

**Second Order Expansions for the Distribution of the
Maximum Likelihood Estimator of the
Fractional Difference Parameter**

By

**Offer Lieberman
Peter C.B. Phillips**

July 2001

COWLES FOUNDATION DISCUSSION PAPER No. 1308



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281**

<http://cowles.econ.yale.edu/>

Second Order Expansions for the Distribution of
the Maximum Likelihood Estimator of the
Fractional Difference Parameter¹

Offer Lieberman

Faculty of Industrial Engineering and Management

Technion–Israel Institute of Technology

and

Peter C.B. Phillips

Cowles Foundation for Research in Economics

Yale University

University of Auckland & University of York

April 10, 2001

¹Work on this paper commenced while Lieberman was visiting the Cowles Foundation during the fall 2000. Lieberman thanks the Cowles Foundation for support and hospitality during this visit. Phillips thanks the NSF for support under Grant No. SBR 97-30295.

ABSTRACT

The maximum likelihood estimator (MLE) of the fractional difference parameter in the Gaussian ARFIMA(0, d , 0) model is well known to be asymptotically $N(0, 6/\pi^2)$. This paper develops a second order asymptotic expansion to the distribution of this statistic. The correction term for the density is shown to be independent of d , so that the MLE is second order pivotal for d . This feature of the MLE is unusual, at least in time series contexts. Simulations show that the normal approximation is poor and that the expansions make significant improvements in accuracy.

Key Words: ARFIMA; Edgeworth expansion; Fractional differencing; Pivotal statistic.

1. INTRODUCTION

The simplest long-memory model is the ARFIMA(0, d , 0) model. In this model the short memory component has a flat spectrum and the long memory component depends on the fractional difference parameter d . If $0 < d < 1/2$ the process is stationary long-memory with hyperbolically decaying autocorrelations, for $d > 1/2$ the process is non-stationary long-memory, for $-1/2 < d < 0$ the process is anti-persistent with autocovariances that sum to zero (producing a zero spectrum at the origin) and for $d = 0$ the process is iid. Given the widely differing characteristics of the process for different d -values it is hardly surprising that its estimation attracted such a great deal of interest over recent years. The literature is now vast and covers many different approaches allowing for parametric structures such as ARFIMA systems and semi-parametric structures where the short memory component is specified in terms of the behavior of its spectrum in a neighbourhood of zero.

The present paper focuses on the maximum likelihood estimator (MLE) of the parameter vector θ in a simple ARFIMA(0, d , 0) model with Gaussian innovations. If the variance $\sigma^2 = 1$ is known, then $\theta = d$. Let $\hat{\theta}$ be the MLE of θ in this case and set $\hat{\delta} = \sqrt{n}(\hat{\theta} - \theta)$. It is well known that $\hat{\delta}$ is asymptotically distributed as $N(0, \frac{6}{\pi^2})$. It is apparent that, unlike the AR(1) model, $\hat{\delta}$ is asymptotically pivotal. That is, the asymptotic distribution of $\hat{\delta}$ is independent of unknown parameters. We are motivated to explore higher order theory for $\hat{\delta}$ for two reasons. The first is to see whether the pivotal character of $\hat{\delta}$ extends to higher orders. The second is to assess the adequacy of the asymptotic distribution in small samples and see how well higher order asymptotic terms correct the discrepancy.

The paper derives an Edgeworth expansions for the distribution of $\hat{\delta}$. While our model is simple it is the leading canonical case and it is the first analytic attempt to extract the explicit form of the Edgeworth approximation of the distribution of the long memory estimator. In this sense it continues in the tradition of Phillips (1977), which developed the explicit form of the Edgeworth expansion of the MLE of the autoregressive coefficient in the canonical first order autoregression. Taniguchi (1984) derived similar expansions for estimators in stationary ARMA models. Using

results of Fox and Taqqu (1987), Dahlhaus (1989) and Lieberman et al. (2001), we extend Taniguchi's work to the long memory ARFIMA(0, d , 0) case. We show that to order $o(n^{-1/2})$ the density expansion is independent of d , so that $\hat{\delta}$ is second order pivotal. This feature of the MLE seems to be rare in time series contexts and so the present case is quite unusual. Simulations indicate that the normal approximation is poor but that the expansions provide significant improvements in accuracy. The expansions are valid asymptotic series in the mathematical sense.

The plan for the rest of the paper is as follows. In section 2 we derive expansions for the cumulant generating function (cgf) of $\hat{\delta}$ for a general, stationary, Gaussian, long-range dependent sequence. The errors of the expansions are $O(n^{-1})$ and $o(n^{-1/2})$, depending on whether exact or approximate cumulants are used. Section 3 applies the results to the fractional Gaussian noise model and provides both density and cdf expansions. Section 4 reports a simulation study evaluating the accuracy of the normal approximation and the expansions.

2. APPROXIMATE CGF OF THE MLE UNDER LONG RANGE DEPENDENCE

Let $\{X_t\}$, $t \in \mathbf{Z}$, be a stationary, zero mean, Gaussian long-memory process, with spectral density $f_\theta(\lambda)$, where $\theta \in \Theta \subset R^p$. The log likelihood is given by

$$\ell(\theta; x) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma_n(f_\theta) - \frac{1}{2} x' \Sigma_n^{-1}(f_\theta) x,$$

where $x = (X_1, \dots, X_n)$ is the sample, and $\Sigma_n(f_\theta)$ is the covariance matrix. The key assumption made about the spectral density is that

$$f_\theta(\lambda) \sim |\lambda|^{-\alpha(\theta)} L_\theta \text{ as } \lambda \rightarrow 0,$$

where $0 < \alpha(\theta) < 1$ and L_θ is slowly varying at zero. A full set of assumptions that assure asymptotic normality is given in Assumptions A0–A9 of Dahlhaus (1989). These assumptions are satisfied by the stationary ARFIMA(p, d, q) model. Under these conditions, $\hat{\delta}$ is asymptotically normal at the usual \sqrt{n} -rate (Dahlhaus, 1989).

In the following we use the summation convention. For brevity, we shall omit the dependence of the null cumulants of the loglikelihood derivatives (LLD's) on n

and on θ . From McCullagh (1987, p. 209), the first three cumulants of $\hat{\delta}$ are given by

$$E(\hat{\delta}^r) = -\frac{1}{\sqrt{n}}\kappa^{r,s}\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})/2 + O(n^{-3/2}) \quad (1)$$

$$\text{cov}(\hat{\delta}^r, \hat{\delta}^s) = \kappa^{r,s} + O(n^{-1}) \quad (2)$$

$$\text{cum}(\hat{\delta}^r, \hat{\delta}^s, \hat{\delta}^t) = \frac{1}{\sqrt{n}}\kappa^{r,i}\kappa^{s,j}\kappa^{t,k}(\kappa_{ijk} - \kappa_{i,j,k}) + O(n^{-3/2}), \quad (3)$$

where

$$\kappa_{r,s} = n^{-1}E(U_r U_s), \quad \kappa_{r,s,t} = n^{-1}E(U_r U_s U_t)$$

$$\kappa_{r,st} = n^{-1}E(U_r U_{st}), \quad U_r = \partial\ell/\partial\theta^r, \quad U_{rs} = \partial^2\ell/\partial\theta^r\partial\theta^s,$$

and so on. In this notation, $\kappa^{r,s}$ is the inverse of the Fisher information matrix. In the statements on the orders of the errors in (1)–(3) it is assumed that the mixed cumulants of the LLD's are $O(1)$. This is indeed the case as we will shortly argue. Higher order cumulants of $\hat{\delta}$ are $O(n^{-1})$ or smaller. It is easy to show (Lieberman et. al., 2000) that

$$\kappa_{r,s} = \frac{1}{2n}\text{tr}(\Sigma^{-1}\Sigma^*)_{r,s} \quad (4)$$

$$\kappa_{r,s,t} = \frac{1}{n}\text{tr}(\Sigma^{-1}\Sigma^*)_{r,s,t} \quad (5)$$

$$\kappa_{r,st} = \frac{1}{2n}\text{tr}(\Sigma^{-1}\Sigma^*)_{(r,st-2r,s,t)}. \quad (6)$$

The first term on the right side of (3) can be recovered by use of the Bartlett identity

$$\kappa_{rst} = -\kappa_{r,st}[3] - \kappa_{r,s,t}. \quad (7)$$

In (4)–(6), $(\Sigma^{-1}\Sigma^*)_{r,s} = \Sigma^{-1}\dot{\Sigma}_r\Sigma^{-1}\dot{\Sigma}_s$, $(\Sigma^{-1}\Sigma^*)_{r,st} = \Sigma^{-1}\dot{\Sigma}_r\Sigma^{-1}\ddot{\Sigma}_{st}$, etc., and where $\dot{\Sigma}_r = \partial\Sigma/\partial\theta^r$, $\ddot{\Sigma}_{st} = \partial^2\Sigma/\partial\theta^s\partial\theta^t$. By Theorem 5.1 of Dahlhaus (1989), $\kappa_{r,s}$, $\kappa_{r,s,t}$ and $\kappa_{r,st}$ are all $O(1)$ with limits

$$\lim_{n \rightarrow \infty} \kappa_{r,s} = \lim_{n \rightarrow \infty} \frac{1}{2n}\text{tr}(\Sigma^{-1}\Sigma^*)_{r,s} = \frac{1}{4\pi} \int_{\Pi} \frac{f_r(\lambda)f_s(\lambda)}{f^2(\lambda)} d\lambda \equiv I_{r,s} \quad (8)$$

$$\lim_{n \rightarrow \infty} \kappa_{r,s,t} = \lim_{n \rightarrow \infty} \frac{1}{n}\text{tr}(\Sigma^{-1}\Sigma^*)_{r,s,t} = \frac{1}{2\pi} \int_{\Pi} \frac{f_r(\lambda)f_s(\lambda)f_t(\lambda)}{f^3(\lambda)} d\lambda \equiv M_{r,s,t} \quad (9)$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \kappa_{r,st} &= \lim_{n \rightarrow \infty} \frac{1}{2n}\text{tr}(\Sigma^{-1}\Sigma^*)_{(r,st-2r,s,t)} = \frac{1}{4\pi} \int_{\Pi} \frac{f_r(\lambda)f_{st}(\lambda)}{f^2(\lambda)} d\lambda \\ &\quad - \frac{1}{2\pi} \int_{\Pi} \frac{f_r(\lambda)f_s(\lambda)f_t(\lambda)}{f^3(\lambda)} d\lambda \equiv J_{r,st}. \end{aligned} \quad (10)$$

In (8)–(10), $f_r(\lambda)$ and $f_{rs}(\lambda)$ stand for the first and second order derivatives of $f_\theta(\lambda)$ with respect to the components θ_r and (θ_r, θ_s) , respectively. In view of (1)–(3) and (7), the joint cgf of $\hat{\delta}$ is

$$\begin{aligned} K_{\hat{\delta}}(\omega) &= \frac{\omega_r \omega_s}{2} \kappa^{r,s} - \frac{\omega_r}{\sqrt{n}} \kappa^{r,s} \kappa^{t,u} (\kappa_{s,t,u} + \kappa_{s,tu}) / 2 \\ &\quad + \frac{\omega_r \omega_s \omega_t}{6\sqrt{n}} \kappa^{r,i} \kappa^{s,j} \kappa^{t,k} (-\kappa_{i,jk}[3] - 2\kappa_{i,j,k}) + O(n^{-1}). \end{aligned} \quad (11)$$

Recalling that $-\kappa^{ij} = \kappa^{i,j}$, the approximate cgf agrees with the one given by Peers and Iqbal (1985, p. 554) to $O(n^{-1})$. We may replace the cumulants appearing in (11) by their asymptotic counterparts as given by (8)–(10). Since the error rate is not established for (8)–(10), we deduce that

$$\begin{aligned} K_{\hat{\delta}}(\omega) &= \frac{\omega_r \omega_s}{2} I^{r,s} - \frac{\omega_r}{\sqrt{n}} I^{r,s} I^{t,u} (M_{s,t,u} + J_{s,tu}) / 2 \\ &\quad + \frac{\omega_r \omega_s \omega_t}{6\sqrt{n}} I^{r,i} I^{s,j} I^{t,k} (-J_{i,jk}[3] - 2M_{i,j,k}) + o(n^{-1/2}). \end{aligned} \quad (12)$$

3. THE GAUSSIAN ARFIMA(0, d , 0) MODEL

In this section we exploit (11)–(12) in deriving density and cdf expansions for the normalized MLE in the Gaussian ARFIMA(0, d , 0) model with unit variance. Denote the true parameter value by d_0 . The spectral density of the process is

$$\begin{aligned} f_d(\lambda) &= \frac{1}{2\pi} |1 - e^{i\lambda}|^{-2d} \\ &= \frac{1}{2\pi} e^{-dc(\lambda)}, \end{aligned}$$

where

$$c(\lambda) = \log[2(1 - \cos \lambda)].$$

In view of (1)–(7) the first three cumulants of $\hat{\delta} = \sqrt{n}(\hat{d} - d)$ are

$$E(\hat{\delta}) = \frac{a_1}{\sqrt{n}} + O(n^{-3/2}),$$

$$\text{var}(\hat{\delta}) = a_2 + O(n^{-1}),$$

$$\kappa_3(\hat{\delta}) = \frac{a_3}{\sqrt{n}} + O(n^{-3/2}),$$

where

$$a_1 = -a_2^2 \left\{ \frac{1}{4n} \text{tr}(\Sigma^{-1} \Sigma^*)_{d,dd} \right\},$$

$$a_2 = \left\{ \frac{1}{2n} \text{tr}(\Sigma^{-1} \Sigma^*)_{d,d} \right\}^{-1},$$

$$a_3 = a_2^3 \left\{ \frac{1}{n} \text{tr}(\Sigma^{-1} \Sigma^*)_{(d,d,d-\frac{3}{2}d,dd)} \right\}.$$

By (8)–(10), $a_i = O(1)$, ($i = 1, 2, 3$). On exponentiation of the cgf

$$K_{\hat{\delta}}(\omega) = \frac{w^2 a_2}{2} + \frac{1}{\sqrt{n}} (\omega a_1 + \frac{w^3 a_3}{6}) + O(n^{-1}),$$

we obtain

$$M_{\hat{\delta}}(\omega) = e^{\frac{w^2 a_2}{2}} \left\{ 1 + \frac{1}{\sqrt{n}} (\omega a_1 + \frac{w^3 a_3}{6}) \right\} + O(n^{-1}). \quad (13)$$

Let $\phi_X(x; 0, \tau)$ be the normal density with mean zero and variance τ evaluated at $X = x$. In the inversion of (13), we will make use of the formulae

$$\tau \omega e^{\omega^2 \tau / 2} = \int e^{w x} x \phi_X(x; 0, \tau) dx,$$

$$(3\tau^2 \omega + \tau^3 \omega^3) e^{\omega^2 \tau / 2} = \int e^{w x} x^3 \phi_X(x; 0, \tau) dx.$$

The density expansion is

$$f_{\hat{\delta}}(x) = \phi_X(x; 0, a_2) \left\{ 1 + \frac{1}{\sqrt{n}} \left[\left(\frac{a_1}{a_2} - \frac{a_3}{2a_2^2} \right) x + \frac{a_3}{6a_2^3} x^3 \right] \right\} + O(n^{-1}). \quad (14)$$

Define

$$b_0 = \frac{1}{n} \text{tr}(\Sigma^{-1} \Sigma^*)_{d,d,d}$$

$$b_2 = \frac{1}{n} \text{tr}(\Sigma^{-1} \Sigma^*)_{(\frac{3}{2}d, dd-d, d, d)}.$$

Integrating (14), the cdf expansion is

$$P(\hat{\delta} \leq x) = \Phi(x a_2^{-1/2}) + \frac{1}{6\sqrt{n}} \phi(x a_2^{-1/2}) (b_0 a_2^{3/2} + b_2 a_2^{1/2} x^2) + O(n^{-1}). \quad (15)$$

We refer to the expansions (14) and (15) as the ‘exact’ Edgeworth expansions.

The expansions (14) and (15) can be further simplified by using the integral approximations (8)–(10). Note that in this model, $(\partial f_d(\lambda) / \partial d) f_d^{-1}(\lambda) = -c(\lambda)$ and $(\partial^2 f_d(\lambda) / \partial d^2) f_d^{-2}(\lambda) = c^2(\lambda)$. This implies that (8)–(10) are independent of d . Now, from Gradshteyn and Ryzhik (1980)

$$\int_{\Pi} c^2(\lambda) d\lambda = \frac{2\pi^3}{3}, \quad (16)$$

$$\int_{\Pi} c^3(\lambda) d\lambda = -24\pi\zeta(3), \quad (17)$$

where $\zeta(\cdot)$ is the Riemann-zeta function. We therefore obtain

$$a_1 = -3\zeta(3)\left(\frac{6}{\pi^2}\right)^2 + o(1), \quad (18)$$

$$a_2 = \frac{6}{\pi^2} + o(1), \quad (19)$$

$$a_3 = -6\zeta(3)\left(\frac{6}{\pi^2}\right)^3 + o(1). \quad (20)$$

Substituting (18)–(20) into (14), the density expansion reduces to

$$f_{\hat{\delta}}(x) = \phi_X\left(x; 0, \frac{6}{\pi^2}\right) \left\{ 1 - \frac{\zeta(3)}{\sqrt{n}}x^3 + o(n^{-1/2}) \right\}. \quad (21)$$

We note that $\zeta(3) \simeq 1.202$. The correction factor $\zeta(3)x^3\sqrt{n}$ in (21) has an exceptionally simple form and does not depend on d . Hence, $\hat{\delta}$ is second order pivotal. As mentioned, this feature of the MLE is unusual in time series models. Finally, integration of (21) yields

$$P(\hat{\delta} \leq x) = \Phi\left(x\frac{\pi}{\sqrt{6}}\right) + \frac{\sqrt{6}\zeta(3)}{\pi\sqrt{n}}\phi\left(x\frac{\pi}{\sqrt{6}}\right) \left\{ x^2 + \frac{12}{\pi^2} \right\} + o(n^{-1/2}), \quad (22)$$

$\Phi(\cdot)$ and $\phi(\cdot)$ being the standard normal cdf and pdf, respectively. We refer to equations (21) and (22) as ‘approximate’ Edgeworth expansions.

Two remarks are in order. First, our results agree with those obtained by Taniguchi (1984, pp. 49–50) who dealt with a uniparameter, circular (short memory) ARMA model. Second, Lieberman et al. (2001) proved the validity of the Edgeworth expansion to the distribution of the MLE of the parameter vector of a stationary, Gaussian, strongly dependent series. While their work is primarily concerned with the question of validity of expansions, they did not explicitly derive the terms in any particular expansion. It follows from Theorem 4 of their paper that the expansions (15) and (22) are valid asymptotic series with error rates holding uniformly in x and in any compact neighborhood of the true parameter d_0 .

4. DISCUSSION AND NUMERICAL EVALUATION

Figures 1 and 2 display kernel estimates of the density of $\hat{\delta}$ in the Gaussian ARFIMA(0, d , 0) model with $n = 20$ and $d = 0.2$ and 0.4 . The MLE’s were calculated by a simple grid search on the interval $(-0.49, 0.49)$ to ensure problem free evaluation

of the covariance matrix. The number of replications was 8,000 and the densities were computed using a normal kernel and a plug-in bandwidth based on Silverman's (1986) rule. It is obvious that the densities are not symmetric. The asymmetry is partly caused by the fact that the maximization is restricted to the $(-0.49, 0.49)$ interval and is therefore more enhanced for the $d = 0.4$ case. This asymmetry is captured, at least in part, by the correction factor in (21). The breakdown of the normal approximation and the 'approximate' Edgeworth expansion are particularly vivid in the $n = 20, d = 0.4$ case (Figure 2). Here, the breakdown occurs not only at the tails, but also at the center of the distribution. The 'exact' Edgeworth expansion is more satisfactory in capturing the overall shape of the distribution in this case and makes a huge improvement at the center of the density but does so at the expense of some fluctuation, including negative density, in the tails.

We move on to evaluate the expansions to the cdf of $\hat{\delta}$. PP plots for the various approximations are provided in Figures 3-6 corresponding to $n = 20, 40$ and $d = 0.2, 0.4$. The simulated cdf is taken to be the benchmark and the closeness of the approximation to the 45 degree line indicates the accuracy of the approximation. We conducted simulation experiments for other cases and our conclusions were not altered. It is apparent that the normal approximation is poor, especially in the upper tail and in the $d = 0.4$ case. The expansion (22) based on Dahlhaus (1989) integral approximations improves significantly over the normal approximation and overall behaves quite well in the $d = 0.2$ case. However, in the $d = 0.4$ case this expansion is poor in the upper tail. The exact Edgeworth expansion (15), on the other hand, is surprisingly accurate and is decidedly superior to the other approximations in both the $n = 20$ and $d = 0.4$ case. The curve corresponding to it traces the 45 degree line very closely in the cases considered.

REFERENCES

- Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*. 17, 1749–1766.
- Fan, J. and Marron, J. S. (1994). Fast implementations of nonparametric curve estimators. *J. Comp. Graph. Statist.* 3, 35?-56.
- Fox, R. and Taqqu, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Prob. The. Rel. Fields* 74, 213–240.
- Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products*. New York: Academic Press.
- Lieberman, O., Rousseau, J. and Zucker, D. M. (2000). Small-sample likelihood-based inference in the ARFIMA model. *Econometric Theory*, 16, 231-248.
- Lieberman, O., Rousseau, J. and Zucker, D. M. (2001). Valid Edgeworth expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent series. Mimeo.
- McCullagh, P. (1987). *Tensor Methods in Statistics*. New York: Chapman and Hall.
- Peers, H. W. and Iqbal, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications. *J. Roy. Statist. Soc. B* 47, 547-554.
- Phillips, P. C. B. (1977). “Approximations to some finite sample distributions associated with a first order stochastic difference equation,” *Econometrica* 45, 463–485.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Taniguchi, M. (1984). Validity of Edgeworth expansions for statistics of time series.
J. Time Ser. Anal. 5, 37–51.

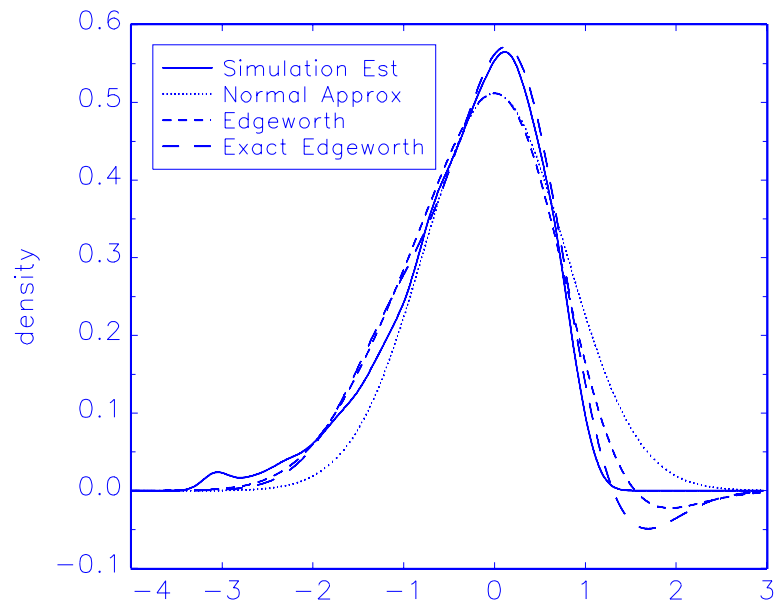


Fig 1. Density of MLE \hat{d} , $d = 0.2$, $n = 20$

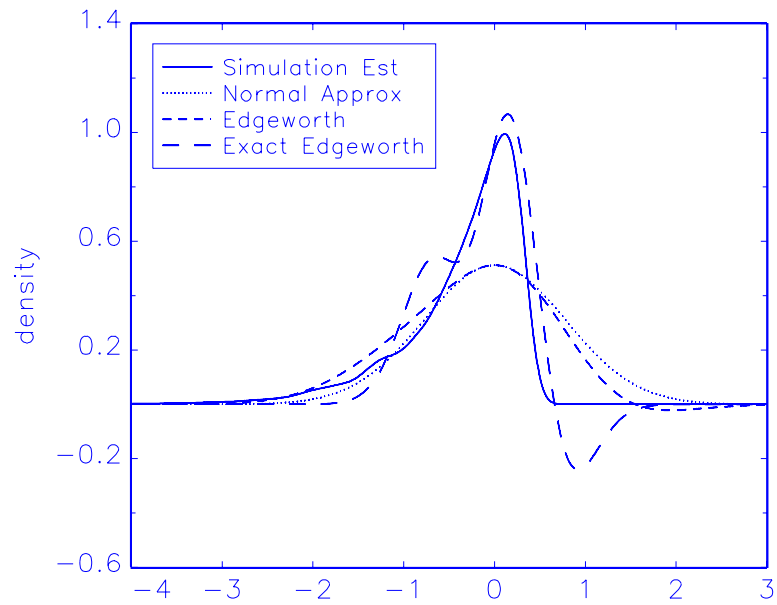


Fig 2. Density of MLE \hat{d} , $d = 0.4$, $n = 20$

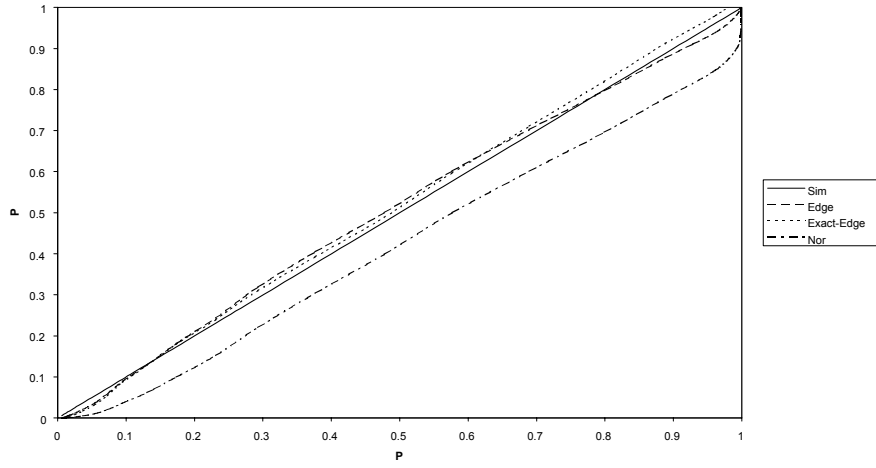


Fig 3. PP plots for the distribution of MLE $\hat{d}, d = 0.2, n = 20$

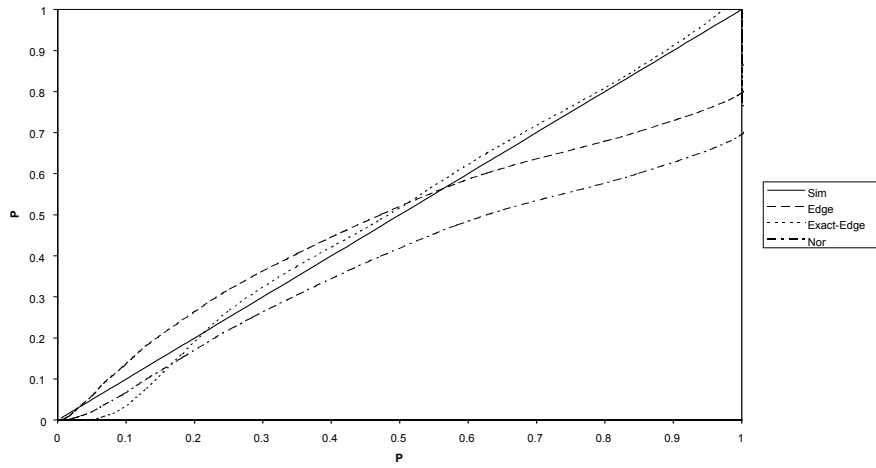


Fig 4. PP plots for the distribution of MLE $\hat{d}, d = 0.4, n = 20$

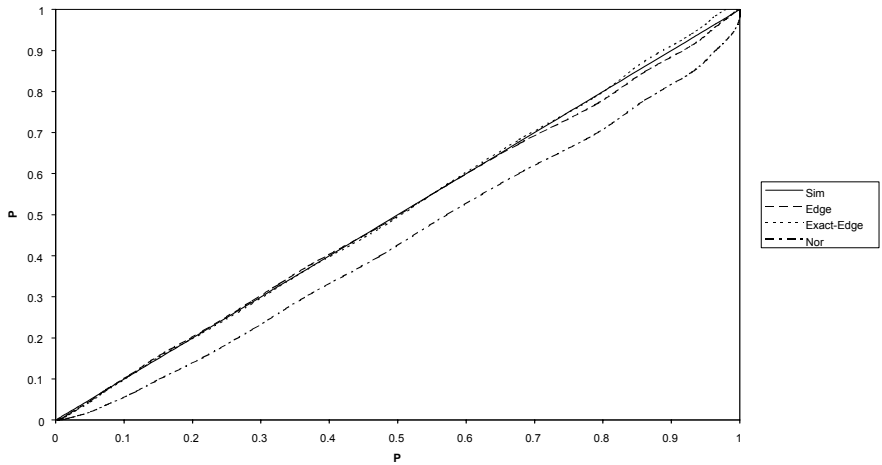


Fig 5. PP plots for the distribution of MLE $\hat{d}, d = 0.2, n = 40$

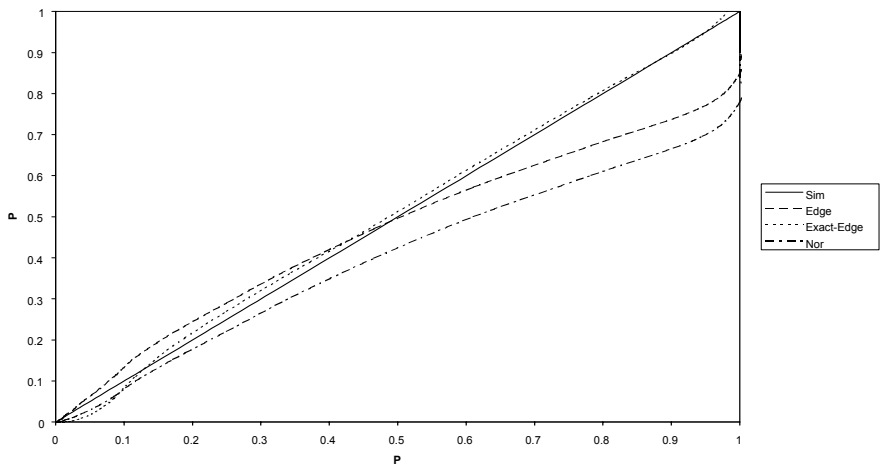


Fig 6. PP plots for the distribution of MLE $\hat{d}, d = 0.4, n = 40$