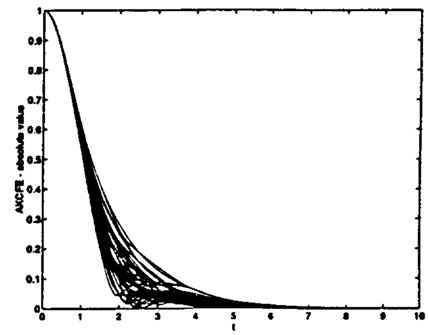

Figure 2: Realisation graph of the AKCFE.

Distn.[1]	$Q_X^{\mathfrak{E}}(t_1)$	$Q_X^{\mathfrak{E}}(t_2)$	$Q_X^{\mathfrak{E}}(t_3)$	Q_X	$Q_X^{\Re}$	$Q_X^{\Im}$	χ^2	W	K^2
$N(0,1)$	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
$U(0,1)$	62.3	99.0	94.4	87.3	88.5	7.5	30.5	97.3	90.1
χ_2^2	99.8	92.4	98.1	100.0	64.5	99.9	95.1	100.0	99.6
χ_8^2	50.6	36.9	47.5	65.1	17.0	75.0	25.8	82.1	63.1
L	41.3	10.7	45.0	70.7	72.1	14.8	29.5	45.9	57.1
$K(1,1)$	90.7	60.8	72.7	88.8	27.7	93.3	44.7	96.9	83.5
$LN(0,1)$	100.0	98.4	99.5	100.0	94.1	100.0	99.6	100.0	100.0
t_2	82.0	57.4	84.0	94.4	94.9	33.3	72.3	87.8	91.0
$T(5)$	31.7	20.5	15.9	37.1	42.1	6.8	23.6	22.9	0.5
$\beta(4,4)$	6.3	17.3	17.1	9.9	11.1	7.3	5.9	13.1	5.8
$T-U$	4.3	10.6	9.6	6.0	6.0	4.7	5.5	7.4	3.3

Table 1: Percentage of realisations rejected for null hypothesis of Gaussianity at 5% level of significance.

[1] U: Uniform, L: Laplace, K: K-distribution, LN: Log-Normal, T: Tukey, $T-U$: Sum of three independent $U(0,1)$ random variables