Statistica Sinica **14**(2004), 1209-1219

SIMULTANEOUS CONCENTRATION BANDS FOR CONTINUOUS RANDOM SAMPLES

Balgobin Nandram and Jai Won Choi

Worcester Polytechnic Institute and National Center for Health Statistics

Abstract: Statisticians routinely plot ordered observations against expected values to validate a model using a random sample. In fact, it is possible to construct a probability plot for a random sample from any continuous distribution function, and this is accommodated by the probability integral transform which facilitates a uniform probability plot of the ordered transformed observations against their expected values. Random variation in the plots is sometimes assessed using pointwise concentration bands. There are two problems with these plots. First, under a given distribution, certain points are much more variable than others. For example, when the distribution is normal, the points nearest the middle of the plot have the smallest variance. Second, the order statistics used in the construction of the plots are correlated. Both problems make the interpretation of the plot difficult. Pointwise concentration bands are, however, inadequate because there will be departures from the expected 45° straight line not only from sampling variation but also from the correlation introduced by ordering the observations. To account for this correlation, we construct simultaneous concentration bands which have exact coverage probability. A comparison is made with the pointwise and Bonferroni concentration bands. An empirical study shows that it is beneficial to construct our exact simultaneous concentration bands, and reasonable departures from the underlying distribution assumption can be detected.

Key words and phrases: Bisection method, correlation, order statistics, probability integral transform.

1. Introduction

It is standard statistical practice to plot ordered values against expected values for random samples (i.e., independent and identically distributed random variables) from continuous distributions to assess a hypothesized probability model. Random variation in the plots is sometimes assessed using pointwise concentration bands. However, ordering of the observations introduces a correlation that is ignored in the pointwise concentration bands. Our task is to construct simultaneous concentration bands by incorporating the correlation. This is accomplished by constructing intervals (with desirable properties) that make up the bands. Our research was motivated when we attempted to fit models to body mass index and bone mineral density in the National Health and Nutrition

Examination Survey, and to mortality for chronic obstructive pulmonary disease in the National Vital Statistics Program at the National Center for Health Statistics.

Let $X_1, \ldots, X_n$ be a random sample of size n from any continuous distribution function and let $X_{(i)}$ represent the ith ordered observation (i.e., $X_{(1)} \leq X_{(2)} \leq \cdots X_{(i)} \leq \cdots \leq X_{(n)}$). Suppose $E(X_{(i)})$ exists. Then we call a plot of the ith ordered observation versus its expected value $E(X_{(i)})$ a *probability plot*. If the X_i are normally distributed, the plot is called a *normal probability plot*, and if the X_i are uniformly distributed, the plot is called a *uniform probability plot*. If the observations follow the prescribed distribution model, the plot is expected to be a 45° straight line. Then the $100(1-\alpha)\%$ pointwise concentration bands, $0 < \alpha < 1$ (e.g., $\alpha = 0.05$ for 95%), are pointwise $100(1-\alpha)\%$ confidence intervals at each observation with the lower and upper end points of the intervals joined to form an upper and a lower band, respectively. Pointwise concentration bands can explain fluctuations arising from random variations.

Let $F(x)$ denote the cumulative distribution function (cdf) of a continuous random variable X, and let $X_1, \ldots, X_n$ be a random sample from $F(\cdot)$. Letting $U_i = F(X_i)$, then U_i, $i = 1, \ldots, n$, are a uniform random sample from $[0, 1]$. Letting $U_{(i)}$ be the ith order statistic, a plot of $E(U_{(i)})$ versus the ordered observed values is a uniform probability plot. It is noteworthy that $E(U_{(i)})$ always exists. Thus, the probability integral transform permits construction of concentration bands for any continuous random sample.

There are two problems with these plots. First, under the given distribution, certain points are much more variable than others. For example, when F is normal, the points nearest the middle of the plot have the smallest variance, and when F is uniform, the points nearest the middle of the plot have the largest variance. Second, the value of the order statistics used in the construction of the plots are correlated. Both problems make the interpretation of the plot difficult. Throughout, we make the standard assumption that the original observations are independent and identically distributed. Using a stabilized probability plot via the arc sine transformation of the $U_{(i)}$, Michael (1983) presented an elegant solution to the first problem. The second problem is largely unresolved and we address it in this paper.

Barnett (1975) discussed the choice of plotting positions for continuous random samples, but he did not study pointwise concentration bands. Quesenberry and Hales (1980) did study them for uniform probability plots. For the normal probability plot, Dempster and Ryan (1985) and Lange and Ryan (1989) used weighted normal plots for random effects models with pointwise bands. Others have considered tests for normality (e.g., Shapiro and Wilk (1965), Filliben (1975)). More recently Brown and Hettmansperger (1996) discussed normal scores, normal plots and tests for normality, but they did not incorporate bands in their probability plots. Others have discussed tests for uniformity (e.g., Quesenberry and Miller (1977), Dudewicz and Meulen (1981)). But these works do not address the issue of correlation introduced by the ordering of the observations. Because of this correlation, pointwise bands are too narrow and, therefore, conservative.

It is of interest to consider simultaneous concentration bands for probability plots. Then, the exact $100(1-\alpha)\%$ concentration bands contain all the observations with probability $1-\alpha$. By using the Bonferroni inequality, a simple adjustment to the pointwise bands can produce simultaneous bands, but Bonferroni bands can be too wide and hence not so useful. Thus, a method to construct simultaneous concentration bands is desirable for model validation.

In an attempt to improve the half-normal probability plot, Zahn (1975 a,b) constructed individual α-level critical values for each of the four largest contrasts (absolute values) in a 2^4 factorial experiment under the null hypothesis that the contrast means are all zero in his Problem G. These critical values are obtained by controlling the probability of a non-zero family (simultaneous) error rate, and they are smaller than the ones originally proposed by Daniel (1959). Then he joined these critical values by a line to yield a guardrail. Our method can be used to construct 95% simultaneous confidence bands for the absolute values for all the contrasts under Problem G. The issue here is just to construct the confidence bands for a random sample drawn from a half-normal probability density function, and this falls under our framework, at least when the error variance is known. (Our method can be extended to cover the case in which the error variance is unknown.) Thus, our method goes beyond that of Zahn (1975 a,b).

In Section 2 we present the theory and method. In Section 3 we use a simulation study to discuss departures from an assumed underlying distribution. There are concluding remarks in Section 4.

2. Theory and Method for Concentration Bands

An important observation (see David (1981) for further details) helps to reduce the computation in our method. Let $U_1, U_2, \ldots, U_n, U_{n+1}, \ldots, U_{2n}$ be a random sample of size $2n$ from a uniform distribution on $[0,1]$, and let $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)} \leq U_{(n+1)} \leq \cdots \leq U_{(2n)}$ be the corresponding order statistics. (For an odd number sample size the results are similar). Then $(U_{(1)}, \ldots, U_{(n)}) \stackrel{d}{=} (1-U_{(2n)}, \ldots, 1-U_{(n+1)})$, where $\stackrel{d}{=}$ means equal in distribution. We call this the *reflection property* for order statistics from a uniform sample and refer to $U_{(n)}$ and $U_{(n+1)}$ as reflection points. It follows that $Pr(a \leq U_{(k)} \leq b) = Pr(1-b \leq$

$U_{(2n-k+1)} \leq 1-a)$, $0 \leq a < b \leq 1$. This is particularly useful because it states that the $100(1-\alpha)$ % point of $U_{(k)}$, which we denote by C_α, is the $100\alpha\%$ point, which is $1-C_\alpha$, of $U_{(2n-k+1)}$.

We introduce the $100(1-\alpha)\%$ pointwise and Bonferroni concentration bands in Section 2.1 and, in Section 2.2, we show how to construct the exact $100(1-\alpha)\%$ simultaneous concentration bands. In Section 2.3 we discuss three special cases.

2.1. Pointwise concentration bands

Let $F_k(x)$ denote the cumulative distribution function (cdf) of the kth order statistic in $[0,1]$ and $f_k(x)$ the corresponding probability density function (pdf). Then the $100(1-\alpha)\%$ highest probability density (HPD) interval (shortest for its probability content) for $U_{(k)}$ is obtained by solving simultaneously the two equations,

$$F_k(b) - F_k(a) = 1-\alpha \text{ and } f_k(b) = f_k(a) \tag{1}$$

(e.g., see Press (1989, pp.30-32)). We have a preference for equal ordinate intervals over equal tail intervals because equal ordinate intervals are HPD intervals for unimodal probability densities.

The equations (1) can be easily solved. For, letting $F_k^{-1}(x)$ be the inverse cdf, we only need to solve

$$f_k(a) - f_k\{F_k^{-1}(F_k(a)+1-\alpha)\} = 0 \text{ such that } 0 \leq a \leq F_k^{-1}(\alpha). \tag{2}$$

The simple bisection method cannot fail in (2). Note that $k=1$ and $k=2n$ are not special cases of (2), because for $k=1$ the HPD interval is $(0, 1-\alpha^{1/2n})$ and, by the reflection property, for $k=2n$ the HPD interval is $(\alpha^{1/2n}, 1)$.

The $100(1-\alpha)\%$ pointwise concentration bands for a random sample is the region $\bigcap_{k=1}^{2n} A_k$ where

$$A_k = \begin{cases} (0, b_1), & k=1 \\ (a_k, b_k), & k=2,\ldots,n, \end{cases}$$

and, by the reflection property,

$$A_k = \begin{cases} (1-b_{2n-k+1}, 1-a_{2n-k+1}), & k=n+1,\ldots,2n-1 \\ (1-b_1, 1), & k=2n, \end{cases}$$

where $Pr(U_{(k)} \in A_k) = 1-\alpha$, and each A_k, $k=2,\ldots,n$, is obtained by (2). Thus, we only need to construct the intervals for the n smallest or largest random variables. We call these pointwise concentration bands QH concentration bands.

The Bonferroni bands, denoted by BO, are simply obtained by setting $Pr(U_{(k)} \in A_k) = 1-\alpha/2n$, $k=2,\ldots,n$. In this case $Pr(\bigcap_{k=1}^{2n} U_{(k)} \in A_k) \geq 1-\alpha$

but, for the pointwise QH bands, $Pr(\bigcap_{k=1}^{2n} U_{(k)} \in A_k) < 1-\alpha$. Thus while the QH bands are too narrow, the BO bands can be too wide. There is evidence that inference based on the Bonferroni Inequality can be very accurate. For example, see Miller (1981, p.254) and Cook and Prescott (1981).

2.2. Exact simultaneous concentration bands

Motivated by the BO concentration bands, we find (a_k, b_k), $k = 1, \ldots, n$ such that

$$F_{k-1}(b_{k-1}) - F_{k-1}(a_{k-1}) = F_k(b_k) - F_k(a_k), \tag{3}$$

$$f_k(a_k) = f_k(b_k), \tag{4}$$

$k = 2, \ldots, n$, and

$$Pr\Big\{ \bigcap_{k=1}^{n} a_k < U_{(k)} < b_k, \bigcap_{k=n+1}^{2n} 1 - b_{2n-k+1} < U_{(k)} < 1 - a_{2n-k+1} \Big\} = 1 - \alpha. \tag{5}$$

Note that the concentration bands consist of pointwise intervals (a) with highest probability density by (4) and (b) the $2n$ intervals have the same probability content by (3). The second term in the probability at (5) follows from the reflection property. Finally note that for the HPD intervals $a_1 = 0$, and that (3) and (4) together imply that a_k and b_k are strictly increasing in k.

It follows from (3) that $b_k = F_k^{-1}\{F_1(b_1) + F_k(a_k)\}$ and $0 \le F_k(a_k) \le 1 - F_1(b_1)$, $k = 2, \ldots, n$. Also, by construction, $F_1(b_1) \ge 1 - \alpha$. Thus, one has

$$F_1^{-1}(1-\alpha) \le b_1 \le F_1^{-1}\left\{1 - \max_{k=2,\ldots,n} F_k(a_k)\right\}, \tag{6}$$

$$0 \le a_k \le F_k^{-1}(\alpha). \tag{7}$$

Note that $\max_{k=2,\ldots,n} F_k(a_k)$ is not necessarily $F_n(a_n)$ (but for equal tail intervals $F_k(a_k) = \alpha/2$, $k = 2, \ldots, n$).

Then the problem is to solve the equations

$$\begin{aligned} &Pr\{0 < U_{(1)} < b_1, \bigcap_{k=2}^{n} a_k < U_{(k)} < F_k^{-1}\{F_1(b_1) + F_k(a_k)\}, \\ &\bigcap_{k=n+1}^{2n-1} 1 - F_{2n-k+1}^{-1}\{F_1(b_1) + F_{2n-k+1}(a_{2n-k+1})\} \le U_{(k)} \le 1 - a_{2n-k-1}, \\ &\qquad 1 - b_1 \le U_{2n} \le 1\} = 1 - \alpha, \end{aligned} \tag{8}$$

$$f_k(a_k) = f_k\{F_k^{-1}(F_1(b_1) + F_k(a_k))\}, \tag{9}$$

subject to (6) and (7).

We show that there is a unique solution $\{b_1,\ldots,a_n\}$ that solves (8) and (9) subject to (6) and (7). First note that if $b_1 = F_1^{-1}(1-\alpha)$, then the probability content of the simultaneous concentration bands is less than $1-\alpha$ since all the intervals have the same probability content. Also, for every fixed $\{a_2,\ldots,a_n\}$, the probability content increases to 1 as b_1 increases to 1. Thus, for every fixed $\{a_2,\ldots,a_k\}$, there is a unique solution for b_1 in (8). Then, since the probability density functions of the $U_{(k)}$ are all unimodal, there is a unique solution for $\{a_2,\ldots,a_n\}$ for each b_1 in (9).

Henceforth, we denote the exact simultaneous concentration bands by NC. By construction, because all intervals are HPD with equal probability content, the widths of the concentration bands in increasing order are QH, NC and BO, with QH bands within NC bands, and NC bands within BO bands. (See Figure 1.)

Note that there are n unknown quantities in (8) and (9). In practice to solve for them, we start with b_1 obtained from the BO bands and, using the bisection method, we solve the $n-1$ equations in (9) for a_k as in (2). Then, again using the bisection method, we solve (8) for b_1 after substituting the values obtained for $\{a_2,\ldots,a_n\}$.

It is easy to obtain the probability in (8) at each step of the bisection method. We obtain a sample of M values of the vector $U_{(1)} \le U_{(k)} \le \cdots \le U_{(n)} \le \cdots \le U_{(2n)}$. Each of these is simply the order statistics from a random sample of $2n$ values from a uniform distribution in $(0,\ 1)$. We simply drew $V_1,\ldots,V_{2n+1} \overset{iid}{\sim} \exp(1)$ and took $U_{(i)} = \sum_{j=1}^{i} V_j / \sum_{j=1}^{2n+1} V_j$, $i=1,\ldots,2n$.

Then, we count the proportion $\hat{p}$ of the M iterates with $0 \le U_{(1)} \le b_1, \bigcap_{k=1}^{n} a_k \le U_{(k)} \le F_k^{-1}\{F_1(b_1)+F_k(a_k)\}, \bigcap_{k=n+1}^{2n-1} 1 - F_{2n-k+1}^{-1}\{F_1(b_1)+F_k(a_k)\} \le U_{(k)} \le 1 - a_{2n-k+1}, 1-b_1 \le U_{2n} \le 1$. We take $M = 10,000$ for good accuracy since, by the Central Limit Theorem with this Monte Carlo sample size, $\hat{p}$ is within 0.0065 of $p = 0.95$ with probability approximately 0.997.

For the 95% concentration bands we constructed, the entire procedure converges very rapidly.

2.3. Three special cases

We describe three special cases, two with closed form answers and the third without; these correspond to $n = 1,2,3$. For $n \ge 3$, closed form answers are difficult to obtain.

For the case of one observation we take the $100(1-\alpha)\%$ QH, NC, and BO intervals to be the same $(\alpha/2, 1-\alpha/2)$. Note that any interval of length $1-\alpha$ is a HPD interval.

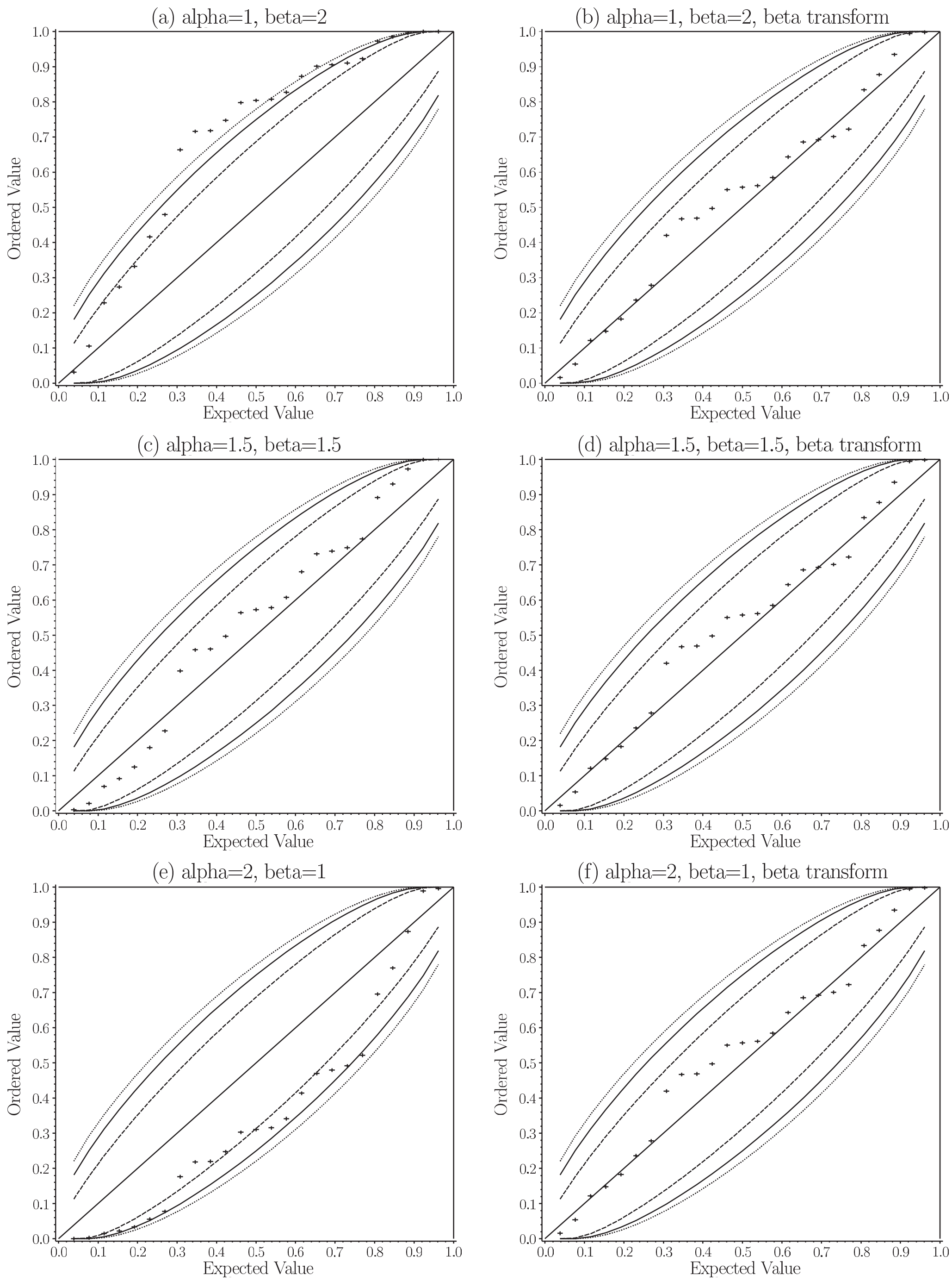

Figure 1. Examples of mild departures from uniformity using beta samples: BO bands (dotted); NC bands and 45° straight line (solid); QH bands (dashed).

For the case of two observations we need b_1 such that $1-\sqrt{\alpha}\le b_1\le 1$ and $Pr\{0\le U_{(1)}\le b_1,\ 1-b_1\le U_{(2)}\le 1\}=1-\alpha$. Now,

$$\begin{aligned}Pr\{0\le U_{(1)}\le b_1,\ 1-b_1\le U_{(2)}\le 1\} &= 2\int_{1-b_1}^{1}\Big\{\int_0^{\min(b_1,u_{(2)})} du_{(1)}\Big\}du_{(2)}\\ &= -2b_1^2+4b_1-1.\end{aligned}$$

It follows that, setting $-2b_1^2+4b_1-1=1-\alpha$, the NC bands are formed from the pointwise intervals $(0,1-\sqrt{\alpha/2})$ and $(\sqrt{\alpha/2},1)$. For the QH bands the intervals are $(0,1-\sqrt{\alpha})$ and $(\sqrt{\alpha},1)$. The BO bands coincide with the NC bands for two observations.

For the case of three observations, analytic solutions are not available, but there is some simplification. We need b_1 and a_2 such that $F_1^{-1}(1-\alpha)\le b_1\le 1,\ 0\le a_2\le F_2^{-1}(\alpha),\ Pr\{0\le U_{(1)}\le b_1,\ a_2\le U_2\le 1-a_2,\ 1-b_1\le U_{(3)}\le 1\}=1-\alpha$ and $F_1(b_1)=F_2(1-a_2)-F_2(a_2)$. From the last, $a_2^3-(3/2)a_2^2+(1/4)(1-b_1)^3=0$ and, letting $\cos\phi=1-(1-b_1)^3$, the solution, denoted by $Q(b_1)$, is the one of $(1/2)+\cos(\phi/2),(1/2)+\cos\{(\phi+2\pi)/3\}$ or $(1/2)+\cos\{(\phi+4\pi)/3\}$ that lies in $(0,F_2^{-1}(\alpha))$. (See CRC tables (1964) for solutions of a cubic equation.) After extensive algebraic manipulation,

$$\begin{aligned}&Pr\{0\le U_{(1)}\le b_1,\ a_2\le U_{(2)}\le 1-a_2,\ 1-b_1\le U_{(3)}\le 1\}\\ &\quad=6\int_{1-b_1}^{1}\Big\{\int_{a_2}^{\min(u_{(3)},1-a_2)}\Big\{\int_0^{\min(b_1,u_{(2)})} du_{(1)}\Big\}du_{(2)}\Big\}du_{(3)}\\ &\quad=3[(b_1-a_2)\{2b_1(1-b_1)-a_2(b_1+a_2)\}+a_2\{2b_1-(b_1+a_2)^2\}].\end{aligned}$$

Thus, we need b_1 such that $(1-Q(b_1))\{2b_1(1-b_1)-Q(b_1)(b_1+Q(b_1))\}+Q(b_1)\{2b_1-(b_1+Q(b_1))^2\}=(1-\alpha)/3$. The BH bands have intervals $(0,1-\sqrt[3]{\alpha})$, $(a_p,1-a_p)$, $(\sqrt[3]{\alpha},1)$ and BO bands have intervals $(0,\sqrt[3]{\alpha/3})$, $(a_B,1-a_B)$, $(\sqrt[3]{\alpha/3},1)$ where, setting

$$\cos\phi=\begin{cases}(1-\alpha)/2, & \text{QH bands}\\ (1-\alpha/3)/2, & \text{BO bands,}\end{cases}$$

a_p and a_B are the values in $(0,F_2^{-1}(\alpha))$ and $(0,F_2^{-1}(\alpha/3))$, respectively, which are one of $(1/2)+\cos(\phi/3)$, $(1/2)+\cos\{(\phi+2\pi)/3\}$ or $(1/2)+\cos\{(\phi+4\pi)/3\}$.

For more than four observations analytical results are difficult to obtain for all three methods, but the QH and BO methods are simpler to compute than the NC method. Still the computational time is small.

3. Empirical Analysis

In this section we characterize departures from an assumed underlying model for a random sample from uniform, beta and Student's t distributions using several simulation runs. In our examples we present the comparisons graphically (in Figure 1, the bands are ordered QH, NC, and BO from the 45° straight line). Note that the bands curve in towards the ends corresponding to the smallest and largest uniform order statistics. In contrast, for a normal probability plot these bands curve away from each other towards the ends corresponding to the smallest and largest normal order statistics in a manner similar to prediction bands. For example, see Dempster and Ryan (1985) and Lange and Ryan (1989) for pointwise bands in weighted normal plots for random effects models.

We assess departures from uniformity by selecting a random sample from the uniform distribution function and then drawing the uniform probability plot. Our six figures (not presented here) show different departures from uniformity at different sample sizes. In the first figure ($n = 10$) the points are all near the lower QH band indicating a departure from uniformity for this sample, but the points are all within the NC bands. In the second figure ($n = 10$) all the points are near and above the upper QH band but again within the NC band. In third figure ($n = 25$) the points tend to cluster about a straight line with slope smaller than 45°. In the fourth figure ($n = 50$) most of the points fall on the lower QH band, but still within the NC band. In the fifth figure ($n = 100$), except in the tails, the points fall approximately on a straight line parallel to the expected 45° straight line and still within the NC bands. Finally, in the sixth figure ($n = 100$) there are at least two runs across the 45° straight line with some points falling outside the QH band but on the NC band. For larger sample size, more points fall near the 45^o straight line.

Next we consider mild and severe departures from uniformity using random samples from beta cdf's. We also assess a random sample from the Student's t cdf using the uniform probability plot.

For mild departures from uniformity we generate samples of sizes 5, 10, 25, 50, 100, 200, 300 from beta(1, 2), beta(1.5, 1.5) and beta(2, 1). For severe departures from uniformity we draw samples from beta(5, 15), beta(10, 10) and beta(15, 5). Then we use the corresponding beta cdf to transform the sample to uniformity.

In Figure 1 we present the mild departures for samples of size 25. The symmetric beta(1.5, 1.5) is the closest to uniformity. Except for a slightly better fit for the transformed case (Figure 1d) over the untransformed case (Figure 1c), there is very little difference between them. However, beta(1, 2) and beta(2, 1) show symmetric departures from uniformity (Figure 1a somewhat symmetric to Figure 1e), the points being plotted near and above the upper bands for beta(1, 2). These problems are resolved (Figures 1b and 1f) by the beta transform.

We consider severe departures for samples of size 50; there are very marked departures even in the symmetric case. The strong unimodality and thin tails in the symmetric case create severe departures from uniformity. The strong unimodality with right skewness and thin left tail in beta(5, 15) and left skewness with thin right tail in beta(15, 5) create severe departures from uniformity. All the departures are resolved by the beta transform, and note that the patterns in Figures 1 persist for all the sample sizes we studied.

We also investigated probability plots for the Student's t family of distributions hoping to discriminate between the Student's t probability density function and the normal. Because it is difficult to discriminate between the normal and the Student's t for small sample sizes, we chose sample sizes of 100, 200 and 300 and, in each case, we ran 5, 10 and 20 degrees of freedom. We generated random samples from the Student's t distribution, and transformed the data using the probability integral transform based on both the Student's t distribution and the standard normal distribution. For 300 observations and 5 degrees of freedom, when the normal transform is used there is poor performance especially in the tails and for ordered values smaller than 0.5, but these are corrected with the appropriate t transform. In general, there are more differences between these plots for small degrees of freedom. For moderate degrees of freedom (e.g., 10) there are very minor differences and for larger degrees of freedom (e.g., 20 and beyond) there are virtually no differences between these plots.

4. Concluding Remarks

We have developed a graphical method to assess the distribution assumption of a continuous random sample. Simultaneous inference is used to address the main issue of correlation among the ordered observations.

The computation consists of two parts. One part requires several root finders and is facilitated by the bisection method. The other part requires a probability calculation facilitated by a Monte Carlo method. A reflection property of the order statistics reduces the number of equations to be solved by half. Computation time is not significant.

While pointwise concentration bands are too narrow and Bonferroni concentration bands can be too wide, our method provides exact $100(1-\alpha)\%$ concentration bands falling between these two. In addition, each confidence interval used in the concentration bands has the highest probability density (shortest interval for its coverage probability) for any continuous distribution.

Acknowledgement

This work started while Balgobin Nandram was the first ASA/NCHS research fellow at the National Center for Health Statistics, January 1999 – June 2000. We are grateful to an associate editor and to two referees for their informative comments.

References

Barnett, V. (1975). Probability plotting methods and order statistics. *Appl. Statist.* **24**, 95-108.

Brown, B. M. and Hettmansperger, T. P. (1996). Normal scores, normal plots, and tests of normality. *J. Amer. Statist. Assoc.* **91**, 1668-1675.

Weast, R. C., Selby, S. M. and Hodgman, C. D. (1964). *CRC Standard Mathematical Tables*, 12th edition (Edited by R. C. Weast, S. M. Selby and C. D. Hodgman), Chemical Rubber Company, Cleveland, Ohio.

Cook, R. D. and Prescott, P. (1981). On the accuracy of the Bonferroni significance levels for detecting outliers in linear models. *Technometrics* **23**, 59-63.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1**, 311-341.

David, H. A. (1981). *Order Statistics*. Wiley, New York.

Dempster, A. and Ryan, L. (1985). Weighted normal probability plots. *J. Amer. Statist. Assoc.* **80**, 845-850.

Dudewicz, E. J. and Meulen, E. C. V. (1981). Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.* **76**, 967-974.

Filliben, J. J. (1975). The probability plot correlation coefficient test of normality. *Technometrics* **17**, 111-117.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *Ann. Statist.* **17**, 624-642.

Michael, J. R. (1983). The stabilization probability plot. *Biometrika* **79**, 11-17.

Miller, R. G. (1981). *Simultaneous Statistical Inference*. 2nd edition. Springer-Verlag, New York.

Press, S. J. (1989). *Bayesian Statistics: Principles, Models and Applications*. Wiley, New York.

Quesenberry, C. P. and Hales, C. (1980). Concentration bands for uniformity plots. *J. Statist. Comput. Simulation* **11**, 41-53.

Quesenberry, C. P. and Miller, F. L. (1977). Power studies of some tests for uniformity. *J. Statist. Comput. Simulation* **5**, 169-191.

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

Zahn, D. A. (1975a). Modifications of and revised critical values for the half-normal plot. *Technometrics* **17**, 189-200.

Zahn, D. A. (1975b). An empirical study of the half-normal plot. *Technometrics* **17**, 201-211.

Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, U.S.A.

E-mail: balnan@wpi.edu

National Center for Health Statistics, Room 915, 6525 Belcrest Road, Hyattsville, MD 20782, U.S.A.

E-mail: jwc7@cdc.gov

(Received January 2003; accepted September 2003)