TECHNICAL REPORT NO. 443

February 1976

FRACTIONAL ORDER STATISTICS, WITH APPLICATIONS

by

Stephen M. Stigler
University of Wisconsin, Madison

Fractional Order Statistics, with Applications

by

Stephen M. Stigler*
University of Wisconsin, Madison

1. Introduction.

The last numbered pages of the four principal volumes of Laplace's Mécanique céleste [1] are, in increasing order, 303, 347, 368, and 382. What is the median number of pages in the volumes of the Mécanique céleste? Simple questions of definition such as this arise in the most elementary statistics courses, and seasoned instructors have developed many ways of handling them, from averaging to get (347 + 368)/2 = 357.5 as a "median", to listing both 347 and 368 as "comedians". In elementary courses this definitional problem is usually of secondary importance (questions such as "what is the population being sampled?" being emphasized), but in more advanced courses and in statistical research, this same simple problem--caused by an inadequate supply of order statistics -- can lead to nagging difficulties.

For example, in studies of robust estimators, intuitively reasonable definitions (such as of "the 10th percentile") that make perfectly good sense in infinite populations may require redefinition for small samples. How does one compute the average of the quartiles of a sample of size 9? What is the 10% trimmed mean of a sample of size 13? In many cases statisticians have invented useful algorithms for bridging the gaps between order statistics: Tukey's hinges can be taken as quartiles for any sample size, and the programs contained in the Princeton Study can be taken as defining all manner of trimmed means [1]. For data analysis, one could scarcely hope for much improvement on such algorithms (although this is a largely neglected research topic), but for theoretical investigations they leave much to be desired.

In the first place, the exact distribution theory of the estimators calculated from these algorithms is usually so complicated as to preclude any but Monte Carlo studies. Even where an exact analytical treatment is feasible, as with the median, strange anomalies may arise that reflect the discreteness of the sample size more than properties of the estimator. For example, Hodges and Lehmann [7] might suggest that when computing the median of an odd sample size one may as well discard an observation at random!

As a second case in point, the large sample theory of (say) a linear function of order statistics can become an annoyingly difficult problem with even the simplest definitional algorithm, due to analytic intractability and the consequent need to employ a variety of techniques of approximation. We shall enlarge upon this point in section 3.

The aim of this paper is to introduce a solution to this dilemma, and discuss a number of applications. The solution we propose could be called "fractional" or "imaginary" order statistics, or even an "order statistics process". From another point of view, this paper may be viewed as suggesting some novel, non-Bayesian applications for a family of probability measures introduced by Ferguson [5, 6] as "Dirichlet Processes" for Bayesian analyses of non-parametric problems. The idea, briefly put, is to consider not just a finite collection, but a continuum of order statistics, notwithstanding a finite "sample size". By this technical device, many of the problems discussed above disappear, and some remarkably simple proofs of known propositions become possible.

It should be emphasized that the "order statistic process" we discuss is a purely technical creation, for use in theoretical investigations and not in data analysis. If one wishes to describe the median number of pages in the

Mécanique céleste, then one had best seek out the fifth, largely supplementary volume (419 pages). But for theoretical studies we shall introduce the 2.5th order statistic from a sample of size 4, not to mention the eth and the πth, by specifying their joint probability distribution.

In the following section we shall review a few needed properties of Ferguson's Dirichlet process, and define the order statistic process. In section 3, the use of this process in large sample theory will be illustrated, in section 4 a possible "small-sample" asymptotic approach is proposed, and in section 5 an application to Yule processes is discussed.

2. Fractional Order Statistics - the Order Statistic Process.

The Dirichlet Process was introduced by Ferguson [5] as a prior distribution for the Bayesian analysis of nonparametric problems. We propose a different use. Basically, a Dirichlet Process is a probability measure on the set of probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ which enjoys the property that the joint distribution of $(P(B_1), \ldots, P(B_k))$ is an ordinary Dirichlet distribution on the k dimensional unit square (see Wilks [17, p. 177]), for any k and any $B_1, B_2, \ldots, B_k, B_{k+1}$ which form a measurable partition of $\mathcal{X}$.

We shall not require the full generality of Ferguson's definition, but will specialize immediately to the case where $\mathcal{X} = [0, 1]$ is the unit interval, with $\mathcal{A}$ = the Borel sets. Also, we shall only consider the random distribution function $F(t) = P([0, t])$ corresponding to P. For this special case the full definition of the Dirichlet process is given by:

Definition 1: We say a random distribution funtion $F(t)$ on $[0, 1]$ is a Dirichlet process indexed by a measure ν on the Borel sets of $[0, 1]$, if, for every $k > 0$ and $0 = t_0 < t_1 < t_2 < \ldots < t_k < 1$, $(F(t_1), F(t_2) - F(t_1), \ldots, F(t_k)-F(t_{k-1}))$ has the Dirichlet distribution with density

$$\propto \left(\prod_{i=1}^{k} x_i^{\nu(t_i) - \nu(t_{i-1}) - 1} \right) \left(1 - \sum_{i=1}^{k} x_i \right)^{\nu(1) - \nu(t_k) - 1},$$

where $\nu(t) = \nu([0, t])$ for any $t \in [0, 1]$.

Ferguson's work [5] proves the existance of such a probability distribution on the distribution functions on $[0, 1]$, for any non-null finite measure ν. In what follows it will be convenient to let $\lambda = \nu(1) = \nu([0, 1])$, and $\alpha(t) = \nu(t)/\lambda$. We shall always suppose $0 < \lambda < \infty$. Then the following properties of $F(t)$ are well-known and easily derived:

(1) For any t, $F(t)$ has a Beta distribution, $\beta(\nu(t), \lambda - \nu(t))$.

(2) $E(F(t)) = \alpha(t)$, $V(F(t)) = \alpha(t)(1-\alpha(t))(\lambda+1)^{-1}$.

(3) For any $0 \leq s \leq t \leq 1$, $\mathrm{cov}(F(s), F(t)) = \alpha(s)(1-\alpha(t))(\lambda+1)^{-1}$.

Our present interest in the Dirichlet process is motivated by the fact that if ν is proportional to Lebesgue measure, that is, if $\alpha(t) = t$ all $t \in [0,1]$, and if $\lambda = n+1$ (n a positive integer), then the vector

$$\left(F\left(\frac{1}{n+1}\right), F\left(\frac{2}{n+1}\right), \ldots, F\left(\frac{n}{n+1}\right)\right)$$

has exactly the same joint distribution as the order statistics $U^{(1)} \leq \ldots \leq U^{(n)}$ of a random sample of size n from a uniform $[0,1]$ distribution. This fact follows immediately from Definition 1 and Wilks [17, pp. 182, 236]. Further, if G is any (fixed) distribution function on $(-\infty, \infty)$, and $G^{-1}(u) = \sup\{x: G(x) \leq u\}$ is the right-continuous inverse function, then under the above conditions (ie. $\alpha(t) = t$, $\lambda = n+1$), the vector

$$\left(G^{-1}\left(F\left(\frac{1}{n+1}\right)\right), \ldots, G^{-1}\left(F\left(\frac{n}{n+1}\right)\right)\right)$$

has exactly the same joint distribution as do the order statistics $X^{(1)} \leq \ldots \leq X^{(n)}$ of a sample of size n from the distribution G. However, unlike the order statistics, $F(t)$ and $G^{-1}(F(t))$ have well-defined joint distributions for all t. This motivates

<u>Definition 2</u>: If $F(t)$ is a Dirichlet process on $[0,1]$ indexed by $\nu(t) = \lambda t$, we shall call $F(t)$ a <u>uniform order statistic process</u>. Similarly, we shall call $G^{-1}(F(t))$ the <u>order statistic process for the distribution G.</u>

Note that λ need not be an integer, although it was motivated by the case $\lambda = n+1$. Regardless of the value of $\lambda > 0$, we can heuristically refer to $G^{-1}(F(t))$ as the $t\lambda$ th order statistic of a sample of size $\lambda - 1$ from G.

While it will not really concern us here, we may recall that Ferguson [5] has shown that with probability one $F(t)$ corresponds to a discrete distribution, thus realizations of $F(t)$ and $G^{-1}(F(t))$ will be step functions. However, as defined here they will be right-continuous and monotone, and their <u>distributions</u> will vary continuously with t.

We now proceed to the discussion of some applications.

3. Applications to large sample theory.

One of several fruitful approaches to the asymptotic behavior of linear functions of order statistics has been what might be called the stochastic process approach. This approach was pioneered by Bickel [2] (who also refers to unpublished work of Hajek) and later developed by Shorack [11, 12, 13] to yield very strong results. Essentially, what is done (in the simplest formulation) is to define a function $Y_n(t) = X^{(i)}$ for $t = i/(n+1)$, and by linear interpolation for other $t \in [1/(n+1), n/(n+1)]$. Then if $H_n(t)$ is a function of bounded variation corresponding to a measure which puts mass c_{in} at $t = i/(n+1)$, a linear function of order statistics $S_n = \Sigma c_{in} X^{(i)}$ can be represented as an integral $S_n = \int_0^1 Y_n(t)\, d\, H_n(t)$. One might then show that $H_n(t) \to$ some limit $H(t)$ in a suitable sense, and that $n^{\frac{1}{2}}(Y_n(t) - G^{-1}(t))$ converges in distribution to a Gaussian process $W(t)$ on $[0, 1]$, and conclude that $n^{\frac{1}{2}}(S_n - \int_0^1 G^{-1}(t)\, d\, H(t))$ converges in distribution to $\int_0^1 W(t) d\, H(t)$, which is normally distributed. The function $Y_n(t)$ is called the quantile function, and one fact contributing to the mathematical complexity of this approach is that, as the distribution of $Y_n(t)$ is somewhat intractable for $t \neq i/(n+1)$, some i, the proof that $n^{\frac{1}{2}}(Y_n(t)-G^{-1}(t))$ converges to $W(t)$ becomes more difficult than might otherwise be the case.

If $\lambda = n + 1$, the order statistic process $G^{-1}(F(t))$ has a distribution which agrees with that of $Y_n(t)$ for t's $= i/(n+1)$, but its distribution is nicer for other t's, and its use permits a simplification in this approach to proving the asymptotic normality of S_n, under moderate regularity conditions[2].

We begin with a theorem concerning the convergence in distribution of Dirichlet processes, which while more general than needed in the present section, may be useful in other applications (such as censored data).

Theorem 1: Let $F(t)$ be a Dirichlet process on $[0, 1]$ indexed by ν, as described by Definition 1. If $\alpha(t) = \nu(t)\lambda^{-1}$ is continuous on $[0, 1]$, $\lambda = \nu(1)$, then

$Z_\lambda(t) = (\lambda+1)^{\frac{1}{2}}(F(t) - \alpha(t))$ converges in distribution to the Gaussian process $Z(t)$ on $[0, 1]$ with $EZ(t) \equiv 0$, $\text{cov}(Z(s), Z(t)) = \alpha(s)(1-\alpha(t))$ for $s \leq t$, as $\lambda \to \infty$, $\alpha(t)$ fixed.

By convergence in distribution, we mean that the probability distributions of $Z_\lambda(t)$ as distributions on the space D of functions which are right continuous and have left-hand limits at all points, converge to the distribution of $Z(t)$, as described in Billingsley [3, chapter 3]. Note that as every realization of $F(t)$ is a distribution function, and thus is right-continuous and monotone, it is in D and thus $Z_\lambda(t)$ is in D.

Proof: By Theorem 15.6 of Billingsley [3, p. 128], it is sufficient to show that (a) for any fixed $t_1, \ldots, t_k \in [0, 1]$ the (finite dimensional) joint distribution of $(Z_\lambda(t_1), \ldots, Z_\lambda(t_k))$ converges to that of $(Z(t_1), \ldots, Z(t_k))$, and (b) $E\{(Z_\lambda(t) - Z_\lambda(t_1))^2 (Z_\lambda(t_2) - Z_\lambda(t))^2\} \leqq (\alpha(t_2) - \alpha(t_1))^2$, for all λ and any $0 \leqq t_1 \leqq t \leqq t_2 \leqq 1$. ((b) implies that the distributions of Z_λ are "tight".) We note that $EZ_\lambda(t) \equiv 0$ all t, $\text{cov}(Z_\lambda(s), Z_\lambda(t)) \equiv \text{cov}(Z(s), Z(t))$ all s, t, by section 2.

Just as is the case with the quantile function, the proof of (a) is easy. In fact, it is accomplished by exactily the same sequence of steps used to prove the joint asymptotic normality of a finite set of quantiles. We omit the details; see Mosteller [10], Wilks [17, p. 271], or David [4, p. 201]. It is in verifying a condition like (b), however, that one encounters the necessity of employing sometimes tedious approximations when dealing with the quantile function. But with the Dirichlet process, (b) follows almost immediately from known properties of the Dirichlet distribution. The left-hand side of the inequality in (b) equals

$$(\lambda+1)^2 E\{(Z_1 - a)^2 (Z_2 - b)^2\}, \text{ where}$$

$$Z_1 = F(t) - F(t_1), \quad a = \alpha(t) - \alpha(t_1), \quad Z_2 = F(t_2) - F(t), \quad b = \alpha(t_2) - \alpha(t_1),$$

and (Z_1, Z_2) has a Dirichlet distribution with parameters $(\lambda a, \lambda b; \lambda(1-a-b))$. Then multiplying out and using standard formulae for the product moments of a Dirichlet distribution (Wilks [17, p. 179]), we find after some algebra that

$$E\{(Z_1-a)^2 (Z_2-b)^2\} = \frac{ab}{(\lambda+1)(\lambda+2)(\lambda+3)} [\lambda + (a+b)(6-\lambda) + 3\lambda ab - 18ab]$$

$$\leqq \frac{ab}{(\lambda+1)(\lambda+2)(\lambda+3)} [\lambda + 6 + \lambda]$$

$$= \frac{2ab}{(\lambda+1)(\lambda+2)}$$

$$\leqq \frac{2ab}{(\lambda+1)^2}$$

$$\leqq \frac{(a+b)^2}{(\lambda+1)^2} .$$

For the first inequality we used the facts $3\lambda ab \leqq \lambda$ and $(a+b)(6-\lambda) \leqq 6$, for the second $\lambda + 2 \geqq \lambda + 1$, and for the third $2ab \leqq (a+b)^2$. Then since $a+b = \alpha(t_2) - \alpha(t_1)$, the final inequality gives (b). The theorem then follows from the previously mentioned theorem in Billinglsley, since $\alpha(t)$ is continuous and monotone.

Q. E. D

Remark: For some applications it is worth noting that essentially the same simple proof can yield more general results. For example, $\alpha(t)$ need not be fixed, it is enough to require that $\nu(t)/\lambda$ converges to a continous limit as $\lambda \to \infty$. Also, the hypothesis that $\alpha(t)$ is continuous can be dispensed with In view of the discussion on page 133 of Billingsley [3] (especially formula (15. 39) the above proof covers this case with only minor adaptation. In fact, essentially the same proof shows weak convergence of the normalized distribution function of a Dirichlet process on the q-dimensional unit cube, by appealing to

Theorem 3 of Bickel and Wichura [18].

Corollary 1: If $F(t)$ is a uniform order statistic process on $[0, 1]$, $(\lambda+1)^{\frac{1}{2}}(F(t) - t)$ converges in distribution to the Gaussian process $W(t)$ on $[0, 1]$ with mean 0 and covariance function $\min(s, t) - st$. ($W(t)$ is sometimes called the Brownian bridge.)

To prove asymptotic normality of S_n we would require the extension of Theorem 1 to the general order statistic processes. We shall present this result only for the case $\alpha(t) \equiv t$ (although the more general case presents no great difficulties), and shall follow essentially the same program as Bickel [2]. The principal difference is that in [2] the sample functions are continuous, here they are only in D. Let $g(x) = \frac{d}{dx} G(x)$.

Theorem 2: If $g(G^{-1}(u))$ is continuous and bounded away from zero on an open interval including $[\epsilon_1, \epsilon_2]$, where $0 < \epsilon_1 < \epsilon_2 < 1$, and $F(t)$ is a uniform order statistic process indexed by λ, then as $\lambda \to \infty$ the process $(\lambda+1)^{\frac{1}{2}}(G^{-1}(F(t)) - G^{-1}(t))$ converges in distribution to $[g(G^{-1}(t))]^{-1} W(t)$ over $[\epsilon_1, \epsilon_2]$, where $W(t)$ is the Brownian bridge of Corollary 1.

Proof: The proof is straightforward, based on the mean value theorem. As $\frac{d}{du} G^{-1}(u) = [g(G^{-1}(u))]^{-1}$ for $\epsilon_1 \leq u \leq \epsilon_2$, $(\lambda+1)^{\frac{1}{2}}(G^{-1}(F(t)) - G^{-1}(t)) = (\lambda+1)^{\frac{1}{2}}(F(t) - t)[g(G^{-1}(\theta(t)))]^{-1}$, where $\theta(t)$ is between $F(t)$ and t. Corollary 1 implies that $\sup_t |F(t) - t| \xrightarrow{P} 0$, thus $\sup_t |\theta(t) - t| \xrightarrow{P} 0$. The conditions of the theorem imply that $[g(G^{-1}(t))]^{-1}$ is uniformly continuous over $[\epsilon_1, \epsilon_2]$. Then Slutsky's theorem implies that the finite dimensional marginal distributions of $(\lambda+1)^{\frac{1}{2}}(G^{-1}(F(t)) - G^{-1}(t))$ converge to those of $[g(G^{-1}(t))]^{-1} W(t)$; further, the uniform continuity of $[g(G^{-1}(t))]^{-1}$ and the necessary and sufficient condition for tightness given by Theorem 15.2 (or 15.3) of Billingsley [3, p. 125] imply that this process inherits the tightness of $Z_\lambda(t)$. The Theorem then follows by Billingsley Theorem 15.1 [3, p. 124].

Q. E. D.

With these results established, one can then proceed to prove the asymptotic normality of S_n under a variety of conditions. For example, the proofs of Bickel [2, section 4] apply with $G^{-1}(F(t))$ replacing the quantile function, with only minor modification.

We should remark that the results concerning S_n that follow from Theorem 2, while applicable in many interesting cases, are much weaker than those found by Shorack [11, 12] using a variant of the stochastic process approach, or Stigler [14, 16] using a projection approach. It remains to be seen whether the civilized behavior of the order statistic process might permit a slight strengthening of Shorack's results.

4. Small Sample Asymptotics.

In section 3 we saw that a linear function of order statistics $S_n = \Sigma c_{in} X^{(i)}$ could be represented as an integral $\int_0^1 G^{-1}(F(t))\, d\, H_n(t)$, where $G^{-1}(F(t))$ is an order statistic process indexed by λ, and $H_n(t) = \Sigma c_{in} I_{[i \leq t(n+1)]}$. For this integral to have the same distribution as S_n it was necessary that $\lambda = n+1$, but the integral in question may make perfectly good sense for any function H_n of bounded variation. In particular, if as $n \to \infty$, $H_n(t) \to H(t)$, we might wish to consider $\int_0^1 G^{-1}(F(t))\, d\, H(t)$ for fixed, finite λ, even λ small.

Why? The answer is simple, and was alluded to in the introduction. The exact distribution of S_n may be rather sensitive to sample size as well as intractable. We have already suggested replacing the sometimes cumbersome quantile function by $G^{-1}(F(t))$, which has a smooth distribution (even if it _is_ a pure jump function with probability one). We now propose replacing the sample size or algorithm dependent weight function $H_n(t)$ by an approximation $H(t)$ and regarding the distribution of the resulting integral $T_n = \int_0^1 G^{-1}(F(t))d\, H(t)$, with $\lambda = n+1$ fixed and small, as a "small sample asymptotic" approximation to the distribution of S_n. We do this in the hope that the distribution of T_n will prove more amenable to analysis than that of S_n, yet for small n provide a better approximation to the distribution of S_n than does the asymptotic normal distribution (c. f. Stigler [16]).

For example, the α-trimmed mean has been defined in many different ways for finite samples. One way is given in [1]; another [15] would take $c_{in} = ([(1-\alpha)n] - [\alpha n])^{-1}$ for $i = [\alpha n] + 1, \ldots, [(1-\alpha)n]$, $c_{in} = 0$ otherwise. In either case H_n is a step function; in either case $H_n(t) \to (1-2\alpha)^{-1}$ if $\alpha < t < 1 - \alpha$, $H_n(t) \to 0$ if $t < \alpha$ or $> 1-\alpha$. Thus we might hope that

$T_n = (1-2\alpha)^{-1} \int_{\alpha}^{1-\alpha} G^{-1}(F(t))\,dt$ would have a distribution that is a useful approximation to that of S_n, and better indicate its performance for small samples than does the limiting normal distribution. Similarly, we might study the behavior of $G^{-1}(F(.5))$ rather than either the exact distribution of the median (whose form depends on whether n is odd or even) or the limiting $N(G^{-1}(.5), (2g(G^{-1}(.5))^{-2})$ distribution. (Note that $G^{-1}(F(.5))$ has exactly the distribution of the median for n odd, for n even it is the "$(n+1)/2$nd order statistic".)

How successful this device is of course depends upon what aspect of the distribution of S_n one is interested in, and on whether G and H are such as to render the distribution of T_n tractable in that aspect. In the following remarks, we concentrate on the variance $V(T_n)$ as an approximation to $V(S_n)$, although when the distribution of S_n is nonnormal $V(S_n)$ may be a misleading measure of the performance of S_n as an estimator (see [1, chapter 5] on this point). We shall proceed heuristically at a number of points, supposing G and H satisfy sufficient regularity conditions that the moments we discuss are finite and given by the integrals in question, and the expansions valid in the range considered.

With this understanding, the variance of T_n is given by

$$V(T_n) = \int_0^1\int_0^1 \operatorname{cov}(G^{-1}(F(s)), G^{-1}(F(t)))\,dH(s)\,dH(t)$$

It is commonly the case with estimators of location parameters that the distribution G is symmetric about a point that may be taken to be zero, and H corresponds to a measure on $[0,1]$ that is symmetric about $.5$ Under these conditions (which are the only ones considered here), we have

$$\int_0^1 G^{-1}(t)\,dH(t) = \int_0^1 EG^{-1}(F(t))\,dH(t) = 0, \text{ and } V(T_n)$$

could also be written

$$V(T_n) = \int_0^1\int_0^1 E[G^{-1}(F(s))\, G^{-1}(F(t))]\, d\,H(s)\, d\,H(t) \quad \text{or}$$

$$= \int_0^1\int_0^1 E\{[G^{-1}(F(s)) - G^{-1}(s)][G^{-1}(F(t)) - G^{-1}(t)]\}\, d\,H(s)\, d\,H(t)\,.$$

If G^{-1} is differentiable with derivatives $\frac{d^k}{dt^k} G^{-1}(t) = g^{(k)}(t)$, then one may consider expanding G^{-1} in a Taylor's series (although other expansions might be better for some distributions) to get

$$V(T_n) = \sum_k \sum_j \int_0^1\int_0^1 \frac{g^{(k)}(s)\, g^{(j)}(t)}{k!\; j!} \cdot f_{jk}(s, t)\, d\,H(s)\, d\,H(s),$$

where $f_{jk}(s,t) = E[(F(s)-s)^j\, (F(t)-t)^k]$, a product central moment for an ordered Dirichlet distribution, expressible in terms of complete gamma functions. When G is symmetric as assumed here, this formidable expression may simplify slightly, since then $g^{(k)}(t) = (-1)^{k+1} g^{(k)}(1-t)$, $f_{jk}(s,t) = (-1)^{j+k} f_{jk}(1-s, 1-t)$, and $g^{(j)}(s)\, g^{(k)}(t)\, f_{jk}(s,t) = g^{(j)}(1-s)\, g^{(k)}(1-t)\, f_{jk}(1-s, 1-t)$.

The implicit hope in performing this expansion is that a very small number of terms will provide an adequate approximation. While the extent to which this hope is realized will not be explored here, we do note the following points in its support:

1. When G is uniform and $G^{-1}(t) = t$, only the first term $k = 1 = j$ is non zero and the expression $(n+2)^{-1} \int\int (\min(s,t) - st)\, d\,H(s)\, d\,H(t)$ gives $V(T_n)$ exactly.

2. The first term $k = 1 = j$ gives $(n+2)^{-1} \int\int [g(G^{-1}(s))\, g\,(G^{-1}(t))]^{-1} [\min(s,t) - st]\, d\,H(s)\, d\,H(t)$ which is just the

asymptotic variance of S_n divided by n+2 rather than the more usual n. Thus additional terms may be viewed as corrections to the approximation given by the asymptotic variance. Incidentally, the change in normalization to $(n+2)^{-1}$ may itself significantly improve the degree of approximation when the $g^{(k)}(t)$ are small for t in the support of H, $k > 1$, as may be the case when trimming is employed.

3. When S_n is just a sample pth percentile, then the approximation suggested is just that based on the probability integral transformation, and goes back at least to Karl Pearson (cf. David [4, p. 65]). For this case there is no novelty in the present approach other than one of the interpretation of a np^{th} order statistic when np is not an integer. For the case of the median, $T_n = G^{-1}(F(.5))$, and if G is symmetric about zero, $g^{(k)}(.5) = 0$ for k even. Then if we keep only the terms corresponding to k=j=1; k=1, j=3; and k=3, j=1 in the above expansion, we find

$$V(T_n) \cong \frac{(g^{(1)}(.5))^2}{4(n+2)} + \frac{g^{(3)}(.5)g^{(1)}(.5)}{3} E(F(.5)-.5)^4$$

$$= V(G, n), \text{ say.}$$

Let

$$V_1(G, n) = (4(n+2))^{-1}(g^{(1)}(.5))^2,$$

the approximation to $V(T_n)$ (and thus to $V(S_n)$) obtained if we only retain the j=k=1 term in the expansion. Then we may compare $V_1(G, n)$, $V(G, n)$, and the more usual formula $V^*(G, n) = (\frac{n+2}{n}) V_1(G, n)$ (the asymptotic variance divided by n) as approximations to $V(S_n)$. Table 1(a-c) presents this comparison for n = 5, 10, 20, 40 and (a) $G = \Phi$, the N(0, 1) distribution, (b) G the standard Cauchy, (c) $G(x) = .75\Phi(x) + .25 S(x)$, where S is the distribution of

ZU^{-1} with Z N(0, 1) and U uniform [0, 1], independent of Z. The "exact" value of $V(S_n)$ was taken from the Princeton Monte Carlo Study [1]. Note that for n odd, $V(S_n) \equiv V(T_n)$; for n even S_n is the average of the comedians (a "quasimedian" in [7]).

n	$nV_1(G, n)$	$nV(G, n)$	$nV^*(G, n)$	$nV(S_n)$
5	1.122	1.220	1.571	1.465
10	1.309	1.411	1.571	1.366
20	1.428	1.505	1.571	1.498
40	1.496	1.544	1.571	1.527

(a) $G = \Phi$, the N(0, 1) distribution.

n	$nV_1(G, n)$	$nV(G, n)$	$nV^*(G, n)$	$nV(S_n)$
5	1.762	2.246	2.467	6.3
10	2.056	2.558	2.467	3.7 or 3.4
20	2.243	2.624	2.467	2.9
40	2.350	2.589	2.467	2.43

(b) G the standard Cauchy.

n	$nV_1(G, n)$	$nV(G, n)$	$nV^*(G, n)$	$nV(S_n)$
5	1.466	1.621	2.052	2.43
10	1.710	1.871	2.052	1.87
20	1.865	1.987	2.052	1.94
40	1.954	2.031	2.052	2.00

(c) $G(x) = .75\,\Phi(x) + .25\,S(x)$, $S(x) = \int_0^1 \Phi(xu^{-1})du$.

Table 1. Comparison of the variance of the median $V(S_n)$ with approximations based on one term of the Taylor's expansion (V_1), on terms through (1, 3) and (3, 1) (V), and on the asymptotic variance (V^*).

On the basis of Table 1 we have at least some encouragement, although the use of the first alone of the expansion does seem to be inferior to its multiple, the asymptotic variance.

As an alternative to the expansion of G^{-1} as a means for determining the variance of $\int G^{-1}(F(t))\, d\, H(t)$, one might focus on a particular family of distributions G, whose inverse cumulatives are particular well-suited for calculation. The most important such examples are Tukey's lambda distributions (cf. Joiner and Rosenblatt [8]). These include the logistic, the uniform, and close approximations to the normal and t-distributions as special cases. Here $G^{-1}(u) = c^{-1}(u^c - (1-u)^c)$, $T_n = c^{-1}\{\int F(t)^c\, d\, H(t) - \int (1-F(t))^c\, d\, H(t)\}$, and if H is symmetric we find

$$V(T_n) = 2c^{-2}\left\{\int_0^1\int_0^1 E(F(s)\, F(t))^c\, d\, H(s)\, d\, H(t) - \int_0^1\int_0^1 E(F(s)(1-F(t)))^c\, d\, H(s)\, d\, H(t)\right\}.$$

This can be expressed in terms of integrals of product moments of Dirichlet random variables, which themselves are ratios of products of complete gamma functions. In this form it may prove feasible to evaluate $V(T_n)$ numerically for a variety of weight functions H. (c. f. Hastings et. al. [19].)

An interesting problem, though unsolved and probably difficult, would be the determination of the optimal weight function H for this "small sample asymptotic" approximation $V(T_n)$ to $V(S_n)$. That is, given G attempt to minimize $V(T_n)$ subject to $\int_0^1 d\, H(t) = 1$ and $\int_0^1 G^{-1}(t)\, d\, H(t) = 0$. This could be attempted either generally or within a restricted family such as Tukey's Lambda family. In this latter case, if $c > 0$ the easy solution is to let H put mass .5 at 0 and at 1 (the "midrange"), but otherwise the solutions, which might provide more efficient estimators than do weights based on the asymptotic variance, are unknown.

5. The Uniform Order Statistic Process, and an Application to the Yule Process.

Going back at least to Laplace, ingenious statisticians have derived an incredible variety of distributional properties for functions of the order statistics $U^{(1)} \leqq \ldots \leqq U^{(n)}$ of a sample from a uniform [0, 1] distribution. David [4] is a good source of examples. Since all of these properties are really properties of the Dirichlet distribution of the sample spacings $U^{(i)} - U^{(i-1)}$, it's not surprising that generalizations of these properties for the uniform order statistic process exist and are easy to find and prove. We shall content ourselves with one example which has an interesting application in finding the passage time distribution of a Yule process.

A well-known property of uniform order statistics is that $[U^{(j)}/U^{(j+1)}]^j$, $j = 1, 2, \ldots, n$, are themselves independent and uniform [0, 1] distributed, where $U^{(n+1)} = 1$. See David [4, P. 19]. If $F(t)$ is a uniform order statistic process indexed by λ, then a generalization of this property is that if $\lambda > 0$, $\beta > -\lambda^{-1}$, and $t_j = \beta + j\lambda^{-1}$, then

$$\left(\frac{F(t_j)}{F(t_{j+1})}\right)^{\lambda t_j}, \quad \text{all } j \text{ such that } 0 < t_j < t_{j+1} \leqq 1,$$

are independently distributed uniform [0, 1]. We omit the easy proof, based on the representation of the Dirichlet distribution in terms of independent gamma distributed random variables. When $\lambda = n + 1$ and $\beta = 0$ this reduces to the earlier property.

The reason for citing this generalization is because it allows a remarkably easy derivation of the distribution of the time to the k^{th} "birth" in a Yule process. For the present purpose we shall define a Yule process as a pure birth process whose birth intensities δ_m increase as $\delta_m = m + \theta$ (the treatment of the case $\delta_m = \phi(m + \theta)$ would follow by a change of scale). Such a counting process may be characterized by a sequence of independent exponentially distributed random variables $W_1, W_2, \ldots$, where $EW_i = \delta_i^{-1}$ and $\sum_{i=1}^{k} W_i$ gives the time of the k^{th} birth.

The distribution of $B_k = \sum_{i=1}^{k} W_i$ can be easily derived by the following trick. First note that $\exp[-\delta_i W_i]$ are independent uniform $[0, 1]$. Then

$$\exp(B_k) = \prod_{i=1}^{k} [\exp\{-\delta_i W_i\}]^{-\delta_i^{-1}}$$

has, by the above with $\lambda = k + \theta$ and $\beta = \theta(k+\theta)^{-1}$, the same distribution as

$$\prod_{i=1}^{k} [F(t_i)/F(t_{i+1})]^{-\lambda t_i \delta_i^{-1}} = (F(t_1))^{-1} ,$$

since then $\lambda t_i \delta_i^{-1} = 1 = t_{k+1}$, and this product telescopes. Thus B_k has the same distribution as $-\ln F(\frac{\theta+1}{\theta+k})$, where $F(\frac{\theta+1}{\theta+k})$ has a beta $(\theta+1, k-1)$ distribution!

Footnotes

1. The given counts are from the first editions [9] of volumes 1-4 (1798-1805), and exclude separately paged supplements to volumes 3 and 4.

2. One way of viewing the relationship between the quantile function $Y_n(t)$ and $G^{-1}(F(t))$ is to note that in the uniform case ($G(x) \equiv x$), the properties of the Dirichlet distribution imply that

   $Y_n(t) = E(F(t) \mid F(i/(n+1))) = X^{(i)}$, $i=1, \ldots, n$), if $\lambda = n+1$.

* This research was sponsored by the National Science Foundation under Grant No. SOC75-02922.

References

[1] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W., Robust estimates of Location: Survey and Advances, Princeton: Princeton University Press, 1972.

[2] Bickel, P. J., "Some contributions to the theory of order statistics", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I (L. LeCam and J. Neyman, eds.), 575-591, Berkeley: University of California Press, 1967.

[3] Billingsley, P., Convergence of Probability Measures, New York: John Wiley and Sons, Inc. 1968.

[4] David, H. A., Order Statistics, New York: John Wiley and Sons, Inc. 1970.

[5] Ferguson, T. S., "A Bayesian analysis of some non-parametric problems," Annals of Statistics 1 (March 1972) 209-230.

[6] Ferguson, T. S., "Prior distributions on spaces of probability measures", Annals of Statistics 2 (July 1974), 615-629.

[7] Hodges, J. L. Jr. and Lehmann, E. L., "On medians and quasi-medians", Journal of the American Statistical Association, 62 (Sept. 1967), 926-31.

[8) Joiner, B. L., and Rosenblatt, J. R., "Some properties of the range in samples from Tukey's symmetric lambda distributions," Journal of the American Statistical Association, 66 (June 1971), 394-399.

[9] Laplace, P. S., Traité de mécanique céleste. First editions of volumes 1-4, 1798-1805, Volume 5, 1825. Paris: Courcier.

[10] Mosteller, F., "On some useful "inefficient" statistics," Annals of Mathematical Statistics, 17 (Dec. 1946), 377-408.

[11] Shorack, G. R., "Asymptotic normality of linear combination of functions of order statistics," Annals of Mathematical Statistics, 40 (Dec. 1969), 2041-2050.

[12] Shorack, G. R., "Functions of order statistics," Annals of Mathematical Statistics, 43 (March 1972) 412-427.

[13] Shorack, G. R., "Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-i. i. d. case," Annals of Statistics, 1 (January 1973), 146-152.

[14] Stigler, S. M., "Linear functions of order statistics," Annals of Mathematical Statistics, 40 (June 1969), 770-788.

[15] Stigler, S. M., "The asymptotic distribution of the trimmed mean," Annals of Statistics, 1 (May 1973), 472-477.

[16] Stigler, S. M., "Linear functions of order statistics with smooth weight functions," Annals of Statistics, 2 (July 1974), 676-693.

[17] Wilks, S. S., Mathematical Statistics, New York: John Wiley and Sons, Inc., 1962.

[18] Bickel, P. J., and Wichura, M. J., "Convergence criteria for multiparameter stochastic processes and some applications," Annals of Mathematical Statistics 42 (Oct. 1971), 1656-1670.

[19] Hastings, C. Jr., Mosteller, F., Tukey, J. W., and Winsor, C. P., "Low moments for small samples: a comparative study of order statistics," Annals of Mathematical Statistics 18 (Sept. 1947), 413-426.