THE USE AND ABUSE OF THE $Q-Q$ PLOT:
SOME ASYMPTOTIC THEORY

by

Anirban DasGupta*
Purdue University

Technical Report #95-30

Department of Statistics
Purdue University

June 1995

*Research supported by NSF grant DMS 9307727. AMS Subject classification: 62E20, 62G30.

◇

THE USE AND ABUSE OF THE $Q-Q$ PLOT:
SOME ASYMPTOTIC THEORY

by

Anirban DasGupta*

Purdue University

Abstract

The $Q-Q$ plot is a standard device for confirming "approximate" normality of a sequence of iid observations. A curious fact from real life empirical experience is that all too frequently the $Q-Q$ plot does really look overall quite linear. This is especially the case if one looks at the central part of a $Q-Q$ plot. Any deviation from an overall linearity is also typically seen at the tailends of the $Q-Q$ plot.

In this article we investigate these empirical experiences and prove a collection of large sample theorems that justify these empirical experiences. We also show how the $Q-Q$ plot can be used as more than a diagnostic; some of our Theorems attempt to indicate how the $Q-Q$ plot may be used to learn about the true underlying population. The large sample theorems are strong laws and laws of the iterated logarithm on some statistics of importance in this context. These include extreme and studentized residuals and the correlation coefficient computed from a $Q-Q$ plot. Some real and simulated data are analyzed for illustration. All proofs are given in an appendix.

*Research supported by NSF grant DMS 9307727. AMS Subject classification: 62E20, 62G30.

1. Introduction. The $Q - Q$ plot is undoubtedly one of the most commonly used tools in applied statistics today. Indeed, it seems to have a universal appeal: frequentists, Bayesians, and practitioners in general really quite routinely use the $Q - Q$ plot, typically as a justification for the further normal theory methods they use. In other words, the $Q - Q$ plot is usually a device for convincing others that there is no red flag: its use is more as a green signal for proceeding with subsequent work. The common standard we apply to ourselves and teach our students appears to be that one should not overreact to wiggles in the plot and as long as the plot gives a visual impression of overall approximate linearity, it is generally safe to proceed with methods that assume normality. We looked up the sections on $Q - Q$ plots in five of the most standard texts on statistical methods: with some variations in the actual words used, this was the standard recommended in all of them. We found the presentation in Moore and McCabe (1993) to be somewhat more critical than in the other texts.

Methods that are very basic and are very widely used automatically face a very real danger: abuse due to routine use. The most widely used methods need the most scrutiny. The purpose of this article is to take a critical look at how we as a profession use the $Q-Q$ plot and teach our students everyday, what is justified and what probably is not, and try to understand what the $Q - Q$ plot says and what it does not say. Why is it that all too frequently we see data sets that produce an approximately linear $Q - Q$ plot? Can we do some mathematics that explains this life experience? Why is it that the central part of a $Q - Q$ plot looks normal so universally? Why is it that if there is a problem, it always appears to be at the tails of the $Q - Q$ plot? And when we see such a problem, can we try to understand what we are seeing and learn from it? For example, if we see a $Q - Q$ plot in which the central part looks very very linear but there is a systematic downward bent at the upper tail, can we actually say something more than just saying that there is a problem? Can we do some mathematics that explains how particular shapes may associate with particular true underlying populations?

In this article, we present some asymptotic theory that tries to answer these types of questions. There are two types of results: strong laws, and a collection of results that go a step beyond the strong laws in the sense of establishing some actual rates of convergence. In arriving at these results, we needed to use two types of results: some classic asymptotic theory like laws of the iterated logarithm for partial sums of iid sequences, plus some asymptotic theory for extreme observations from iid sequences. For the classical asymptotic theory we needed, Serfling (1980) is a standard reference. For the asymptotic theory of extremes we needed, we recommend Galambos (1978) as a very useful reference.

The following is a brief nontechnical outline of the subsequent sections: in Section 2, we give a collection of Theorems that address our life experience that we often see $Q-Q$ plots that look remarkably linear, especially so in the central part of the plot. In Section 3 we ask how we can learn from $Q-Q$ plots about the true underlying population; this is done by means of a collection of strong laws for extreme residuals; studentized residuals are considered in section 4. Two illustrative examples are presented in Section 5. Section 6 is an appendix containing the proofs.

2. Why $Q-Q$ Plots Tend to Look Linear.

2.1. Notation. The exact definition of a $Q-Q$ plot varies a little from source to source. The asymptotic nature of our results makes this distinction unimportant. We will define the $Q-Q$ plot as a plot of the pairs $(z_{(i-\frac{1}{2})/n}, X_{(i)})$, where z_α is the 100 αth quantile of the $N(0,1)$ distribution and $X_{(i)}$ is the ith sample order statistic (in other places, $z_{(i-\frac{1}{2})/n}$ may for instance be replaced by $z_{i/(n+1)}$: this is what we mean by the difference in the definition at different places). As usual, Φ will denote the CDF of the $N(0,1)$ distribution. Also we will write simply z_i in place of $z_{(i-\frac{1}{2})}/n$.

As an index of how linear a plot on the plane looks to the human eye, we will use the coefficient of linear correlation between the coordinates in the plot. In the context of $Q-Q$ plots, we therefore have the following index of linearity:

$$r_n = \frac{\sum\limits_{i=1}^{n} z_i(X_{(i)} - \bar{X})}{\sqrt{\sum\limits_{i=1}^{n} z_i^2 \cdot \sum\limits_{i=1}^{n} (X_i - \bar{X})^2}} = \frac{\sum\limits_{i=1}^{n} z_i X_{(i)}}{\sqrt{\sum\limits_{i=1}^{n} z_i^2 \cdot \sum\limits_{i=1}^{n} (X_i - \bar{X})^2}} \left(\because \sum_{z=1}^{n} z_i = 0\right) \tag{2.1}$$

As we stated earlier, we will also try to understand the central part of the $Q-Q$ plot separately from the whole plot itself; this will therefore necessitate a separate index of linearity for the central part and a specification of what we mean by the "central part".

This is done by defining, for $0 < \alpha < .5$, the trimmed correlation

$$r_\alpha = r_{\alpha,n} = \frac{\sum_{i=k+1}^{n-k} z_i X_{(i)}}{\sqrt{\sum_{i=k+1}^{n-k} z_i^2 \cdot \sum_{i=k+1}^{n-k} (X_{(i)} - \bar{X}_k)^2}}; \tag{2.2}$$

where $k = [n\alpha]$, and $\bar{X}_k$ is the trimmed mean.

This definition therefore says that r_α is the correlation in the $Q - Q$ plot when 100 α% of the points are trimmed from each end of the plot. Therefore, according to our empirical experience, we tend to see a larger value of r_α if α is larger because this corresponds to more trimming. We will assume in this article that the true underlying CDF F is continuous, although it is not necessary to assume this for most of the results. For previous work in the spirit of our results, one should see Filliben (1975), LaBrecque (1977), Anderson and Darling (1954), Shapiro and Wilk (1965) and Seber (1984) for some unified information.

2.2. Almost Sure Limit of r_n and r_α.

Theorem 2.1. Let $X_1, X_2, \ldots$ be iid observations from a distribution with CDF F with finite variance $\sigma^2 = \sigma^2(F)$. Then $r_n \to \rho(F)$ with probability 1, where

$$\rho(F) = \frac{\int_0^1 F^{-1}(x)\Phi^{-1}(x)dx}{\sigma}. \tag{2.3}$$

Proof. The proof is deferred till the appendix.

There is an interesting lower bound on the limiting correlation $\rho(F)$ for symmetric F that comes out of (2.3).

Corollary 1. Let F be any CDF with a point of symmetry, say θ, and in addition with a finite variance σ^2. Then

$$\rho(F) \geq \sqrt{\frac{2}{\pi}} \frac{E_F|X-\theta|}{\sigma}.$$

In addition, for any $0 < p < \frac{1}{2}$,

$$\begin{aligned}\rho(F) &\geq \sqrt{2}\frac{(E_F|X-\theta|^p)^{\frac{1}{p}} \cdot (\Gamma(\frac{1-2p}{2-2p}))^{\frac{p-1}{p}}}{\sigma} \\ &= \frac{\sqrt{2}}{\sigma}c(p) \quad \text{(say)}\end{aligned}$$

and hence,

$$\rho(F) \geq \max\{\sqrt{\frac{2}{\pi}}\frac{E_F|X-\theta|}{\sigma}, \quad \frac{\sqrt{2}}{\sigma}\sup_{0<p<\frac{1}{2}} c(p)\}. \qquad (*)$$

Discussion of Corollary 1. Corollary 1 is valuable in the following sense: direct use of formula (2.3) requires evaluation of the quantiles of F. Except in quite special cases, there are usually no closed form expressions for the quantiles. Thus (2.3) has to undergo a numerical integration or Monte Carlo followed by a numerical evaluation of the quantiles themselves. In contrast, Corollary 1 only involves the moments of F which require only the density for which almost always there will be a closed form expression in practice; in fact even the moments may have closed form formulae. In addition, the strengthened inequality (*) stated at the end gives sharper lower bounds on $\rho(F)$. Because of all of these reasons, Corollary 1 is useful.

Notice that Theorem 2.1 needs finiteness of the variance of F. One may suspect that if the underlying CDF does not have a finite variance, then there will be a drastic departure from overall linearity at the tailends of the $Q-Q$ plot, resulting in a small value of the correlation r_n. The following result justifies this intuitive expectation.

Proposition 2.2. Let $\mathcal{F}$ be the family of normal scale mixtures; i.e., $X \underset{F}{\overset{L}{=}} \mu + \sigma\tau Z$, where μ, $\sigma > 0$ are any arbitrary fixed constants, $Z \sim N(0,1)$ and $\tau \geq 0$ is distributed independently of Z. Then

$$\inf_{F\in\mathcal{F}} \rho(F) = 0.$$

Again, a proof is postponed till the appendix.

Interestingly, the limiting correlation $\rho(F)$ may be close to zero even when F is not at all thick tailed, but in fact very concentrated. The intuitive explanation is that if F is so concentrated that it acts like a point mass, then the corresponding $Q-Q$ plot is likely to look like a flat line, as though the exact value of z_i had no effect on $X_{(i)}$. The actual mathematics is more intricate than this simplistic explanation because σ in the denominator of the formula for $\rho(F)$ is also very small if F acts like a point mass. The following example is a good illustration.

Example 1. Let $X \underset{F}{\overset{L}{=}} U^n$ where $U \sim$ Uniform $[0,1]$ and $n > 0$. We will ultimately let

$n \to \infty$. Then, from (2.3),

$$\begin{aligned}\rho(F) &= \int^1 x^{\frac{1}{n}} \Phi^{-1}(x)dx \cdot \frac{n+1}{n}\sqrt{2n+1} \\ &= \frac{n+1}{n}\sqrt{2n+1} \int_0^{\frac{1}{2}} \left\{(1-x)^{\frac{1}{n}} - x^{\frac{1}{n}}\right\} \Phi^{-1}(1-x)dx\end{aligned} \tag{2.4}$$

Using now the fact that $\Phi^{-1}(1-x) = O(\sqrt{\log \frac{1}{x}})$ as $x \to 0$, and making the change of variable $y = \log \frac{1}{x}$, one gets that

$$\begin{aligned}&\int_0^{\frac{1}{2}} \left\{(1-x)^{\frac{1}{n}} - x^{\frac{1}{n}}\right\} \Phi^{-1}(1-x)dx \\ &= O(\frac{1}{n}) \text{ since } \int_0^{\infty} e^{-y} y^{\frac{3}{2}} dy < \infty\end{aligned} \tag{2.5}$$

Combining (2.4) and (2.5), one has

$$\rho(F) = O(\frac{1}{\sqrt{n}}) \to 0 \text{ as } n \to \infty.$$

Discussion and application of the Theorem. First a curious technical point. Since $\rho(F)$ is the almost sure limit of a correlation sequence, it follows $|\rho(F)| \leq 1$ for any F with 2 moments. This can actually be proved by a direct and tricky argument, and is a nice inequality by itself. Actually, it can be shown that $\rho(F) \geq 0$, since $F^{-1}(\alpha)$ and $\Phi^{-1}(\alpha)$ are both nondecreasing functions of α: essentially the fact that the covariance of two nondecreasing functions of a (uniform) random variable must be nonnegative. This of course makes sense since the correlation sequence r_n is positive itself by construction.

Notice also that symmetry of F was not assumed in the Theorem above. Thus this Theorem can be used to quantify how linear to the eye the $Q-Q$ plot will look for skewed populations as well. The following table gives the computed values of the limiting correlation $\rho(F)$ for a number of standard distributions. We find it really quite amazing how large the linear correlations are across this broad spectrum of examples. In the following table, F_n denotes the CDF of the maximum of n iid $N(0,1)$ variables.

Table 1.

F	$\rho(F)$
Normal	1
Uniform	$\sqrt{\frac{3}{\pi}} = .977205$
Double Exponential	.981051
$t(3)$	.90077
$t(5)$	.983236
$t(7)$	.993375
Tukey distribution (.9N(0,1)+.1N(0,3))	.970633
Logistic	.966273
F_2	.986636
F_4	.985113
F_{10}	.982345
Extreme value (with density $\exp(-\exp(-x))\exp(-x)$)	.98

Of these, the last four distributions are skewed and the others are symmetric (a location and scale can be added in an arbitrary way in all of these examples because we are considering correlations). These numbers obtained by an application of Theorem 2.1 give an idea of why we tend to see linear looking $Q-Q$ plots in our empirical experience. Also see Tables 2 and 3 where the value of r_n is reported for 25 simulated samples of size $n = 30$ each for the uniform and the Double Exponential case. This also leads to the fact that if $Q-Q$ plots are to help in deciding whether one has normal data or not, overall linearity should not be the standard of judgement. One has to go into finer features of the $Q-Q$ plot, like its shape at the tails. But first let us understand the qualitative behavior as regards visual linearity of the central part of a $Q-Q$ plot. The following Theorem is an analog of Theorem 2.1.

Theorem 2.3. Let $0 < \alpha < .5$. Suppose $X_1, X_2, \ldots, X_n$ are iid observations from a CDF

F on the real line. Let $\mu_\alpha = \frac{\int\limits_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dF(x)}{2\alpha}$. Then $r_\alpha \to \rho_\alpha(F)$ with probability 1, where

$$\rho_\alpha(F) = \frac{\int_\alpha^{1-\alpha} F^{-1}(x)\Phi^{-1}(x)dx}{\sqrt{\int_\alpha^{1-\alpha}\{\Phi^{-1}(x)\}^2 dx \cdot \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)}(x-\mu_\alpha)^2 dF(x)}} \tag{2.6}$$

The Proof of Theorem 2.2 is very similar to that of the preceding Theorem 2.1. However, a brief outline of the proof is indicated for completeness in the appendix.

In all cases that we tried, the limiting correlation of the central part, i.e. $\rho_\alpha(F)$ of Theorem 2.3, was found to be increasing in α. In other words, the more one trims the more linear the Q-Q plot looks, probably an anticipated fact. We believe that an analytical proof of this monotonicity in α should not be difficult although we didn't try it. It is our expectation that simply calculus should suffice for this monotonicity result under some conditions on F. Figure 1 is a plot of $\rho_\alpha(F)$ for F = Uniform and the Double Exponential. One is tempted to conclude that uniform data look more normal than other types of data in the central part of a $Q-Q$ plot. The other interesting fact one sees from Figure 1 is that even 5% trimming from both ends of the plot results in a correlation as large as about .994 in each case. The central part of a $Q-Q$ plot is likely to look nearly perfectly linear in large samples with just a small amount of trimming from each end, again signifying that the information about the population can only come from finer features of the $Q-Q$ plot.

Table 2

Correlation in $Q-Q$ plot for simulated uniform data

($n = 30$, 25 sets)

.962	.966	.969	.987	.970
.994	.973	.975	.969	.969
.977	.979	.982	.981	.969
.989	.975	.973	.985	.976
.986	.964	.975	.972	.981

Table 3

Correlations in $Q - Q$ plot for simulated Double Exponential data

($n = 30$, 25 sets)

.920	.952	.978	.857	.990
.963	.962	.975	.985	.974
.921	.980	.956	.934	.955
.992	.993	.942	.925	.960
.986	.973	.989	.907	.975

The following Table is given to help understand the effect of trimming on the linearity of the $Q - Q$ plot.

Table 4.

F	Limiting correlation		
	No trimming	5% trimming	10% trimming
Uniform	.977205	.994937	.997846
Double Exp	.981051	.994055	.996242
$t(3)$	.90077	.998365	.999601
Logistic	.966273	.999526	.999816

3. Extreme residuals. The findings of the preceding section indicate that finer features of a $Q - Q$ plot such as its shape at the tailends are necessary to look at over and above just the overall linearity of the $Q - Q$ plot. In order to make the point that tailends of the $Q - Q$ plot are instructive, we have included simulated $Q - Q$ plots for data arising from uniform, normal,Double Exponential and $t(6)$ distributions. Note the following interesting features in these simulated $Q - Q$ plots (Figures 2 through 6):

a) For the uniform data, the last point is below the least squares line which is given on the same plot;

b) For the normal data, the last point is above this same line;

c) For Double Exponential data, it is above the line as well, but seems to be more above than it was for the normal data;

d) For the $t(6)$ data, of which there are two simulated plots, in both cases the last point is above the line, but the magnitude by which they are above are noticeably different in the two simulated sets.

Admittedly, these are one time simulations and perhaps mean nothing. Interestingly, we will now state a collection of almost sure results that vindicate some of the the general phenomena we noticed in these limited simulations. We give these one at a time. But first let us clearly explain notation and exactly what we are doing.

Consider the cluster of points $(z_i, X_{(i)})$ that form the $Q-Q$ plot. Fit the least squares line of fit for these pairs which has the equation

$$\hat{X}_{(i)} = a + bz_i$$

where

$$b = b_n = \frac{\sum_{i=1}^{n} z_i X_{(i)}}{\sum_{i=1}^{n} z_i^2} \tag{3.1}$$

and $a = \bar{X} - b\bar{z} = \bar{X}$ since the $\{z_i\}$ add to zero.

The residuals are defined in the usual way:

$$e_i = \text{ Observed } X_{(i)} - \text{ Predicted } \hat{X}_{(i)}.$$

In particular, the residual corresponding to the very last point in the $Q - Q$ plot is the extreme residual

$$e_n = X_{(n)} - \bar{X} - z_n.b.$$

The almost sure results we are now stating are about the behavior of these quantities e_n. The proof for the normal case needed the most work. We will outline these proofs in the appendix. There is a unifying thread in all of these proofs for the different distributions: e_n involves in its expression three types of quantities: the maximum order statistic $X(n)$, the sample mean $\bar{X}$, and the slopes b_n. The unifying thread in the proofs is that $X_{(n)}$ is handled by quite delicate limit results in Extreme value theory, $\bar{X}$ is handled by more standard laws of the iterated logarithm for sample means, and b_n is handled by limit results on L statistics, by and large results due to Wellner. We will now state the almost sure results concerning e_n.

Theorem 3.1. For each of $F =$ uniform, Double Exponential, there is a sequence c_n such that the extreme residual $e_n \sim c_n$, i.e., $e_n/c_n \to 1$ with probability 1. The sequence c_n are as follows:

$$c_n = -\sqrt{\log n}/\sqrt{2\pi} \qquad \text{for } F = \text{uniform } [0,1];$$

$$c_n = \log n \qquad \text{for } F = \text{Double Exponential};$$

for $F = N(0,1)$, $\limsup\{\frac{e_n\sqrt{2\log n}}{\log\log n}\} \leq 1$ and $\liminf\{\frac{e_n\sqrt{2\log n}}{\log\log n}\} \geq 0$;

For $F = t(m)$ where m denotes the degree of freedom of the t distribution, the following holds:

$$\limsup e_n/n^{\frac{1}{m}} = \infty \text{ with probability } 1$$

$$\liminf e_n/n^{\frac{1}{m}} = 0 \text{ with probability } 1.$$

Discussion of Theorem 3.1. The sign of c_n is positive for Double Exponential data, i.e., the observed point is going to be more than the predicted and hence above the least squares line. In the uniform case, the sign of c_n is actually reversed: so now one is going to see the last point BELOW the line. And for data from t distributions, there is too much fluctuation from sample to sample and there is no almost sure limit as the zero $\liminf$ and infinite $\limsup$ testify to. Note that each assertion is intriguingly consistent with what we saw in the simulated $Q-Q$ plots discussed above. We find this very appealing and instructive ourselves. We were not able to strengthen the result for the normal case to the existence of a limit.

4. Studentized residuals and learning about the true population from the $Q-Q$ plot. The results on extreme residuals presented in Section 3 give some qualitative understanding of the shape of the $Q-Q$ plot as a function of the true underlying CDF for a few standard cases. There is one fundamental difficulty in learning about the true CDF by visual examination of the residuals; it is that residuals are not scale invariant. It is therefore valuable to study the scaled residuals as well so that for every distribution within a given location-scale family one sees the same asymptotic behavior. A standard method towards this end is to consider the studentized residuals

$$se_i = \frac{e_i}{\sqrt{\frac{1}{n}\sum_{j=1}^{n} e_j^2}}. \tag{4.1}$$

More formally, take any number α such that $0 < \alpha < 1$, and take $i = [n\alpha]$ (the integer part of $n\alpha$); define now the studentized residual indexed by α as

$$e^*_\alpha = se_i.$$

Then it turns out that for each α in the interval (0,1), there is an almost sure limit for e^*_α provided the true CDF F has two moments. The following result gives the expression for this almost sure limit.

Theorem 4.1. Let $X_1, X_2, \ldots, X_n$ be iid observations from a CDF F on the real line. Assume F has two moments. Then e^*_α converges with probability 1 to

$$s(\alpha) = \frac{F^{-1}(\alpha) - (\int_0^1 F^{-1}(x)\Phi^{-1}(x)dx)\Phi^{-1}(\alpha)}{\sqrt{\sigma^2(F) - (\int_0^1 F^{-1}(x)\Phi^{-1}(x)dx)^2}} \tag{4.2}$$

The derivation of this expression is given in the appendix.

Evaluation of $s(\alpha)$ for a given F will usually require numerical computing in the sense it would not be possible to write $s(\alpha)$ in terms of elementary functions. In certain special cases, however, basically closed form expressions for $s(\alpha)$ are possible. For example, if F is Uniform, then $s(\alpha) = (2\alpha - 1 - \Phi^{-1}(\alpha)/\sqrt{\pi})/\sqrt{(1/3 - 1/\pi)}$ and if F is Double Exponential, then $s(\alpha) = -3.64961(\log(2(1-\alpha)) + 1.387416\Phi^{-1}(\alpha))$ for $\alpha \geq \frac{1}{2}$ and 3.64961 $(\log(2\alpha) - 1.387416\Phi^{-1}(\alpha))$ for $\alpha \leq \frac{1}{2}$. The constants with decimals are approximate, of course. Figure 7 gives a plot of the function $s(\alpha)$ for three common cases: uniform, Double Exponential, and the t distribution with 3 degrees of freedom. The 3 degree of freedom was chosen as a type of extreme slow tail with a finite variance. Other degrees of freedom can also be studied. Other choices of F should also be studied. These plots promise to be useful in the following way: given a real data set, one computes the studentized residuals, omitting a few extreme ones due to the assumption of a fixed α in the Theorem. As a rule of practice, may be the middle 95% of the residuals could be computed. If a smoothed version of the observed studentized residuals generally matches the theoretical $s(\alpha)$ plot for some standard F, then we have indeed learned from the $Q-Q$ plot more than whether the data are normal or not normal. We have now better reasons to believe that the true underlying CDF is the one whose theoretical $s(\alpha)$ plot resembles the observed smoothed studentized residual plot. Smoothing can be done in many standard ways. We have simply used the List Plot command in Mathematica on the list of observed studentized residuals. Certainly more sophisticated smoothing can be also done. We will present one real data set and one simulated data set for illustration.

5. Examples.

Example 1. We will illustrate the potential for typical use of Theorem 4.1 by considering the so called water salinity data set. This is a data set of 28 observations on the salinity level in North Carolina's Pamlico sound; the observations are biweekly. As far as we know, this data set was first studied in Ruppert and Carroll (1980); it was subsequently analyzed in Hettmansperger (1987), Rousseeuw and Leroy (1987) and Ruppert and Carroll (1985). There is a highly enjoyable description of the original background of this data set in Ruppert and Carroll (1985). Note that the basic goal of the statistical analysis done in these papers with the salinity data set was different. Note also that there is clearly a time series character inherent in this data set and there is apriori no reason to model this data as iid realizations of a fixed CDF.

If we proceed anyway for the time being and calculate the studentized residuals after fitting the ordinary least squares regression of the sample order statistics on the $N(0,1)$ percentiles, then something interesting happens. The observed studentized residuals (joined by line segments) are plotted in Figure 8. Now we recognize this plot to resemble the theoretical $s(\alpha)$ plot for the Uniform family. For purposes of comparison, this theoretical plot is superimposed on the same plot. Note that in the theoretical $s(\alpha)$ plot for the Uniform distribution in Figure 7, the ordinates go to 6 in the range plotted, but in the plot under consideration for the given data set, the studentized residuals do not go that far; this is because Figure 7 was plotted for between .001 and .999, a much larger range than 2/28 to 27/28 which is the effective range in the observed residual plot. We find this interesting; we have undoubtedly a time series character in the data set. Yet a model of iid observations from a Uniform distribution may also be something to look at, it appears. Note that this analysis does not say anything about the location and the scale of the Uniform distribution. These have to be estimated in a serious actual study by using standard estimation methods. Perhaps robust alternative estimates may also be used. We will not go into that aspect of the analysis since that is not the goal of this article.

Example 2. We will now give a simulated example for 60 simulated observations from a standard Double Exponential distribution. The purpose of the example is to see whether the simulated plot indeed looks like the theoretical plot. This is important for purposes of identification. Since the steps leading to the plots are exactly the same as in Example 1, we omit discussions of that aspect. Figures 9 and 10 give the observed studentized residuals together with the theoretical plot corresponding to the Double Exponential case. We find the similarity striking indeed.

6. Appendix.

Proof of Theorem 2.1. The idea of the proof is actually very simple. One handles the numerator and the denominator separately. To do this, however, one multiplies the numerator by $1/n$ and each term within the square root sign in the denominator is also multiplied by $1/n$.

Once this is done, the first term in the denominator converges to $\int_0^1 \{\Phi^{-1}(x)\}^2 dx$ since it is a Riemann sum of that integral; the integral is seen to be 1. Thus the first term within the square root converges to 1.

The second term within the square root in the denominator converges with probability 1 to the variance of the underlying distribution. On taking the square root, the variance becomes the standard deviation.

The numerator now remains to be handled. This is essentially done in Example A* in page 279 in Serfling (1980). Let us briefly explain.

Using the same notation as in Serfling (1980), one uses the double sequence $t_{ni} = (i - 1/2)/n$, and uses $J(t) = \Phi^{-1}(t)$, with Φ denoting the $N(0,1)$ CDF as usual. Thus J is everywhere continuous and the growth condition of Example A* is satisfied with any $r > 1$, in particular $r = 2$. Note $r = 2$ is a legitimate choice because we have assumed a finite variance. Finally, one must also have $\max |t_{ni} - i/n| \to 0$, which is trivial, and the existence of a constant $a > 0$ such that the condition $a. \min\{i/n, 1 - i/n\} \leq t_{ni} \leq 1 - a. \min\{i/n, 1 - i/n\}$ holds. This is also valid in particular with $a = 1/2$. This establishes the assertion of Theorem 2.1.

Proof of Corollary 1: We can clearly assume that $\theta = 0$. Thus, the numerator of

(2.3) equals $2\int_{\frac{1}{2}}^{1} F^{-1}(x)\Phi^{-1}(x)dx$. Denote $F^{-1}(x)$ by $g(x)$ and $\Phi^{-1}(x)$ by $h(x)$. Thus the numerator is $E[g(X)h(X)]$ where X is Uniform $[\frac{1}{2}, 1]$. The first inequality in Corollary 1 follows on using $\mathrm{Cov}(f, g) \geq 0$ since f, g are both monotone increasing. The second inequality follows from Holder's inequality

$$E(fg) \geq (Ef^p)^{\frac{1}{p}}(Eg^q)^{\frac{1}{q}}$$

for $0 < p < 1$, with $\frac{1}{p} + \frac{1}{q} = 1$. The restriction $p < \frac{1}{2}$ is needed for finiteness of (Eg^q) in this context.

The inequality (*) is an obvious consequence of the first two inequalities.

Proof of Proposition 2.2. We will prove this result by construction. Indeed, take F to be the CDF of the t distribution with m degrees of freedom, $m > 2$. Thus F has a finite variance. We will ultimately let $m \to 2$. Note that F is well known to belong to the family $\mathcal{F}$ under consideration.

The denominator σ in the formula for $\rho(F)$ is known to be $\sqrt{\frac{m}{m-2}}$. The numerator

$$\begin{aligned}&\int_0^1 F^{-1}(x)\Phi^{-1}(x)dx\\ &= 2\int_{\frac{1}{2}}^1 F^{-1}(x)\Phi^{-1}(x)dx.\end{aligned}$$

Due to the facts that $F^{-1}(x) = O(\frac{1}{(1-x)^{\frac{1}{m}}})$ and $\Phi^{-1}(x) = O(\sqrt{\log\frac{1}{1-x}})$ as $x \to 1$, it follows that the above integral is $O(1)$ as $m \to \infty$ and hence $\rho(F) = O(\sqrt{m-2}) \to 0$ as $m \to 2$.

Proof of Theorem 2.3. There is no essential difference in the proofs of Theorem 2.1 and 2.3. The normalization in each term in the numerator and the denominator is still $1/n$. On this normalization, the first term in the denominator is again a Riemann sum; the second term in the denominator is a trimmed variance and converges with probability 1 to a trimmed variance, that is, the variance of the distribution obtained on truncating F at the αth and $(1-\alpha)$th percentiles respectively. The numerator is handled in essentially the same way as in Theorem 2.1. The function $J(t)$ is now piecewise continuous, having two discontinuities, since it is equal to zero outside of a compact interval. The double sequence t_{ni} and the constant 'a' are the same as in Theorem 2.1.

Proof of Theorem 3.1. First consider the case when $F =$ Normal $(0,1)$. The proof is somewhat involved and the main facts used in the proof and the main steps are outlined below. The following facts are used:

a) $n\bar{X}/\sqrt{n \log\log n}$ is a relatively compact sequence: this is a consequence of the usual law of the iterated logarithm for sums of iid random variables;

b) $\sqrt{n}(b_n - \mu)/\sqrt{2 \log\log n}$ is also a relatively compact sequence for appropriate μ, which is actually equal to 1 in the case $F =$ Normal $(0,1)$: this is available for example in Wellner (1977).

c) z_n is of the following order:

$$1 - z_n/\sqrt{2\log n} = \log\log n/(4\log n) + o(\frac{\log\log n}{\log n}). \tag{6.1}$$

This can be seen by using the fact that the tail area in the $N(0,1)$ distribution, $1 - \Phi(x)$, is of the order $\phi(x)/x$ as $x \to \infty$, where ϕ is the $N(0,1)$ density.

d) Also needed is the following fact:

The maximum order statistic $X_{(n)}$ for the $N(0,1)$ distribution satisfies the following:

For any $\delta > 0$ and for all large n, with probability 1,

$$a_n^* - \beta_n[\log\log\log n + \log(1+\delta)] \le X_{(n)} \le a_n^* + (1+\delta)\beta_n \log\log n \tag{6.2}$$

where $a_n^* = \sqrt{2\log n} - \frac{\log\log n + \log(4\pi)}{2\sqrt{2\log n}}$ and $\beta_n = 1/\sqrt{2\log n}$.

This can be seen as inequality (51) in page 228 in Galambos (1978).

Now the stated assertion on $e_n \frac{\sqrt{2\log n}}{\log\log n}$ is obtained by writing

$$\begin{aligned} &e_n\sqrt{2\log n}/\log\log n \\ &= X_{(n)}\sqrt{2\log n}/\log\log n - \bar{X}\sqrt{2\log n}/\log\log n - z_n(b_n - 1)\sqrt{2\log n}/\log\log n \\ &\quad - z_n\sqrt{2\log n}/\log\log n. \end{aligned} \tag{6.3}$$

In the second and the third term one now uses the relative compactness facts about $\bar{X}$ and b_n listed above, on carefully multiplying and dividing by the same normalizing constant so that relative compactness obtains. For example, the second term is the term that involves

$\bar{X}$. In this term one multiplies and divides by $n/\sqrt{n \log\log n}$ in order that the relative compactness fact listed above is usable. One does likewise for the term involving b_n.

In the first term one uses both inequalities listed as fact d) above; the upper bound is used to show that $\limsup e_n\sqrt{2\log n}/\log\log n$ is less than or equal to $1+\delta$ for any given $\delta > 0$, and hence is less than or equal to 1. The lower bound of fact d) is then used to show that the lim inf is greater than or equal to 0.

Finally, the order for z_n listed as fact c) is used in the fourth term of the above expression (6.3).

The proof for the case $F =$ Uniform $[0,1]$ follows the same type of argument and is the easiest of all the cases. The only thing one needs to be careful about is that in the relative compactness result for b_n listed as fact b) above, the constant μ is evaluated to be $1/(2\sqrt{\pi})$ for the $F =$ Uniform $[0,1]$ case. The compact support of the Uniform distribution makes the rest of the proof easy.

The proof for the case $F =$ Double Exponential also follows the same main steps. Now one has to use the following fact about Double Exponential distribution:

The maximum order statistic $X_{(n)}$ for the Double Exponential distribution satisfies $X_{(n)}/\log n \to 1$ with probability 1. This fact can be deduced either from Theorem 4.4.4 in page 230 of Galambos (1978) or alternatively can also be seen from Example 4.3.3 in page 224 of the same book. The rest of the argument now follows the normal case steps exactly.

The case when F is a t distribution is an exception due to the different phenomenon it demonstrates. We need to assume that (the degree of freedom) $m > 2$ in order that the stated t distribution has a finite variance. A finite variance is needed for the relative compactness argument we have been using.

The central fact responsible for this distinct phenomenon in the case of t distributions is the following fact about t distributions:

The maximum order statistic $X_{(n)}$ of a t distribution satisfies the following two facts:

Fact 1: $\limsup X_{(n)}/n^{\frac{1}{m}} = \infty$ with probability 1

Fact 2: $\liminf X_{(n)}/n^{\frac{1}{m}} = 0$ with probability 1.

Fact 1 is a consequence of Theorem 4.4.1 in page 225 of Galambos (1978); in this Theorem one shows very easily that (44) holds for every positive number k and therefore (45) of Theorem 4.4.1 in Galambos (1978) follows.

Fact 2 is a consequence of Theorem 4.3.1 in page 214 of Galambos. In this Theorem, take as the sequence u_n, $u_n = \gamma n^{\frac{1}{m}}$, where γ is any fixed positive real number. Then it is seen without difficulty that the sum in (21) of that theorem in Galambos (1978) diverges. Therefore it follows from this Theorem that $X_{(n)}/n^{\frac{1}{m}} \leq \gamma$ infinitely often with probability 1. But γ was taken as any arbitrary fixed positive number; the stated assertion in Fact 2 above therefore follows.

Once these two facts are established, the rest of the argument is again the same as the normal case argument. In fact all the main steps follow directly once Facts 1 and 2 above are observed.

Proof of Theorem 4.1. The numerator in the studentized residual equals $X_{(i)} - \bar{X} - b_n z_i$. We can assume without loss of generality that the mean of F is zero in this Theorem. From standard asymptotic theory, $X_{(i)}$ converges with probability 1 to $F^{-1}(\alpha)$ since $i = [n\alpha]$ and F was assumed to be continuous. $\bar{X}$ converges with probability 1 to zero, and b_n, the slope of the least squares line also has an almost sure limit by virtue of facts described in the proof of Theorem 2.1 and indeed this limiting value equals $\int_0^1 F^{-1}(x)\Phi^{-1}(x)dx = \mu$ (say).

The average sum of squared residuals is seen on standard algebra to be equal to $s^2 - b_n^2 \cdot \frac{1}{n}\sum_{i=1}^{n} z_i^2$, which has the almost sure limit $\sigma^2 - \mu^2$. Thus both the numerator and the denominator in the studentized residual are handled and from this the Theorem follows.

Acknowledgement. Peter Bickel mentioned to me during a visit to the MSRI that uniform data tend to look like normal on a $Q - Q$ plot. Mary Ellen Bock and David Moore made a number of helpful suggestions. I have had useful discussions with P. K. Sen on the possibility of central limit theorems and I am deeply grateful to Herman Rubin for constantly listening to me and for verifying a number of my results.

References

1. Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit, *JASA*, **49**, 765–769.

2. Filliben, J. J. (1975).The probability plot correlation coefficient test for normality, *Technometrics*, **17**, 111–117.

3. Galambos, J.(1978). The Asymptotic Theory of Extreme Order Statistics, John Wiley, New York.

4. Hettmansperger, T. P. (1987). Why not try a robust regression, *Austr. Jour. Stat.*, **29**, 1–18.

5. LaBrecque, J.(1977). Goodness of fit tests based on nonlinearity in probability plots, *Technometrics*, **19**, 293–306.

6. Moore, D. S. and McCabe, G. P. (1993).Introduction to the Practice of Statistics, Freeman, New York.

7. Rousseeuw, P. J. and Leroy, A. (1987). Robust Regression and Outlier Detection, John Wiley, New York.

8. Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model, *JASA*, **75**, 828–838.

9. Ruppert, D. and Carroll, R. J. (1985). Transformations in regressions: a robust analysis, *Technometrics*, **27**, 1–12.

10. Seber, G. A. F. (1984). Multivariate observations, John Wiley, New York.

11. Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics, John Wiley, New York.

12. Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611.

13. Wellner, J. A. (1977). A law of the iterated logarithm for functions of order statistics, *Ann. Statist.*, **5**, 481–494.