Biometrika (1999), **86**, 4, *pp.* 909–921

Printed in Great Britain

Two-sample quantile tests under general conditions

By MICHAEL R. KOSOROK

Department of Statistics, Biostatistics & Medical Informatics, University of Wisconsin, K6/428 Clinical Science Center, 600 Highland Avenue, Madison, Wisconsin 53792, U.S.A.

kosorok@biostat.wisc.edu

Summary

A simple, nonparametric two-sample test for equality of a given collection of quantiles is developed which can be applied to a variety of empirical distribution functions, including the Kaplan–Meier estimator, a self-consistent estimator for doubly-censored data and an estimator for repeated measures data. The null hypothesis tested is that the quantiles are equal but other aspects of the distributions may differ between the two samples. This procedure can also be applied to quantile testing in group sequential clinical trials with staggered patient entry. A simple simulation study demonstrates that the moderate sample size properties of this procedure are reasonable.

Some key words: Doubly-censored data; Empirical distribution function; Group sequential methods; Kaplan–Meier estimator; Kernel density estimation; Minimum dispersion statistic; Nonparametric methods.

1. Introduction

Populations can often be usefully compared in terms of quantiles. In children with cystic fibrosis, the 10th percentiles of height and weight are important clinical boundaries between healthy and possibly nutritionally compromised patients (Farrell et al., 1997). Since the median height or weight is associated with normal growth status, it may be meaningful with cystic fibrosis clinical trials to compare two treatment groups on the basis of both the 10th and 50th percentiles of height and/or weight. In addition, censoring or truncation can preclude estimation of entire distribution functions and an examination of a collection of quantiles is a reasonable alternative. Use of quantiles also offers robustness against outliers.

Recently, several promising nonparametric two-sample median comparison procedures for censored survival data have been developed. The earliest of these, proposed by Wang & Hettmansperger (1990), requires either the two-sample shift model assumption or estimation of the involved densities. To avoid density estimation, Su & Wei (1993) developed a minimum dispersion statistic based on the Kaplan–Meier estimator; see also Basawa & Koul (1988). The fact that Su & Wei's statistic is easily computed and asymptotically chi-squared is appealing, but their analytical approach cannot be directly applied to group sequential clinical trials with staggered patient entry. Keaney & Wei (1994) manage to solve this difficult problem by using an interesting extension of the resampling procedure of Parzen, Wei & Ying (1994).

In this paper, we develop a nonparametric two-sample test for equality of a given collection of quantiles which can be applied to a variety of empirical distribution functions, including both the Kaplan–Meier estimator, Turnbull's (1974) self-consistent survival estimator for doubly-censored data and an estimator for repeated measures data. For instance, the collection of quantiles examined can consist of several interim estimates of the median, or some other quantile, in a group sequential clinical trial setting involving two treatment arms; or it can consist of several quantiles to be compared between two groups at one analysis time, such as the 10th and 50th percentiles of height or weight. Under the null hypothesis of equal quantiles, the proposed test statistic is asymptotically normal with a covariance which can be consistently estimated. Not only does the proposed test procedure provide an analytical alternative to the resampling method of Keaney & Wei (1994), but it applies to more general censoring schemes and to a variety of empirical distribution functions.

Although one-sample quantile estimation procedures for right-censored data have been extensively studied (Brookmeyer & Crowley, 1982; Doss & Gill, 1992; Efron, 1981; Emerson, 1982; Li et al., 1996; Padgett, 1986; Reid, 1981; Simon & Lee, 1982; Slud, Byar & Green, 1984), we develop in § 2 one-sample results for empirical distribution estimators which are sufficiently general to allow for double censoring and for several kinds of empirical distribution estimator. Methods of density estimation are examined in § 3, while § 4 utilises the results of § 3 to formulate the proposed asymptotically distribution-free test statistics. A simple simulation study of the proposed methods is given in § 5 to evaluate moderate sample size properties.

2. Distribution theory for quantiles

For a distribution function $F:\Re\mapsto[0,1]$, let $F^{-1}:[0,1]\mapsto\Re$ be the usual inverse distribution function defined by letting $F^{-1}(p)\equiv\inf\{x:F(x)\geqslant p\}$. The estimators of $F^{-1}(p)$, for a given p, which we shall use take the form $\hat{F}^{-1}(p)$, where $\hat{F}:\Re\mapsto[0,1]$ is an appropriate estimator of F. We shall require that F and $\hat{F}$ satisfy the following conditions at $t=F^{-1}(p)$ for some strictly positive, increasing, normalising sequence $\{c_n, n\geqslant 1\}$ with $\lim_{n\to\infty} c_n=\infty$.

Condition 1. Distribution function F has a density f in a neighbourhood of t such that f is continuous at t and $0<f(t)<\infty$.

Condition 2. As $n\to\infty$, $c_n\{\hat{F}(t)-F(t)\}$ converges in distribution to a bounded random variable with continuous distribution function.

Condition 3. For every $\varepsilon>0$,

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty} \mathrm{pr}\left\{\sup_{s:|t-s|<\delta} c_n|\hat{F}(t)-F(t)-\hat{F}(s)+F(s)|>\varepsilon\right\}=0.$$

Condition 1 is fairly standard (Serfling, 1980, p. 77), while Conditions 2 and 3 are not quite so standard but will permit the greater generality we are seeking. Condition 1 with the asymptotic stochastic equicontinuity in Condition 3 together yield sufficient continuity of $\hat{F}$ to allow the weak convergence of $\hat{F}$ in Condition 2 to be carried through to $\hat{F}^{-1}(p)$. Although Condition 3 appears difficult to verify, we will see in several examples given later in this section that it is satisfied by a wide variety of distribution function estimators.

The following lemma allows us to establish an asymptotic equivalence between the centred quantile estimator and the centred distribution function estimator.

LEMMA 1. *Suppose F satisfies Condition* 1 *and $\hat{F}$ satisfies Conditions* 2 *and* 3 *at $t = F^{-1}(p)$ for a specified $p \in (0, 1)$. Then*

$$c_n\{\hat{F}^{-1}(p) - F^{-1}(p)\} = -\frac{c_n[\hat{F}\{F^{-1}(p)\} - p]}{f\{F^{-1}(p)\}} + o_p(1).$$

The proof of Lemma 1 depends on the following lemma.

LEMMA 2. *Assume the conditions of Lemma* 1 *are satisfied, although Condition* 2 *is not needed at this point, and let $t \equiv F^{-1}(p)$. Then, for any compact $K \subset \Re$,*

$$\sup_{u \in K} |c_n\{\hat{F}(t + u/c_n) - \hat{F}(t)\} - f(t)u| = o_p(1),$$

Proof. We have

$$|c_n\{\hat{F}(t + u/c_n) - \hat{F}(t)\} - f(t)u| \leqslant c_n|\hat{F}(t + u/c_n) - F(t + u/c_n) - \hat{F}(t) + F(t)| + |c_n\{F(t + u/c_n) - F(t)\} - f(t)u|,$$

and the result follows from Conditions 1 and 3 and the compactness of K. □

In fact, Lemma 2 gives a stronger result than necessary, but this stronger result will prove useful later on.

Proof of Lemma 1. Denote by $I_{\{A\}}$ the indicator of the event A. For each $x \in \Re$,

$$I_{\{\hat{F}^{-1}(p) \leqslant x\}} = I_{\{\hat{F}(x) \geqslant p\}}$$

by the definition of $\hat{F}^{-1}$. However, this implies that

$$I_{\{c_n[\hat{F}^{-1}(p) - F^{-1}(p)] \leqslant x\}} = I_{\{c_n[\hat{F}\{F^{-1}(p) + x/c_n\} - p] \geqslant 0\}}; \tag{1}$$

but

$$c_n\left[\hat{F}\left\{F^{-1}(p) + \frac{x}{c_n}\right\} - p\right] = f\{F^{-1}(p)\}x + c_n[\hat{F}\{F^{-1}(p)\} - p] + R_n(x),$$

where $R_n(x)$ converges to zero in probability, as $n \to \infty$, by Lemma 2. Condition 2 now gives us that, for all n large enough, the right-hand side of (1) is equal to

$$I_{\{-c_n[\hat{F}\{F^{-1}(p)\} - p] \leqslant x f\{F^{-1}(p)\}\}} + o_p(1),$$

and the result follows. □

Sometimes it is useful to modify a distribution function estimator slightly, for example by connecting jump points.

Condition 4. For every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \text{pr}\left\{\sup_{s: |t-s| < \delta} c_n|\tilde{F}(s) - \hat{F}(s)| > \varepsilon\right\} = 0.$$

It is fairly easy to show that, if at $t = F^{-1}(p)$, for $p \in (0, 1)$, a modification $\tilde{F}$ of a distribution function estimator $\hat{F}$ satisfies Condition 4, then $\tilde{F}$ satisfies Conditions 2 and 3 if and only if $\hat{F}$ satisfies Conditions 2 and 3. The following lemma shows us that Condition 4 is satisfied for a variety of modifications made by connecting jump points.

LEMMA 3. *Let $\hat{F}$ be a right-continuous, piecewise constant estimator of F, where F and*

$\hat{F}$ satisfy Conditions 1, 2 and 3 at $t = F^{-1}(p)$ for a chosen $p \in (0, 1)$; and let $\tilde{F}$ be the modification of $\hat{F}$ made by connecting adjacent jump points of F with straight lines. Then Condition 4 is satisfied.

Proof. By Condition 1, there exists a $\delta > 0$ such that F is continuous and f is strictly positive on $(t - 2\delta, t + 2\delta)$. Thus, for n large enough, the probability that a jump point of $\hat{F}$ occurs in both $(t - 2\delta, t - \delta)$ and $(t + \delta, t + 2\delta)$ can be made arbitrarily close to 1 by Conditions 2 and 3. Hence

$$\sup_{s:|t-s|<\delta} c_n|\tilde{F}(s) - \hat{F}(s)| \leqslant \sup_{s:|t-s|<2\delta} c_n|\hat{F}(s) - \hat{F}(s-)|,$$

and the result follows from reapplication of Condition 3. □

Example 1: *Right-censored data.* For right-censored data, the theory in Su & Wei (1993) establishes the conclusions of our Lemma 1. However, we include it here for completeness and also to establish the conclusions of Lemma 3 for the Kaplan–Meier estimator. Let (T_i, C_i), for $i = 1, \ldots, n$, be independent pairs of failure and censoring times, where the distribution function for T_i is F. The data available in this setting are the pairs (X_i, δ_i), for $i = 1, \ldots, n$, where $X_i = T_i \wedge C_i$, $\delta_i = I_{\{X_i = T_i\}}$, and $x \wedge y$ is the minimum of x and y. Let

$$\bar{Y}(s) \equiv \sum_{i=1}^{n} I_{\{X_i \geqslant s\}}$$

be the number at risk at time s, and assume that $n^{-1}\bar{Y}(s)$ converges in probability, as $n \to \infty$, to $\pi(s)$ uniformly on the interval $[0, u]$, where u is larger than $t = F^{-1}(p)$ for a chosen p. Let $\hat{F} = 1 - \hat{S}$, where $\hat{S}$ is the usual Kaplan–Meier estimator.

If F is absolutely continuous and if $\pi(s) > 0$, for all s in some neighbourhood of t, then Theorem 6.3.1 of Fleming & Harrington (1991) yields weak convergence in the Skorohod topology of $n^{\frac{1}{2}}\{\hat{F}(.) - F(.)\}$ to a tight, continuous Gaussian process on the interval $[0, t + \varepsilon]$, for some $\varepsilon > 0$. Thus, Conditions 1, 2 and 3 are satisfied. Gill (1980, Ch. 4) shows, by use of a random time transform, that F need not be absolutely continuous, except for our purposes we still need Condition 1 satisfied at t, and $n^{\frac{1}{2}}\{\hat{F}(.) - F(.)\}$ still converges weakly to a tight Gaussian process which is continuous in a neighbourhood of t but need not be continuous over all of $[0, t + \varepsilon]$. The results of Lemmas 1, 2 and 3 will thus all follow, since $\hat{F}$ is right-continuous and piecewise constant.

Example 2: *Doubly-censored data.* The distribution of quantiles estimated from doubly-censored data has not been obtained in previous work. However, the results of Chang (1990) establish the weak convergence results for a self-consistent estimator, $\hat{F}$, of the distribution function for doubly-censored data; this is essentially Turnbull's (1974) self-consistent estimator. Doubly-censored data in this context arise from independent and identically distributed triplets (X_i, Y_i, Z_i), for $i = 1, \ldots, n$, where all random variables are nonnegative and the X_i, with distribution function F, are failure times of interest while Y_i and Z_i, with $\text{pr}(Z_i \leqslant Y_i) = 1$, are the double-censoring variables. We only observe X_i if it falls in the random interval $[Z_i, Y_i]$; otherwise, we know only whether $X_i < Z_i$ or $X_i > Y_i$. Thus the data we actually observe consist of the independent and identically distributed pairs (W_i, δ_i), where $W_i = (X_i \wedge Y_i) \vee Z_i$,

$$\delta_i = \begin{cases} 1 & \text{if } Z_i \leqslant X_i \leqslant Y_i, \\ 2 & \text{if } X_i > Y_i, \\ 3 & \text{if } X_i < Z_i, \end{cases}$$

and where $x \vee y$ is the maximum of x and y.

If we assume F is continuous and satisfies Condition 1 at $t = F^{-1}(p)$, and if Chang's (1990) Conditions A1–A6 are appropriately satisfied, then Chang's Theorem 3.1 yields that $n^{\frac{1}{2}}\{\hat{F}(.) - F(.)\}$ converges weakly in the Skorohod topology to a tight, continuous Gaussian process, over an interval containing $t + \varepsilon$, for some $\varepsilon > 0$. Thus Conditions 2 and 3 are satisfied. Chang's Conditions A1–A6 are realistic and not difficult to verify in practice, but they are somewhat tedious and we will omit them. An inspection of the Fredholm integral equation which defines $\hat{F}$ as given in § 2 of Chang (1990), where our $\hat{F}$ equals Chang's $1 - S_X^{(n)}$, yields that $\hat{F}$ is almost surely right-continuous and piecewise constant, for all n sufficiently large. Thus the results of Lemmas 1, 2 and 3 follow.

Example 3: *Repeated measures.* When repeated measurements are taken from several individuals, an empirical distribution function and quantile estimator can still be constructed but the dependence within individuals must be accounted for. Assume that n independent and identically distributed individuals, each with m_i repeated measures X_{ij}, for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$, are sampled, and that all X_{ij} have common marginal distribution F. If F is absolutely continuous in some neighbourhood of t, and if we assume $m_i = m < \infty$ for all individuals, it is straightforward to show that F and

$$\hat{F}(x) \equiv (mn)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} I_{\{x \geqslant X_{ij}\}}$$

satisfy Conditions 1 and 2. The only difficult condition to satisfy here is Condition 3. To see how this is done, note that $\hat{F}$ can be written

$$\hat{F}(x) = \sum_{j=1}^{m} m^{-1} \frac{\sum_{i=1}^{n} I_{\{x \geqslant X_{ij}\}}}{n},$$

and that each of the m components are empirical distribution functions based on independent and identically distributed data; thus the stochastic equicontinuity of the usual empirical distribution function carries over to $\hat{F}$. Similar results can be obtained when the m_i are allowed to vary from individual to individual while restricting m_i such that $1 \leqslant m_i \leqslant m$ for some finite m.

3. Kernel density estimation

In this section, we develop kernel density estimators which are needed for estimating the variance of the foregoing quantile statistics. The following lemma gives the main result of this section.

Lemma 4. *Suppose F satisfies Condition* 1 *and $\hat{F}$ satisfies Conditions* 2 *and* 3 *at $t = F^{-1}(p)$ for a specified $p \in (0, 1)$. Suppose also that K is a nonnegative kernel with bounded total variation, Lebesgue integral* 1 *and compact support on $\Re$; that the window width b_n is* *positive with* $b_n = o_p(1)$ *and* $(c_n b_n)^{-1} = O_p(1)$; *and that the estimator* $t_n = t + o_p(1)$. *Then*

$$\int_{\Re} b_n^{-1} K\left(\frac{t_n - x}{b_n}\right) d\hat{F}(x) = f(t) + o_p(1).$$

The proof of Lemma 4 depends on the following lemma.

LEMMA 5. *Assume the conditions of Lemma* 4 *are satisfied. Assume also that* $u \neq v$ *are real numbers. Then*

$$b_n^{-1}\left\{\frac{\hat{F}(t_n - ub_n) - \hat{F}(t_n - vb_n)}{v - u}\right\} = f(t) + o_p(1).$$

Proof. We have

$$\left| b_n^{-1}\left\{\frac{\hat{F}(t_n - ub_n) - \hat{F}(t_n - vb_n)}{v - u}\right\} - f(t)\right|$$
$$\leqslant \frac{b_n^{-1}}{|v-u|}\,|\hat{F}(t_n - ub_n) - F(t_n - ub_n) - \hat{F}(t_n - vb_n) + F(t_n - vb_n)|$$
$$+ \left|\frac{b_n^{-1}}{v-u}\{F(t_n - ub_n) - F(t_n - vb_n)\} - f(t)\right|,$$

where the first term on the right-hand side is $o_p(1)$, by Conditions 2 and 3, while the second term converges to zero, as $n \to \infty$, by Condition 1. □

Proof of Lemma 4. Let the support of K be denoted by $T \subset \Re$ with $U(T) \equiv \inf(T)$ and $V(T) \equiv \sup(T)$. The conditions on K ensure that, for every $\varepsilon > 0$, there exists an approximation K_ε composed of a finite linear combination of rectangular kernels each with support on a subset of T such that $\sup_{t \in T} |K_\varepsilon(t) - K(t)| \leqslant \varepsilon$. Let K_ε^* denote the rectangular kernel $K_\varepsilon^*(x) = \{V(T) - U(T)\}^{-1} I_{\{x \in [U(T), V(T)]\}}$. Then

$$\left|\int_{\Re} b_n^{-1} K\left(\frac{t_n - x}{b_n}\right) d\hat{F}(x) - f(t)\right| \leqslant \left|\int_{\Re} b_n^{-1} K_\varepsilon\left(\frac{t_n - x}{b_n}\right) d\hat{F}(x) - f(t)\right|$$
$$+ \varepsilon\{V(T) - U(T)\}\int_{\Re} b_n^{-1} K_\varepsilon^*\left(\frac{t_n - x}{b_n}\right) d\hat{F}(x),$$

where Lemma 5 can be used to establish that the first term on the right-hand side is $o_p(1)$, while the second term converges to $\varepsilon\{V(T) - U(T)\} f(t)$. The result now follows since ε is arbitrary. □

The minimum dispersion statistic of Su & Wei (1993) for right-censored data, discussed in the Introduction, appears to permit two-sample inference on quantiles while bypassing the need to estimate the density. However, this is not the case; as suggested by a referee, it can be shown that the minimum dispersion statistic implicitly uses a kernel estimator with window width $O_p(n^{-\frac{1}{2}})$. Such a window width satisfies the criteria given in Lemma 4 and would thus work, at least asymptotically, for any empirical distribution estimator satisfying Conditions 2 and 3. An advantage of such an approach is that the choice of kernel and window width would be automatic.

Let G be the distribution function for a second sample with corresponding empirical distribution estimator $\hat{G}$ satisfying Conditions 1, 2 and 3 at a chosen $t = F^{-1}(p) = G^{-1}(p)$, under the null hypothesis of equal p-quantiles. Rather than utilise a generalisation of the minimum dispersion statistic, one could simply use the density estimators

$$\frac{\hat{F}\{\hat{G}^{-1}(p)\}-p}{\hat{G}^{-1}(p)-\hat{F}^{-1}(p)}, \quad \frac{\hat{G}\{\hat{F}^{-1}(p)\}-p}{\hat{F}^{-1}(p)-\hat{G}^{-1}(p)}. \tag{2}$$

These estimators thus have similar properties to the implicit density estimators associated with minimum dispersion statistics in that they are automatic, have window widths $O_p(n^{-\frac{1}{2}})$ and otherwise satisfy the criteria of Lemma 4. However, these window widths lead to suboptimal convergence. Better choices, such as window widths of order $O_p(n^{-1/5})$, are available for the usual independent and identically distributed data setting (Hall et al., 1991; Silverman, 1986, Ch. 3).

4. Multivariate two-sample tests

4·1. *General theory*

We now present a general framework for multivariate two-sample quantile tests. In this framework we have two independent samples, one containing information about the distribution function F and the other containing information about the distribution function G. We are interested in testing the null hypothesis

$$H_0: F^{-1}(p_j)=G^{-1}(p_j) \quad (j=1,\ldots,J),$$

where $p_1,\ldots,p_J$ are not necessarily distinct. The statistics of interest to us are the quantile pairs $\{\hat{F}_j^{-1}(p_j), \hat{G}_j^{-1}(p_j)\}$ for $j=1,\ldots,J$, where the distribution function estimators $\hat{F}_j$, for $j=1,\ldots,J$, come from the sample corresponding to the distribution function F while the estimators $\hat{G}_j$, for $j=1,\ldots,J$, come from the sample corresponding to G. If we have a single distribution function estimator from each sample but are interested in testing equality of J distinct quantiles, then $\hat{F}_1=\ldots=\hat{F}_J=\hat{F}$ and $\hat{G}_1=\ldots=\hat{G}_J=\hat{G}$ but $p_1,\ldots,p_J$ are distinct. If, on the other hand, we wish to construct group sequential boundaries for testing equality of a single quantile, $F^{-1}(p)=G^{-1}(p)$, then $p_1=\ldots=p_J=p$, but the pairs $\{\hat{F}_j^{-1}(p), \hat{G}_j^{-1}(p)\}$ come from data obtained at J different analysis times. Even more complex settings are possible; for example, we could test equality of a collection of quantiles at several different analysis times. We have the following corollary.

Corollary 1. *Assume H_0 obtains and the following conditions are satisfied.*

(i) *We require that F, $\hat{F}_j$, G and $\hat{G}_j$ satisfy Conditions* 1, 2 *and* 3 *at $t=F^{-1}(p_j)=G^{-1}(p_j)$, for $j=1,\ldots,J$.*

(ii) *The chosen kernel density estimator $\hat{f}_j$ of $f\{F^{-1}(p_j)\}$ and $\hat{g}_j$ of $g\{G^{-1}(p_j)\}$ satisfy the conditions of Lemma* 4, *for $j=1,\ldots,J$.*

(iii) *The vector of statistics $(c_n[\hat{F}_1\{F^{-1}(p_1)\}-p_1],\ldots,c_n[\hat{F}_J\{F^{-1}(p_J)\}-p_J])^{\mathrm{T}}$ is asymptotically a zero-mean Gaussian process with covariance matrix $\Phi=(\phi_{jk})$. Also, the vector of statistics $(c_n[\hat{G}_1\{G^{-1}(p_1)\}-p_1],\ldots,c_n[\hat{G}_J\{G^{-1}(p_J)\}-p_J])^{\mathrm{T}}$ is asymptotically a zero-mean Gaussian process with covariance matrix $\Gamma=(\gamma_{jk})$. Furthermore, these vectors are asymptotically independent.*

(iv) *The covariance estimator $\hat{\Phi}=(\hat{\phi}_{jk})$ is consistent for Φ; and the covariance estimator $\hat{\Gamma}=(\hat{\gamma}_{jk})$ is consistent for Γ.*

Then the vector of statistics $(c_n\{\hat{F}_1^{-1}(p_1)-\hat{G}_1^{-1}(p_1)\},\ldots,c_n\{\hat{F}_J^{-1}(p_J)-\hat{G}_J^{-1}(p_J)\})^{\mathrm{T}}$ is asymptotically a zero-mean Gaussian process with covariance matrix $\Psi=(\psi_{jk})$, where Ψ can *be consistently estimated by* $\hat{\Psi} = (\hat{\psi}_{jk})$, *where, for each j and k,*

$$\hat{\psi}_{jk} \equiv \frac{\hat{\phi}_{jk}}{\hat{f}_j \hat{f}_k} + \frac{\hat{\gamma}_{jk}}{\hat{g}_j \hat{g}_k}.$$

The proof is an immediate consequence of Lemmas 1 and 4.

4·2. *Examining a collection of quantiles*

As illustrated by the height and weight example given in the Introduction, it is sometimes useful to compare two populations in terms of a finite collection of quantile probabilities. It is known that Canadian and U.S. cystic fibrosis patients differ in many ways (Lai et al., 1998), but if the 10th and 50th percentiles are in agreement the same diagnostic boundaries distinguishing small, an indication of potential malnutrition, and normal-sized patients from other patients could be utilised in both countries.

Let $\hat{F}$ and $\hat{G}$ be the distribution function estimates from samples 1 and 2, respectively. After selecting the J quantile probabilities to be compared $(p_1, \ldots, p_J)$, let $\hat{F}_j(p_j) \equiv \hat{F}(p_j)$ and $\hat{G}_j(p_j) \equiv \hat{G}(p_j)$, for $j = 1, \ldots, J$. The methods of § 4·1 can now be readily applied to this setting, provided we can obtain the necessary covariance estimators $\hat{\Phi}$ and $\hat{\Gamma}$. To illustrate this in a simple setting, assume that $\hat{F}$ and $\hat{G}$ are the usual empirical distribution function estimators for independent and identically distributed observations; essentially, Example 1 from § 2 is applicable for each sample if we assume no censoring. Let n_1 be the sample size associated with $\hat{F}$, and let n_2 be the sample size associated with $\hat{G}$, with $n \equiv n_1 + n_2$ and $c_n \equiv n^{\frac{1}{2}}$. Assume that n_1/n converges to an element in $(0, 1)$, that Condition 1 is satisfied at the points $F^{-1}(p_j)$, for $j = 1, \ldots, J$, for both F and G, and that the null hypothesis $H_0 : F^{-1}(p_j) = G^{-1}(p_j)$, for $j = 1, \ldots, J$, obtains.

In this setting, $\hat{\phi}_{jk} = n(p_j \wedge p_k - p_j p_k)/n_1$ and $\hat{\gamma}_{jk} = n(p_j \wedge p_k - p_j p_k)/n_2$ are consistent estimators of ϕ_{jk} and γ_{jk}, respectively; and Conditions 2 and 3 are readily satisfied for $\hat{F}$ and $\hat{G}$ at the chosen quantiles. To estimate the densities, we could either use $O_p(n^{-\frac{1}{2}})$ window estimators of the form given in (2), or optimal-order estimators such as the $O_p(n^{-1/5})$ window estimators

$$\hat{f}_j \equiv \int_{\Re} n_1^{1/5} \hat{Q}_F^{-1} K \left\{ \frac{\hat{F}_j^{-1}(p_j) - x}{n_1^{-1/5} \hat{Q}_F} \right\} d\hat{F}(x), \tag{3}$$

$$\hat{g}_j \equiv \int_{\Re} n_2^{1/5} \hat{Q}_G^{-1} K \left\{ \frac{\hat{G}_j^{-1}(p_j) - x}{n_2^{-1/5} \hat{Q}_G} \right\} d\hat{G}(x), \tag{4}$$

where $\hat{Q}_F$ and $\hat{Q}_G$ are twice the estimated interquartile ranges of F and G, respectively, and where the kernel is triangular:

$$K(x) = \begin{cases} x + 1 & \text{if } x \in [-1, 0], \\ 1 - x & \text{if } x \in (0, 1], \\ 0 & \text{if } |x| > 1. \end{cases}$$

Corollary 1 now gives all that is necessary to obtain an asymptotically chi-squared statistic with J degrees of freedom for testing H_0.

Now consider the slightly more complex setting of Example 3, where the data for each sample consist of independent and identically distributed clusters of fixed size m. Let n_1 be the number of clusters in the sample with marginal distribution F, labelled as sample 1, and let n_2 be the number of clusters in the sample with marginal distribution G, labelled as sample 2, with $n=n_1+n_2$ and $c_n=n^{\frac{1}{2}}$, and with n_1/n converging to an element in (0, 1) as before. Also let the data in sample 1 be X_{il} for $l=1,\ldots,m$ and $i=1,\ldots,n_1$, and the data in sample 2 be Y_{il} for $l=1,\ldots,m$ and $i=1,\ldots,n_2$. As in Example 3, we will use the empirical distribution estimators

$$\hat{F}(x)=(mn_1)^{-1}\sum_{i=1}^{n_1}\sum_{l=1}^{m} I_{\{x\geqslant X_{il}\}},\quad \hat{G}(x)=(mn_2)^{-1}\sum_{i=1}^{n_2}\sum_{j=1}^{m} I_{\{x\geqslant Y_{ij}\}}.$$

We have previously demonstrated that Conditions 2 and 3 can be easily satisfied for $\hat{F}$ and $\hat{G}$, thus establishing part (i) of Corollary 1; and it is not difficult to establish part (ii) for the repeated measures analogue of the kernel density estimators given in 3 and 4. Part (iii) is also easily established, and the covariance estimators

$$\hat{\phi}_{jk}=n\times n_1^{-1}\sum_{i=1}^{n_1}\left(\left[m^{-1}\sum_{l=1}^{m} I_{\{\hat{F}^{-1}(p_j)\geqslant X_{il}\}}-\hat{F}\{\hat{F}^{-1}(p_j)\}\right]\times\left[m^{-1}\sum_{l=1}^{m} I_{\{\hat{F}^{-1}(p_k)\geqslant X_{il}\}}-\hat{F}\{\hat{F}^{-1}(p_k)\}\right]\right),$$

$$\hat{\gamma}_{jk}=n\times n_2^{-1}\sum_{i=1}^{n_2}\left(\left[m^{-1}\sum_{l=1}^{m} I_{\{\hat{G}^{-1}(p_j)\geqslant Y_{il}\}}-\hat{G}\{\hat{G}^{-1}(p_j)\}\right]\times\left[m^{-1}\sum_{l=1}^{m} I_{\{\hat{G}^{-1}(p_k)\geqslant Y_{il}\}}-\hat{G}\{\hat{G}^{-1}(p_k)\}\right]\right)$$

readily satisfy part (iv). Hence Corollary 1 again gives all that is necessary to obtain an asymptotically chi-squared statistic with J degrees of freedom for testing H_0.

4·3. *Group sequential clinical trials with staggered entry*

We now consider two-arm survival clinical trials with possibly staggered entry of patients and several interim analyses. This is the setting considered by Keaney & Wei (1994). We will assume that the two arms are statistically independent. In this context, if Conditions 1, 2 and 3 are satisfied for each arm separately, and if the assumptions of Corollary 1 are satisfied, then the methods of § 4·1 can be applied directly. We will assume that testing the equality of a single quantile, with quantile probability p, is of interest. For now, we will focus on the first arm only. Let the distribution function for this arm be F, and assume that Condition 1 is satisfied for F at $t=\xi\equiv F^{-1}(p)$. The notation of Example 1 of § 2 needs to be extended to allow for multiple interim analyses.

Suppose we have $K<\infty$ interim analyses with altogether n independent and identically distributed subjects. The data available in this setting are the pairs (X_{ij},δ_{ij}), for $i=1,\ldots,n$ and $j=1,\ldots,K$, where $X_{ij}=T_i\wedge C_{ij}$, $\delta_{ij}=I_{\{X_{ij}=T_i\}}$, T_i is the true failure time for individual i, and where C_{ij} is the effective censoring time for subject i and interim analysis j. All time is measured from enrolment; hence, if individual i has not been enrolled by the jth interim analysis, then $C_{ij}=0$. This censoring notation incorporates random loss-to-follow-up as well as several other censoring schemes. We will also assume that the information is cumulative over j, in the sense that $C_{ij}\leqslant C_{ik}$ for all $1\leqslant j<k\leqslant K$ and $i=1,\ldots,n$. Let $N_{ij}(t)\equiv I_{\{X_{ij}\leqslant t,\delta_{ij}=1\}}$ and $Y_{ij}(t)\equiv I_{\{X_{ij}\geqslant t\}}$, and define

$$\bar{N}_j(t)\equiv\sum_{i=1}^{n} N_{ij}(t),\quad \bar{Y}_j(t)\equiv\sum_{i=1}^{n} Y_{ij}(t).$$

Let $\hat{S}_j$ be the Kaplan–Meier estimator at interim analysis j, for $\hat{F}_j = 1 - \hat{S}_j$, and $\hat{\xi}_j \equiv \hat{F}_j^{-1}(p)$. We will also assume, as in § 2, that $n^{-1}\bar{Y}_j(t)$ converges in probability, as $n \to \infty$, to $\pi_j(t)$ uniformly on the interval $[0, u]$, where u is larger than ξ. From the results presented in Example 1, it is clear that, if $\pi_1(t) > 0$, for all t in some neighbourhood of ξ, then $\hat{F}_j$ satisfies Conditions 2 and 3 for $j = 1, \ldots, K$. Thus the only remaining work to be done before the results of § 4·2 can be utilised is to establish joint asymptotic normality of $(n^{\frac{1}{2}}\{\hat{F}_1(\xi) - p\}, \ldots, n^{\frac{1}{2}}\{\hat{F}_K(\xi) - p\})^{\mathrm{T}}$, and to develop a consistent estimator of the associated covariance matrix. This is accomplished in the following lemma.

LEMMA 6. *Assume that $\pi_1(t) > 0$ for all t in some neighbourhood of ξ, which implies the same for $\pi_2, \ldots, \pi_K$. Then*

(i) $(n^{\frac{1}{2}}\{\hat{F}_1(\xi) - p\}, \ldots, n^{\frac{1}{2}}\{\hat{F}_K(\xi) - p\})^{\mathrm{T}}$ *is asymptotically a zero-mean Gaussian process with covariance matrix $\Phi = (\phi_{jk})$, where*

$$\phi_{jk} = \begin{cases} \phi_{kk} & \text{if } j \leqslant k, \\ \phi_{jj} & \text{if } j > k, \end{cases}$$

for $j = 1, \ldots, K$ and $k = 1, \ldots, K$;

(ii) *the estimator $\hat{\Phi}$, with diagonal elements*

$$\hat{\phi}_{jj} \equiv \hat{S}_j^2(\hat{\xi}_j) \int_0^{\hat{\xi}_j} \frac{d\bar{N}_j(t)}{\{\bar{Y}_j(t) - \Delta\bar{N}_j(t)\}\bar{Y}_j(t)},$$

and off-diagonal elements

$$\hat{\phi}_{jk} = \begin{cases} \hat{\phi}_{kk} & \text{if } j < k, \\ \hat{\phi}_{jj} & \text{if } j > k, \end{cases}$$

for $j = 1, \ldots, K$ and $k = 1, \ldots, K$, is consistent for Φ.

Proof. As pointed out by a referee, the Kaplan–Meier estimator is an efficient score statistic, and the results of Theorem 1 of Scharfstein, Tsiatis & Robins (1997) can be applied to obtain part (i). To obtain part (ii), we can apply the martingale calculations given in § 3·2, especially p. 104, of Fleming & Harrington (1991) to establish that, for all t in a neighbourhood of ξ for which $\pi_j(t) > 0$, the variance of $n^{\frac{1}{2}}\{\hat{F}_j(t) - F(t)\}$ can be consistently estimated by

$$\hat{V}(t) \equiv \hat{S}_j^2(t) \int_0^t \frac{n d\bar{N}_j(s)}{\{\bar{Y}_j(s) - \Delta\bar{N}_j(s)\}\bar{Y}_j(s)}.$$

The result now follows from the consistency of $\hat{\xi}_j$ and the continuity of $\hat{V}(t)$ in a neighbourhood of ξ. □

If a statistically independent second sample with failure time distribution function G and sample size n^* also satisfies Conditions 1, 2 and 3 as well as the assumptions of Lemma 6, then all of the assumptions of Corollary 1 will be satisfied and the results of § 4·1 will follow, provided $n/(n + n^*)$ converges to an element in (0, 1).

5. A SIMULATION STUDY

The following small simulation study is meant to demonstrate that the moderate sample size properties of the proposed chi-squared statistic for examining a collection of quantiles are reasonable. We examine the statistical test proposed in § 4·2 for the collection of quantiles consisting of the 25th and 75th percentiles in both the independent and repeated measures data settings. Three different distributions are examined, denoted by D_1, D_2 and D_3. Of these, D_1 is a mixture distribution, having a $\frac{1}{3}$ probability of being a positive exponential with standard deviation 1 and having a $\frac{2}{3}$ probability of being a negative exponential with standard deviation 2; a random variable with distribution D_2 is simply the negative of a random variable with distribution D_1; and a random variable with distribution D_3 is simply a random variable with distribution D_2 minus the constant 1·674. The 25th and 75th percentiles of D_1 are −0·940 and 0·087; the 25th and 75th percentiles of D_2 are −0·087 and 0·940; and the 25th and 75th percentiles of D_3 are the same as the corresponding percentiles of D_1.

To evaluate type I error and power for moderate sample sizes in the independent data setting, we generated 5000 pairs of samples for small, $n_1 = n_2 = 20$, and large, $n_1 = n_2 = 200$, samples for three different scenarios. Kernel estimators were used with both the $O_p(n^{-1/5})$ window width of (3) and (4) as well as the $O_P(n^{-\frac{1}{2}})$ window width of (2). In the first scenario, both samples in the pairs were generated from the D_1 distribution; for the second scenario, the first sample of each pair was generated from the D_1 distribution while the second sample was generated from the D_3 distribution. The first scenario allows us to evaluate the type I error when the two distributions being compared are identical, while the second scenario allows us to evaluate the type I error when the 25th and 75th percentiles match but the densities do not. In both scenarios, the null hypothesis of equal 25th and 75th quantiles obtains. For the third scenario, the D_1 and D_2 distributions were compared to get some idea of power under an alternative hypothesis.

For each pair of samples generated, the statistical test proposed in § 4·2 was computed and compared against the $\alpha = 0{\cdot}05$ level critical value for a chi-squared distribution with 2 degrees of freedom. We used 5000 replications to ensure that our Monte Carlo error for estimating a probability in the region of 0·05 is about $0{\cdot}003 = (0{\cdot}05 \times 0{\cdot}95/5000)^{\frac{1}{2}}$. For the quantile estimates, we utilised the modified empirical distribution function made by connecting jump points as presented in Lemma 3.

This entire procedure was duplicated for repeated measures data with cluster sizes of 4. To generate each cluster, the 4 elements of the cluster either shared the same random variable, with probability $\frac{1}{3}$, or consisted of 4 independent and identically distributed random variables, with probability $\frac{2}{3}$, where all random variables had the same distribution. In this setting, each cluster contains the same amount of information as two independent random variables; for this reason, the small sample size scenarios used $n_1 = n_2 = 10$ independent clusters per group while the large sample size scenarios used $n_1 = n_2 = 100$ independent clusters per group. The results of this simulation study are summarised in Table 1.

Under the null hypothesis, this test procedure tends to be conservative for kernel window widths of $O_p(n^{-\frac{1}{2}})$ yet less conservative for window widths of $O_p(n^{-1/5})$ when sample sizes are small. However, this conservatism goes away with increasing sample size. It also appears that conservatism under the null hypothesis is associated with a decrease in power under the alternative hypothesis. The use of the more optimal window width really does seem to improve the small-sample performance. The reason for the generally better performance under the D_1 versus D_3 scenario is unclear, but it may be because the information for the least informative comparison between quantiles is greater in the D_1 versus D_3 comparison than in the D_1 versus D_1 comparison.

While the results of this simulation study are encouraging, it is unclear whether or not

Table 1. *Simulated power for tests of size* $\alpha = 0{\cdot}05$ *based on* 5000 *replications*

Distributions compared	Kernel window length	Independent data: Sample size per group	Independent data: Simulated power	Repeated measures: Sample size per group	Repeated measures: Simulated power
D_1 versus D_1	$O_p(n^{-1/5})$	20	0·030	10	0·025
		200	0·044	100	0·035
	$O_p(n^{-1/2})$	20	0·001	10	0·007
		200	0·026	100	0·025
D_1 versus D_3	$O_p(n^{-1/5})$	20	0·060	10	0·047
		200	0·052	100	0·053
	$O_p(n^{-1/2})$	20	0·004	10	0·008
		200	0·034	100	0·030
D_1 versus D_2	$O_p(n^{-1/5})$	20	0·558	10	0·516
		200	1·000	100	1·000
	$O_p(n^{-1/2})$	20	0·305	10	0·347
		200	1·000	100	1·000

the bootstrap approach of Keaney & Wei (1994) for right-censored data may be more effective in small clinical trials than the analytical method proposed in § 4·3. A careful simulation comparison of these two approaches, building upon the results of § 4·3, along with additional analytical insights, would be of great practical value.

ACKNOWLEDGEMENT

This research was supported by a grant from the U.S. National Cancer Institute. The author would like to thank Dr Jeffrey Douglas for several helpful discussions and the editor and a referee for several helpful suggestions.

REFERENCES

BASAWA, I. V. & KOUL, H. L. (1988). Large-sample statistics based on quadratic dispersion. *Int. Statist. Rev.* **56**, 199–219.

BROOKMEYER, R. & CROWLEY, J. (1982). A confidence interval for the median survival time. *Biometrics* **38**, 29–41.

CHANG, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18**, 391–404.

DOSS, H. & GILL, R. D. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *J. Am. Statist. Assoc.* **87**, 869–77.

EFRON, B. (1981). Censored data and the bootstrap. *J. Am. Statist. Assoc.* **76**, 312–9.

EMERSON, J. D. (1982). Nonparametric confidence intervals for the median in the presence of right censoring. *Biometrics* **38**, 17–27.

FARRELL, P. M., KOSOROK, M. R., LAXOVA, A., SHEN, G., KOSCIK, R. E., BRUNS, T., SPLAINGARD, M. & MISCHLER, E. H. (1997). Nutritional benefits of newborn screening for cystic fibrosis. *New Engl. J. Med.* **337**, 963–9.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

GILL, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tracts 124. Amsterdam: Mathematisch Centrum.

HALL, P., SHEATHER, S. J., JONES, M. C. & MARRON, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* **78**, 263–9.

KEANEY, K. M. & WEI, L. J. (1994). Interim analyses based on median survival times. *Biometrika* **81**, 279–86.

LAI, H.-C., COREY, M., FITZSIMMONS, S., KOSOROK, M. R. & FARRELL, P. M. (1999). Comparison of growth status in patients with cystic fibrosis between the United States and Canada. *Am. J. Clin. Nutr.* **69**, 531–8.

LI, G., HOLLANDER, M., MCKEAGUE, I. W. & YANG, J. (1996). Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data. *Ann. Statist.* **24**, 628–40.

PADGETT, W. J. (1986). A kernel-type estimator of a quantile function from right-censored data. *J. Am. Statist. Assoc.* **81**, 215–22.

PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.

REID, N. (1981). Estimating the median survival time. *Biometrika* **68**, 601–8.

SCHARFSTEIN, D. O., TSIATIS, A. A. & ROBINS, J. M. (1997). Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. *J. Am. Statist. Assoc.* **92**, 1342–50.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

SIMON, R. & LEE, Y. J. (1982). Nonparametric confidence limits for survival probabilities and median survival time. *Cancer Treat. Rep.* **66**, 37–42.

SLUD, E. V., BYAR, D. P. & GREEN, S. B. (1984). A comparison of reflected versus test-based confidence intervals for the median survival time, based on censored data. *Biometrics* **40**, 587–600.

SU, J. Q. & WEI, L. J. (1993). Nonparametric estimation for the difference or ratio of median failure times. *Biometrics* **49**, 603–7.

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Am. Statist. Assoc.* **69**, 169–73.

WANG, J. L. & HETTMANSPERGER, T. P. (1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data. *J. Am. Statist. Assoc.* **85**, 529–36.

[*Received April* 1998. *Revised November* 1998]