GENERALIZED PEARSON–FISHER CHI–SQUARE GOODNESS–OF–FIT TESTS, WITH APPLICATIONS TO MODELS WITH LIFE HISTORY DATA

by

Gang Li and Hani Doss
Purdue University | Florida State University

Technical Report # 92-48C

Department of Statistics
Purdue University

October, 1992

△

Generalized Pearson-Fisher Chi-Square Goodness-of-Fit Tests, With Applications To Models With Life History Data

Gang Li
Purdue University

Hani Doss
Florida State University

Abstract

Suppose that $X_1, \ldots, X_n$ are i.i.d. $\sim F$, and we wish to test the null hypothesis that F is a member of the parametric family $\mathcal{F} = \{F_\theta(x);\ \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^q$. The classical Pearson-Fisher chi-square test involves partitioning the real axis into k cells $I_1, \ldots, I_k$ and forming the chi-square statistic $X^2 = \sum_{i=1}^k \frac{(O_i - nF_{\hat\theta}(I_i))^2}{nF_{\hat\theta}(I_i)}$, where O_i is the number of observations falling into cell i and $\hat\theta$ is the value of θ minimizing $\sum_{i=1}^k \frac{(O_i - nF_\theta(I_i))^2}{nF_\theta(I_i)}$. We obtain a generalization of this test to any situation for which there is available a nonparametric estimator $\hat F$ of F for which $n^{\frac{1}{2}}(\hat F - F) \xrightarrow{d} W$ where W is a continuous zero mean Gaussian process satisfying a mild regularity condition. We allow the cells to be data dependent. Essentially, we estimate θ by the value $\hat\theta$ that minimizes a "distance" between the vectors $(\hat F(I_1), \ldots, \hat F(I_k))$ and $(F_\theta(I_1), \ldots, F_\theta(I_k))$, where distance is measured through an arbitrary positive definite quadratic form, and then form a chi-square type test statistic based on the difference between $(\hat F(I_1), \ldots, \hat F(I_k))$ and $(F_{\hat\theta}(I_1), \ldots, F_{\hat\theta}(I_k))$. We prove that this test statistic has asymptotically a chi-square distribution with $k - q - 1$ degrees of freedom, and point out some errors in the literature on chi-square tests in survival analysis. Our procedure is very general and applies to a number of well-known models in survival analysis, such as right censoring and left truncation. We apply our method to deal with questions of model selection in the problem of estimating the distribution of the length of the incubation period of the AIDS virus using the CDC's data on blood-transfusion related AIDS. Our analysis suggests some models that seem to fit better than those used in the literature.

Keywords and phrases: goodness-of-fit test, Pearson-Fisher chi-square test, chi-squared statistic, left truncation, right censoring, Aalen model.

Research supported by Air Force Office of Scientific Research Grant 90-0202.

1 Introduction and Summary

Let $X_1, \ldots, X_n$ be i.i.d. from a distribution function F. To test the null hypothesis that F is equal to a completely specified distribution function F_0, K. Pearson (1900) introduced the now classical chi-square test, which involves partitioning the real line into k cells and forming the chi-square statistic $X^2 = \sum_{i=1}^k (O_i - np_i)^2/np_i$, where O_i is the number of observations falling into cell i and np_i is the expected value of O_i under the null hypothesis. Pearson showed that for large n, the distribution of X^2 is approximately chi-square with $k-1$ degrees of freedom. It is rare that one wants to test the null hypothesis that F equals a completely specified F_0. The more common situation is that we wish to test the null hypothesis H_0 that F is a member of a certain parametric family F_θ, $\theta \in \Theta$, where Θ is an open subset of $\mathbb{R}^q$. In this case, $p_1, \ldots, p_k$ $(k > q)$ are functions of θ_0, the true value of θ, and are no longer known. Fisher (1922, 1924) showed that if θ_0 is estimated by the value minimizing $\sum_{i=1}^k (O_i - np_i(\theta))^2/np_i(\theta)$, then X^2 has, for large n, approximately a chi-square distribution with $k-q-1$ degrees of freedom. The estimate of θ_0 obtained in this fashion is called the minimum chi-square estimator. It is important to note that Fisher's result is valid only if θ_0 is estimated by the minimum chi-square estimator (or an estimator asymptotically equivalent to it). Chernoff and Lehmann (1954) observed that if θ_0 is estimated by the more efficient maximum likelihood estimator based on the whole sample, then the asymptotic distribution of X^2 is that of $\sum_{i=1}^{k-q-1} Z_i^2 + \sum_{i=k-q}^{k-1} \lambda_i Z_i^2$, where the Z_i's are independent from the normal distribution with mean 0 and standard deviation 1, $0 \le \lambda_i \le 1$, and the λ_i's depend on the unknown θ_0. Thus, use of the more efficient maximum likelihood estimator enables us to "partially recoup" the q lost degrees of freedom. This is, however, at the cost of complicating the analysis since the limit distribution is neither tabulated nor independent of θ_0.

Since Fisher's (1922, 1924) papers, there has been sustained interest in the general problem of testing goodness of fit of a parametric family, and in chi-square tests in particular, and in recent years much of this interest has focused on models arising in survival analysis which are more complicated than the one in which we observe $X_1, \ldots, X_n$ i.i.d. $\sim F$. The reason for this interest is that in many situations, there are physical reasons that indicate specific parametric families. Exponential distributions arise in a very large number of contexts; extreme value distributions arise frequently in reliability theory because they are the limiting distributions of the lifelength of series or parallel systems with a large number of identically distributed components. There are also cases where preliminary nonparametric studies suggest a specific parametric model. If a goodness-of-fit test can lead the investigator to accept a certain parametric model, then this can lead to statistical procedures that are substantially more efficient than those based on nonparametric models. Moreover, the analysis is then more parsimonious and so easier to understand, and can enable some inferences, for example about tail behavior, that are impossible under a nonparametric model. See Miller (1983).

In survival analysis, the data are often not completely observed. For example, a very common situation is that of right censoring, where for some values of l, X_l is not observed, but it is known only that $X_l > c_l$ where c_l is observed. In the expression for X^2, we therefore do not have access to O_i. Suppose now that we have an estimator $\hat{F}$ of

F with the property that, whether or not the parametric model holds, we have

$$n^{1/2}(\hat{F} - F) \xrightarrow{d} W \tag{1.1}$$

for some process W, where the convergence is in an appropriate Skorohod space. Let $I_1, \ldots, I_k$ denote the k cells in the partition of $\mathbb{R}$ and define

$$\boldsymbol{\zeta}(\theta) = n^{1/2}\Big(\hat{F}(I_1) - p_1(\theta), \ldots, \hat{F}(I_k) - p_k(\theta)\Big). \tag{1.2}$$

To assess the fit of the parametric family it is natural to take an estimator $\hat{\theta}$ of θ_0 and consider the measure of discrepancy $\boldsymbol{\zeta}(\hat{\theta})$. If $D(\theta)$ is a symmetric matrix, then one can form the vector $\boldsymbol{\xi}(\theta) = D(\theta)\boldsymbol{\zeta}(\theta)$ and estimate θ_0 by the parameter value minimizing the quadratic form $\boldsymbol{\xi}'(\theta)\boldsymbol{\xi}(\theta) = \boldsymbol{\zeta}'(\theta)D^2(\theta)\boldsymbol{\zeta}(\theta)$. Such an estimator is called a minimum chi-square estimator. The purpose of this paper is to show that some of the original ideas in Fisher (1922, 1924) can be pushed through to obtain classes of chi-square goodness-of-fit tests in a very general framework: We show that if

A (1.1) holds and W is a continuous Gaussian process such that

$$\mathrm{Cov}(W(t_1), \ldots, W(t_{k-1})) \text{ is nonsingular whenever } t_1 < t_2 < \cdots < t_{k-1}, \tag{1.3}$$

B $\mathrm{Cov}(W(t_1), \ldots, W(t_{k-1}))$ can be consistently estimated for all $t_1 < t_2 < \cdots < t_{k-1}$,

C $\hat{\theta}$ is the value of θ minimizing $\boldsymbol{\zeta}'(\theta)D^2(\theta)\boldsymbol{\zeta}(\theta)$, where $D(\theta)$ is positive definite for all θ and satisfies some mild regularity conditions (see Section 2.1)

then $\boldsymbol{\xi}(\hat{\theta}) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, where Σ is a nonnegative definite matrix of rank $k - q - 1$, and which can be consistently estimated by an estimator $\hat{\Sigma}$. Let $\hat{\Sigma}^\dagger$ denote the Moore-Penrose inverse of $\hat{\Sigma}$. We also show that $\boldsymbol{\xi}'(\hat{\theta})\hat{\Sigma}^\dagger\boldsymbol{\xi}(\hat{\theta}) \xrightarrow{d} \chi^2_{k-q-1}$.

We observe that for the classical case considered by Pearson and Fisher, the diagonal matrix $D(\theta) = \mathrm{diag}\Big((p_1(\theta))^{-\frac{1}{2}}, \ldots, (p_k(\theta))^{-\frac{1}{2}}\Big)$ has the property that $D^2(\theta_0)$ is a generalized inverse of the limiting covariance matrix of $\boldsymbol{\zeta}(\theta_0)$ (i.e. the covariance matrix for the multinomial distribution). In Section 2.3, we consider the natural special case where $D(\theta_0)$ is the square root of a generalized inverse of the limiting covariance matrix of $\boldsymbol{\zeta}(\theta_0)$.

Our results apply to a fixed partition of $\mathbb{R}$, or to a partition where the cell boundaries are chosen as a function of the data. This gives rise to chi-square tests that are very easy to use.

It is difficult to find examples which violate Condition (1.3) provided neither F nor $\mathrm{Var}(V)$ have flat spots (i.e. under simple and natural nondegeneracy conditions). In Section 2 we show that (1.3) is satisfied whenever W has the form $W \stackrel{d}{=} (1 - F) \cdot V$ where V is a Gaussian martingale, a form intimately connected with Aalen's multiplicative intensity model. See Andersen and Borgan (1985) for a description and review. Here, we will say only that this is an extremely important model which encompasses a very wide range of situations arising in survival analysis, including quite general forms of censoring (censoring by fixed constants, Type II censoring, and the important special case of random censoring), random truncation models, and of course the i.i.d. setup described earlier. These models are described and discussed in Section 3. There, we show that in the special case where we observe i.i.d. observations from a distribution function F, our test statistic reduces to the original Pearson-Fisher chi-square test statistic. For the case where the data undergo Type II censoring, we obtain a test studied by Mihalko and Moore (1980).

Hjort (1990) has also developed tests of goodness of fit of a parametric family in the framework of the Aalen model. His approach involves hazard functions and their cumulatives, and for the case where we observe i.i.d. observations $X_1, \ldots, X_n$, his tests do not reduce to the classical Pearson-Fisher chi-square test. To describe his approach, let $\lambda_\theta(t) = F'_\theta(t)/\bar{F}_\theta(t)$ be the hazard rate and $\Lambda_\theta(t) = \int_0^t \lambda_\theta(s)ds$ be the cumulative hazard rate. Let $\hat{\Lambda}^{nonp.}$ be the standard Nelson-Aalen estimator of Λ_{θ_0}; this is a non-parametric estimator which is valid whether or not the parametric model holds. Also let $\hat{\Lambda}^{par.}(t) = \int_0^t \lambda_{\hat{\theta}}(s)ds$, where $\hat{\theta}$ is the maximum likelihood estimate of θ_0. Hjort's approach involves comparing $\hat{\Lambda}^{nonpar.}$ and $\hat{\Lambda}^{par.}$. More specifically he establishes, using the well-developed theory of counting processes, that under the null hypothesis, for a large class of weight functions K_n, if $H_n(t) = \int_0^t K_n(s)d(\hat{\Lambda}^{nonpar.} - \hat{\Lambda}^{par.})(s)$, then $H_n(t)$ converges in distribution to some process $H(t)$. Results of this sort are very often used to obtain chi-square tests because of the difficulties in getting a handle on the distribution of the process $H(t)$. For the cells $I_1, \ldots, I_k$ in the partition of $\mathbb{R}$, let $\Delta H_n(I_i) = \int_{I_i} dH_n(s)$. Then the vector $\boldsymbol{w} = (\Delta H_n(I_1), \ldots, \Delta H_n(I_k))$ satisfies $\boldsymbol{w} \xrightarrow{d} \mathcal{N}(0, R)$ where R is a possibly singular matrix. Let $\hat{R}$ be a consistent estimate of R and $\hat{R}^-$ be a generalized inverse of $\hat{R}$. He proposes the test statistic $\boldsymbol{w}'\hat{R}^-\boldsymbol{w}$ and shows that this has a limiting chi-square distribution with degrees of freedom equal to the rank of R. Unfortunately, the rank of R depends on the model under consideration and on the parametric family in question, and so additional work is required for each new application. We point out that each weight function gives rise to a chi-square test, and if we specialize Hjort's results to the random censorship model, then for a particular choice of K_n the test is the same as one proposed independently by Akritas (1988). We discuss Hjort's paper further in Section 5, where we also point out some errors in the literature. Our approach is different from Hjort's. We take as our starting point any model for which Conditions A and B hold, and we produce a test statistic which we show is asymptotically chi-square with number of degrees of freedom *always* equal to $k - q - 1$. This extends the applicability of the Pearson-Fisher chi-square test while retaining its simplicity.

There is a large literature on goodness-of-fit tests in survival analysis. A recent review is contained in Hollander and Peña (1990).

The rest of the paper is organized as follows. Section 2 gives the statements of our main results. In Section 3 we show how our test reproduces some tests already present in the literature, and we apply our procedure to obtain chi-square tests for some well-known models, including models with right-censored data and those with left-truncated data. In Section 4 we apply our test to deal with some questions of model selection in the problem of estimating the distribution of the length of the incubation period of the AIDS virus using data on blood-transfusion related AIDS. We use our procedure to examine some parametric assumptions made in the literature. Our analysis suggests some parametric models that appear to fit better. Section 5 gives proofs of our main theoretical results, and discusses some problems that need to be addressed when establishing that a quadratic form has an asymptotic chi-square distribution. In the Appendix we show that minimum chi-square estimators satisfy the regularity conditions needed for our main results to hold.

2 The Generalized Pearson-Fisher χ^2 Test

Assume that $\hat{F}$ is a nonparametric estimator of F satisfying

$$n^{\frac{1}{2}}(\hat{F} - F) \xrightarrow{d} W \quad \text{in} \quad D[\epsilon, M], \tag{2.1}$$

where W is a continuous Gaussian process with zero mean, $D[\epsilon, M]$ is the standard Skorohod space on $[\epsilon, M]$, and $-\infty \leq \epsilon < M \leq \infty$ (in survival analysis, ϵ will usually be greater than or equal to 0). We shall develop a chi-square statistic for testing the null hypothesis and our statistic will be based on $\hat{F}$ and any estimator of θ_0 asymptotically equivalent to a minimum chi-square estimator. Note that θ_0, the true value of θ, is of course unknown; however, the limiting distribution of our test statistics will not depend on the value of θ_0.

2.1 Notation and Assumptions

For each n, let $-\infty = a_0^{(n)} < a_1^{(n)} < \cdots < a_k^{(n)} = \infty$ be a partition of the real line such that each cell boundary $a_i^{(n)} \equiv a_i(\hat{F})$ is a functional of $\hat{F}$ and converges in probability to a constant a_i, where $\epsilon < a_1, \quad a_{k-1} < M$. Let

$$p_i(\theta) = F_\theta(a_i) - F_\theta(a_{i-1}) \tag{2.2}$$

for all $i = 1, \ldots, k$ and $\theta \in \Theta$. We assume that $p_i(\theta) > 0$ for all θ. Define the covariance matrix

$$\Sigma^{(1)}(\theta, \boldsymbol{t}) = \text{Cov}(W(t_1), \ldots, W(t_{k-1})) \quad \text{when} \quad F = F_\theta, \tag{2.3}$$

and assume that

$$\Sigma^{(1)} = \Sigma^{(1)}(\theta_0, \boldsymbol{a}) \quad \text{is nonsingular.} \tag{2.4}$$

This condition is weaker than (1.3), but is in fact all that we will need. Define

$$p_i^{(n)}(\theta) = F_\theta(a_i^{(n)}) - F_\theta(a_{i-1}^{(n)}) \tag{2.5}$$

and

$$\hat{p}_i = \hat{F}(a_i^{(n)}) - \hat{F}(a_{i-1}^{(n)}), \tag{2.6}$$

and let $\boldsymbol{p}(\theta)$, $\boldsymbol{p}^{(n)}(\theta)$, and $\hat{\boldsymbol{p}}$ denote the vectors corresponding to (2.2), (2.5), and (2.6). (Note: We are assuming tacitly that $\hat{F}(-\infty) = 0$ and that $\hat{F}(\infty) = 1$. If this is not the case, then $\hat{p}_1$ and $\hat{p}_k$ must be defined as $\hat{p}_1 = \hat{F}(a_1^{(n)})$ and $\hat{p}_k = 1 - \hat{F}(a_{k-1}^{(n)})$). Denote $\boldsymbol{a} = (a_1, \ldots, a_{k-1})'$, $\boldsymbol{a}^{(n)} = (a_1^{(n)}, \ldots, a_{k-1}^{(n)})'$ and let $\text{N}(\boldsymbol{a})$ be a neighborhood of $\boldsymbol{a}$ in $\mathbb{R}^{k-1}$. Assume that $D(\theta, \boldsymbol{t})$ is a $k \times k$ symmetric matrix whose elements are (known) functions of $(\theta, \boldsymbol{t})$ on $\Theta \times \text{N}(\boldsymbol{a})$ and satisfies the following regularity conditions

R1 $D(\cdot, \cdot)$ is continuous at $(\theta_0, \boldsymbol{a})$, R2 $D^{-1}(\cdot,\cdot)$ exists and is bounded on $\Theta \times \mathrm{N}(\boldsymbol{a})$,

R3 $\frac{\partial}{\partial\theta}(D^2(\theta,\boldsymbol{t}))$ exists at every $(\theta,\boldsymbol{t}) \in \Theta \times \mathrm{N}(\boldsymbol{a})$ and $\frac{\partial}{\partial\theta}(D^2(\theta,\boldsymbol{t}))$ is continuous at $(\theta_0,\boldsymbol{a})$.

The choice of $D(\theta,\boldsymbol{t})$ is discussed in Section 2.3.

We use $D(\theta)$ and $D_n(\theta)$ to denote the matrices $D(\theta,\boldsymbol{a})$ and $D(\theta,\boldsymbol{a}^{(n)})$, respectively. Define

$$\boldsymbol{\zeta}_n(\theta) = n^{\frac{1}{2}}(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta)) \quad \text{and} \quad \boldsymbol{\xi}_n(\theta) = D_n(\theta)\boldsymbol{\zeta}_n(\theta). \tag{2.7}$$

We shall use $\boldsymbol{\xi}_n(\theta)$ to construct a statistic to test the null hypothesis.

There are two important special cases in the above framework. When fixed cells are used (i.e. the $a_i^{(n)}$'s are independent of the data), then the $\hat{p}_i$'s are random quantities while the $p_i^{(n)}$'s are deterministic. This is the standard setup used in the classical Pearson test. It is usually more useful to take the quantiles of $\hat{F}$ as cell boundaries. In this case, the $\hat{p}_i$'s are deterministic quantities while the $p_i^{(n)}$'s depend on the sample. One can show that $\hat{F}^{-1}(u) \xrightarrow{P} F^{-1}(u)$ for each $u \in (F^{-1}(\epsilon), F^{-1}(M))$, if (2.1) holds. (Doss and Gill (1990) give a stronger result about weak convergence of the process $n^{\frac{1}{2}}(\hat{F}^{-1}(u) - F^{-1}(u))$.) One of the advantages of using random cells is that one can then ensure that the $\hat{p}_i$'s are not too small.

The following assumptions are made throughout the paper.

A1 $F_\theta(x)$ is continuously differentiable in θ and x.

A2 The matrix

$$\frac{\partial \boldsymbol{p}(\theta)}{\partial \theta'} = \begin{pmatrix} \frac{\partial p_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial p_1(\theta)}{\partial \theta_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_k(\theta)}{\partial \theta_1} & \cdots & \frac{\partial p_k(\theta)}{\partial \theta_q} \end{pmatrix}_{k\times q}$$

is of rank q for all $\theta \in \Theta$.

A3 The estimate $\hat{\theta}$ satisfies

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = \left(C'C\right)^{-1} C'\boldsymbol{\xi}_n(\theta_0) + o_p(1) \tag{2.8}$$

where $C = D(\theta_0)\dfrac{\partial \boldsymbol{p}(\theta_0)}{\partial \theta'}$.

Remark 2.1 *Suppose that $\hat{\theta}$ is a minimum chi-square estimator, i.e.*

$$\hat{\theta} \text{ is the value of } \theta \text{ minimizing } \boldsymbol{\xi}_n'(\theta)\boldsymbol{\xi}_n(\theta). \tag{2.9}$$

In the case where we have completely observed data, fixed cells, and $D(\theta)$ is chosen to be the diagonal matrix $D(\theta) = diag\big((p_1(\theta))^{-\frac{1}{2}}, \ldots, (p_k(\theta))^{-\frac{1}{2}}\big)$, it is well known (and not difficult to see) that $\hat{\theta}$ satisfies (2.8); see e.g. Section 30.3 of Cramér (1946). This is still true in the general situation. A proof is given in Lemma A.1 in the Appendix. We prefer to take (2.8) rather than (2.9) as our condition on $\hat{\theta}$ because of the slight increase in generality.

Note that

$$\boldsymbol{p}^{(n)}(\hat{\theta}) = \boldsymbol{p}(\theta_0) + o_p(1) \quad \text{and} \quad \frac{\partial \boldsymbol{p}^{(n)}(\hat{\theta})}{\partial \theta'} = \frac{\partial \boldsymbol{p}(\theta_0)}{\partial \theta'} + o_p(1) \tag{2.10}$$

by Assumptions A1 and A3. Also, the existence of the matrix inverse in (2.8) is guaranteed by Assumption A2.

From now on, 1 denotes a column vector of 1's where the dimension is taken from context.

2.2 Main Theorems and Construction of the Test Statistic

Theorems 1, 2, and 3 below give our main findings. Theorem 1 gives the limiting distribution of $\boldsymbol{\xi}_n(\hat{\theta})$. The proof is based on the original ideas of Fisher. Theorem 2 gives the rank of the asymptotic covariance matrix and also gives the ranks of certain natural estimates of it. Theorem 3 states that the quadratic form which is the test statistic has an asymptotic chi-square distribution with $k-q-1$ degrees of freedom. This theorem follows directly from Theorem 1 and both parts of Theorem 2, and constitutes our main result.

Theorem 1 *Let $\boldsymbol{\xi}_n(\theta)$ be defined by (2.7) where $D(\cdot,\cdot)$ satisfies R1–R3, and assume A1–A3. If (2.1) holds, then under H_0*

$$\boldsymbol{\xi}_n(\hat{\theta}) \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{2.11}$$

where $\Sigma = PD(\theta_0)J\Sigma^{(1)}J'D(\theta_0)P$, $\Sigma^{(1)}$ is defined in (2.4), $P = I - C(C'C)^{-1}C'$, and

$$J = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{pmatrix}_{k\times(k-1)} . \tag{2.12}$$

Theorem 2 *If (2.4) and the conditions of Theorem 1 hold, then*

1. *$rank(\Sigma) = k - q - 1$.*
2. *Let $\hat{\Sigma}^{(1)}$ be a consistent estimator of $\Sigma^{(1)}$ and let $\hat{\Sigma}$ be obtained by replacing $\boldsymbol{p}(\theta_0)$, $d\boldsymbol{p}(\theta_0)/d\theta'$, $D(\theta_0)$ and $\Sigma^{(1)}$ with $\boldsymbol{p}^{(n)}(\hat{\theta})$, $d\boldsymbol{p}^{(n)}(\hat{\theta})/d\theta'$, $D_n(\hat{\theta})$ and $\hat{\Sigma}^{(1)}$ respectively in $\Sigma = PD(\theta_0)J\Sigma^{(1)}J'D(\theta_0)P$. Then the consistent estimator $\hat{\Sigma}$ satisfies*

$$rank\,(\hat{\Sigma}) \xrightarrow{P} k - q - 1.$$

Let $\hat{\Sigma}^{\dagger}$ and $\Sigma^{\dagger}$ denote the Moore-Penrose inverses of $\hat{\Sigma}$ and Σ respectively. We shall see in Section 5 that Theorem 2 implies that $\hat{\Sigma}^{\dagger} \xrightarrow{P} \Sigma^{\dagger}$, and this fact underlies the simple asymptotic distribution of our test statistic.

Theorem 3 *Define $Q = \boldsymbol{\xi}_n'(\hat{\theta})\, \hat{\Sigma}^{\dagger}\, \boldsymbol{\xi}_n(\hat{\theta})$. Then, under H_0,*

$$Q \xrightarrow{d} \chi^2_{k-q-1} \quad as \quad n \to \infty.$$

This theorem enables us to test the null hypothesis and obtain p-values in the usual way.

2.3 Choice of the Quadratic Form

The procedure described above produces a class of test statistics based on different choices of the matrix $D(\theta, \boldsymbol{t})$ used in (2.7). Following are some important examples that may be used in practice.

Example 1 Take $D(\theta, \boldsymbol{t}) = I_k$ to be the identity matrix. Then, regularity conditions R1–R3 are satisfied and the test statistic is formed with $\boldsymbol{\xi}_n(\theta) = n^{\frac{1}{2}}(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta))$.

Example 2 Take

$$D(\theta) = \operatorname{diag}\Big((p_1(\theta))^{-\frac{1}{2}}, \ldots, (p_k(\theta))^{-\frac{1}{2}}\Big). \tag{2.13}$$

(More precisely, $D(\theta, \boldsymbol{t}) = \operatorname{diag}((F_\theta(t_1))^{-\frac{1}{2}}, \ldots, (1 - F_\theta(t_{k-1}))^{-\frac{1}{2}})$). Then, regularity conditions R1–R3 are satisfied. The test statistic is thus formed with

$$\boldsymbol{\xi}_n(\theta) = \Big(\frac{n\hat{p}_1 - np_1^{(n)}(\theta)}{(np_1^{(n)}(\theta))^{\frac{1}{2}}}, \ldots, \frac{n\hat{p}_k - np_k^{(n)}(\theta)}{(np_k^{(n)}(\theta))^{\frac{1}{2}}}\Big),$$

which is the vector used in the classical i.i.d. Pearson-Fisher setting.

The main advantage of the above examples is their simplicity. We need only minimize a simple quadratic form to obtain an estimate of θ. However, in general there is nothing special about the matrix (2.13). The motivation for its use is that it is the square root of a generalized inverse of the multinomial covariance matrix. In more detail, we have in the classical i.i.d. case (with fixed cells)

$$\boldsymbol{\zeta}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, M(\theta_0)),$$

where $M(\theta_0) = \operatorname{diag}(p_1(\theta_0), \ldots, p_k(\theta_0)) - \boldsymbol{p}(\theta_0)\boldsymbol{p}'(\theta_0)$. Recall that if A is an arbitrary matrix, then a generalized inverse is any matrix A^- that satisfies $AA^-A = A$. It is easy to see that for $D(\theta)$ given by (2.13), $D^2(\theta_0)$ is a generalized inverse of $M(\theta_0)$ (we omit the arguments θ_0 and $\boldsymbol{a}$ for convenience):

$$\begin{aligned} MD^2M &= (\operatorname{diag}(p_1, \ldots, p_k) - \boldsymbol{p}\boldsymbol{p}')\operatorname{diag}\Big(\frac{1}{p_1}, \ldots, \frac{1}{p_k}\Big)(\operatorname{diag}(p_1, \ldots, p_k) - \boldsymbol{p}\boldsymbol{p}') \\ &= (I - \boldsymbol{p}\mathbf{1}')(\operatorname{diag}(p_1, \ldots, p_k) - \boldsymbol{p}\boldsymbol{p}') \\ &= \operatorname{diag}(p_1, \ldots, p_k) - \boldsymbol{p}\boldsymbol{p}' - \boldsymbol{p}\boldsymbol{p}' + \boldsymbol{p}\boldsymbol{p}' \\ &= \operatorname{diag}(p_1, \ldots, p_k) - \boldsymbol{p}\boldsymbol{p}' \\ &= M. \end{aligned}$$

(Of course, this was not the motivation used by Pearson and Fisher.)

Let us now turn to the general case and to make our explanations simpler, we temporarily continue to consider only the case of fixed cells. Let $\boldsymbol{\eta}_n^{(1)}$ be the vector of length $k-1$ defined by

$$\boldsymbol{\eta}_n^{(1)} = n^{\frac{1}{2}}(\hat{F}(a_1) - F_{\theta_0}(a_1), \ldots, \hat{F}(a_{k-1}) - F_{\theta_0}(a_{k-1})).$$

Note that $\boldsymbol{\zeta}_n(\theta_0) = J\boldsymbol{\eta}_n^{(1)}$, and that $\operatorname{rank}(J) = k-1$ (the matrix formed by the first $k-1$ rows of J is lower triangular, and so its determinant is easily seen to be 1, which implies that $\text{rank}(J) \geq k-1$; since J is a $k \times (k-1)$ matrix, its rank is at most $k-1$). Thus, by (2.1) and (2.3), we have

$$\boldsymbol{\zeta}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, M(\theta_0)) \quad \text{where } M(\theta) = J\Sigma^{(1)}(\theta, \boldsymbol{a})J'. \tag{2.14}$$

Thus, it seems natural to use for the matrix $D(\theta)$ a square root of a generalized inverse of $M(\theta)$. In fact, the quadratic form $\boldsymbol{\zeta}'(\theta)M^-(\theta)\boldsymbol{\zeta}(\theta)$ is invariant under the choice of $M^-(\theta)$, since $\boldsymbol{\zeta}(\theta)$ is in the space spanned by the columns of $M(\theta)$. This is because $M'(\theta)1 = J\Sigma^{(1)}(\theta)J'1 = J\Sigma^{(1)}(\theta)0 = 0$; furthermore, $\boldsymbol{\zeta}'(\theta)1 = 0$. Since $\text{rank}(M(\theta)) = k-1$ by (2.4), we see that $\boldsymbol{\zeta}(\theta)$ is in the space spanned by the columns of $M(\theta)$. Thus, by Lemma 5.1 in Section 5, the quadratic form $\boldsymbol{\zeta}'(\theta)M^-(\theta)\boldsymbol{\zeta}(\theta)$ is uniquely defined.

In the case of random cells, (2.14) is still true; see (5.5) in the proof of Theorem 1.

In the development of our theory, we need the matrix $D(\cdot,\cdot)$ to satisfy Conditions R1–R3, essentially that $D^2(\cdot,\cdot)$ be an *invertible* generalized inverse of M, which also has some *continuity and differentiability properties in θ and $\boldsymbol{t}$*. We produce such a $D(\cdot,\cdot)$ in the next example.

Example 3 Let $M(\theta,\boldsymbol{t}) = J\Sigma^{(1)}(\theta,\boldsymbol{t})J'$. Since $\text{rank}(J) = k-1$, the $(k-1)\times(k-1)$ matrix $J'J$ is invertible, and so we may write

$$M(\theta,\boldsymbol{t}) = \left[J(J'J)^{-\frac{1}{2}}\right]\left[(J'J)^{\frac{1}{2}}\Sigma^{(1)}(\theta,\boldsymbol{t})(J'J)^{\frac{1}{2}}\right]\left[(J'J)^{-\frac{1}{2}}J'\right] = GA(\theta,\boldsymbol{t})G', \tag{2.15}$$

where G and $A(\theta,\boldsymbol{t})$ denote the matrices appearing in the first and second sets of brackets, respectively. Then $H_{k\times k} = [G, 1/\sqrt{k}]$ is an orthogonal matrix since

$$H'H = \begin{pmatrix} G'G & G'1/\sqrt{k} \\ 1'G/\sqrt{k} & 1'1/k \end{pmatrix} = \begin{pmatrix} I_{k-1} & 0 \\ 0' & 1 \end{pmatrix} = I_k.$$

Moreover,

$$M(\theta,\boldsymbol{t}) = GA(\theta,\boldsymbol{t})G' = H\begin{pmatrix} A(\theta,\boldsymbol{t}) & 0 \\ 0' & 0 \end{pmatrix}H'. \tag{2.16}$$

Since $A(\theta,\boldsymbol{t})$ is invertible, we may define

$$D(\theta,\boldsymbol{t}) = H\begin{pmatrix} A^{-\frac{1}{2}}(\theta,\boldsymbol{t}) & 0 \\ 0' & 1 \end{pmatrix}H'. \tag{2.17}$$

(Strictly speaking, we take $A^{\frac{1}{2}}(\theta,\boldsymbol{t})$ to be a positive definite square root of $A(\theta,\boldsymbol{t})$ to ensure continuity). Then, regularity conditions R1–R3 reduce to (mild) regularity conditions on $\Sigma^{(1)}(\theta,\boldsymbol{t})$. We also see that $D^2(\theta,\boldsymbol{t})$ is a generalized inverse of $M(\theta,\boldsymbol{t})$ since

$$MD^2M = H\begin{pmatrix} A & 0 \\ 0' & 0 \end{pmatrix}H'H\begin{pmatrix} A^{-1} & 0 \\ 0' & 1 \end{pmatrix}H'H\begin{pmatrix} A & 0 \\ 0' & 0 \end{pmatrix}H' = H\begin{pmatrix} A & 0 \\ 0' & 0 \end{pmatrix}H' = M$$

(here we have dropped the notation $(\theta,\boldsymbol{t})$ for brevity). Hence, we estimate θ_0 by the parameter value that minimizes

$$\boldsymbol{\xi}_n'(\theta)\boldsymbol{\xi}_n(\theta) = \boldsymbol{\zeta}_n'(\theta)D^2(\theta,\boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\theta) = \boldsymbol{\zeta}_n'(\theta)M^-(\theta,\boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\theta),$$

and construct a test statistic as in Section 2.2.

We know by Theorem 3 that for this special choice of D the test statistic has an asymptotic χ^2_{k-q-1} distribution. The next result states that for this D, the test statistic essentially reduces to a simpler quadratic form.

Proposition 2.1 *Let $\hat{\theta}$ be the parameter value that minimizes the (well-defined) quadratic form $\boldsymbol{\zeta}_n'(\theta)M^-(\theta,\boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\theta)$. Let $D(\theta,\boldsymbol{t})$ be given by (2.17), let Q be defined as in Theorem 3, and let*

$$\tilde{Q} = \boldsymbol{\zeta}_n'(\hat{\theta})M^-(\hat{\theta},\boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\hat{\theta}).$$

Assume (2.1), A1 and A2, and that $\Sigma^{(1)}(\cdot,\cdot)$ and $\partial\Sigma^{(1)}(\cdot,\cdot)/\partial\theta$ are continuous at $(\theta_0,\boldsymbol{a})$, and $\Sigma^{(1)}(\cdot,\cdot)$ is bounded in $\Theta \times N(\boldsymbol{a})$, where $N(\boldsymbol{a})$ is a neighborhood of $\boldsymbol{a}$ in $\mathbb{R}^{k-1}$. Then $D(\theta,\boldsymbol{t})$ satisfies R1–R3, and Q satisfies

$$Q = \tilde{Q} + O_p(n^{-1}) \quad \text{under } H_0. \tag{2.18}$$

In particular,

$$\tilde{Q} \xrightarrow{d} \chi^2_{k-q-1} \quad \text{as} \quad n \to \infty. \tag{2.19}$$

Remark 2.2 *One might suspect that $Q = \tilde{Q}$; this is not the case, however, and the $O_p(n^{-1})$ term in (2.18) is really needed.*

Remark 2.3 *We have not been able to prove (2.19) directly. The only way that we know of for obtaining (2.19) is to apply (2.18) together with Theorem 3.*

Remark 2.4 *The choice of $D(\cdot,\cdot)$ is clearly an interesting and important problem. One can conjecture that choosing $D(\cdot,\cdot)$ to be the square root of a generalized inverse of $M(\cdot,\cdot)$ will lead to some optimality properties. We have not investigated this question.*

3 Applications to Commonly Arising Models

We first give a result that shows that the nonsingularity assumption (2.4) is satisfied for a class of models arising in survival analysis.

Proposition 3.1 *If the process W in (2.1) has the form $W \stackrel{d}{=} \bar{F} \cdot V$ where $\bar{F} = 1 - F$ and V is a Gaussian martingale for which the variance function $v(t) = Var(V(t))$ satisfies $[v(a_i) - v(a_{i-1})] > 0$ for all i, then (2.4) is satisfied.*

Proof Note that the $(i,j)^{th}$ element of $\Sigma^{(1)}$ is $\sigma_{ij}^{(1)} = \bar{F}_{\theta_0}(a_i)\bar{F}_{\theta_0}(a_j)v(\min(a_i,a_j))$. We will show that

$$\det(\Sigma^{(1)}) = \prod_{i=1}^{k-1} \left(\bar{F}_{\theta_0}(a_i)\right)^2 \left(v(a_i) - v(a_{i-1})\right) > 0. \tag{3.1}$$

To see (3.1), we consider the linear operator $\mathcal{L}_i$ acting on $(k-1)\times(k-1)$ matrices which adds $-\frac{\bar{F}_{\theta_0}(a_i)}{\bar{F}_{\theta_0}(a_{i-1})} \times$ row $(i-1)$ to row i. This operator does not change the determinant of a matrix. Applying $\mathcal{L}_{k-1}, \mathcal{L}_{k-2}, \ldots, \mathcal{L}_2$ successively to $\Sigma^{(1)}$ reduces $\Sigma^{(1)}$ to an upper triangular matrix with diagonal elements $[\bar{F}_{\theta_0}(a_i)]^2[v(a_i) - v(a_{i-1})]$, $i = 1, \ldots, k-1$. This implies (3.1) immediately.

3.1 The Pearson-Fisher Test, with Random Cells

Let $X_1, \ldots, X_n$ be i.i.d. $\sim F$ and let $\hat{F}$ be the empirical distribution function. Then (2.1) holds with $W \stackrel{d}{=} \bar{F} \cdot V$ where V is the Gaussian martingale with variance function $v(t) = F(t)/\bar{F}(t)$. As explained in Section 2.3, $M(\theta, \boldsymbol{a}^{(n)}) = J\Sigma^{(1)}(\theta, \boldsymbol{a}^{(n)})J' = \text{diag}(p_1^{(n)}(\theta), \ldots, p_k^{(n)}(\theta)) - \boldsymbol{p}^{(n)}(\theta)\boldsymbol{p}^{(n)\prime}(\theta)$, for which $\text{diag}(1/p_1^{(n)}(\theta), \ldots, 1/p_k^{(n)}(\theta))$ is a generalized inverse. Thus, the chi-square statistic $\tilde{Q}$ constructed in Proposition 2.1 coincides exactly with the classical Pearson-Fisher statistic, except that the cells may be random.

For the development of Pearson-Fisher test with random cells, see e.g. Čebyšev (1971), Moore (1971), Moore and Spruill (1975), and Pollard (1979).

3.2 Chi-Square Tests for Left-Truncated Data

Let (X, Y) be a pair of independent nonnegative random variables with distribution functions F and G respectively. Random left-truncated data consists of n i.i.d. draws, $(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)$, from the conditional distribution of (X, Y), given that $Y < X$. Here, X is called the random variable of interest and Y is called the truncation variable, and the objective is to make inference on F. Left truncation arises when individuals come under observation only some known time after the natural time origin of the phenomenon under study. That is, for any given individual, had failure occurred before the truncation variable in question, variables pertaining to that individual would not have been recorded.

This kind of data arises frequently in medical survival studies when one wants to study the length of survival after the start of the disease: If X denotes the time elapsed between the onset of the disease and death, and if the followup period starts Y units of time after the onset of the disease, then clearly X is left truncated by Y.

Certain studies on AIDS give rise to a slightly different form of the random truncation model, and in Section 4 we illustrate the methods of this paper in an analysis of a data set from the Center for Disease Control (CDC) that is used to study the latency of the AIDS virus.

Random truncation models arise also in fields other than survival analysis. For a general overview of the model and references to the literature see Woodroofe (1985).

The Product Limit Estimate and its Asymptotics Nonparametric estimation of F based on left-truncated data was first studied by Lynden-Bell (1971) who proposed the product-limit estimate (PLE) $\hat{F}$ given by

$$1 - \hat{F}(x) = \prod_{l=1}^{n} \left(1 - \frac{I(X_l^* \leq x)}{J(X_l^*)}\right) \tag{3.2}$$

where $J(t) \equiv \sum_{l=1}^{n} I(Y_l^* < t \leq X_l^*)$. Keiding and Gill (1990) have shown that this estimator is the nonparametric maximum likelihood estimate of F for the model in which F and G are completely unknown, provided that $J(X_{(l)}^*) > 1$ for $l = 1, \ldots, n-1$.

Weak convergence results for the PLE were later established by Woodroofe (1985) and Keiding and Gill (1990). Here we follow the notation of Keiding and Gill (1990). For convenience, we assume that F and G are continuous, and that ess $\sup(F) = \infty$, ess $\inf(G) = 0$, and $\alpha \equiv P(Y < X) > 0$.

Weak convergence of the process $n^{\frac{1}{2}}(\hat{F}-F)$ involves delicate problems near 0, and to obtain weak convergence in $D[0,\infty]$ one must impose the condition

$$\int_0^\infty \frac{dF(t)}{G(t)} < \infty \tag{3.3}$$

(see Section 5 of Woodroofe (1985) and Section 5.2 of Keiding and Gill (1990)). This is a rather restrictive condition and for that reason Keiding and Gill (1990) consider, for fixed $\epsilon > 0$, the process $n^{\frac{1}{2}}(\hat{F}^\epsilon - F^\epsilon)$ where $1-\hat{F}^\epsilon(t) = (1-\hat{F}(t))/(1-\hat{F}(\epsilon))$ and $1-F^\epsilon(t) = (1-F(t))/(1-F(\epsilon))$ for $t \geq \epsilon$ (i.e. $\hat{F}^\epsilon(t) = P_{\hat{F}}(X \leq t \,|\, X > \epsilon)$ in an obvious notation, and similarly for $F^\epsilon(t)$). They prove weak convergence of this process in $D[\epsilon,\infty]$.

Theorem 3.1 (Keiding and Gill, 1990) *Let $\epsilon > 0$, and assume that for all $t \geq \epsilon$ $P(Y < t \leq X \mid Y < X) > 0$. Then*

$$n^{\frac{1}{2}}(\hat{F}^\epsilon - F^\epsilon) \xrightarrow{d} \bar{F}^\epsilon \cdot V \quad in \quad D[\epsilon,\infty], \tag{3.4}$$

where $V(\cdot)$ is a continuous Gaussian martingale with zero mean and variance function

$$v(t) = \alpha \int_\epsilon^t \frac{dF(u)}{G(u)[1-F(u)]^2}.$$

A consistent estimator of $v(t)$ for $t \geq \epsilon$ is

$$\hat{v}(t) = \int_\epsilon^t nJ(s)^{-2} dN(s),$$

where $N(t) = \sum_{l=1}^n I(X_l^* \leq t)$.

The Chi-Square Test Statistics As in Section 2, we first form a data dependent partition. Then we choose a symmetric matrix $D(\theta, \boldsymbol{t})$ satisfying R1–R3 and construct $\boldsymbol{\zeta}_n(\theta)$ and $\boldsymbol{\xi}_n(\theta)$ from F_θ^ϵ and $\hat{F}^\epsilon$. We estimate the true parameter θ_0 by the value $\hat{\theta}$ of θ which minimizes $\boldsymbol{\xi}_n'(\theta)\boldsymbol{\xi}_n(\theta)$ and estimate v by the $\hat{v}$ defined in Theorem 3.1. Then the test statistic $Q = \boldsymbol{\xi}_n'(\hat{\theta})\hat{\Sigma}^\dagger\boldsymbol{\xi}_n(\hat{\theta})$ obtained from $\hat{F}^\epsilon$, $\hat{\theta}$, $F_{\hat{\theta}}^\epsilon$, and $\hat{v}(t)$ has limiting null distribution χ^2_{k-q-1}.

Remark 3.1 *In Section 2.3 we mentioned that it may be natural to use for the matrix $D(\cdot,\cdot)$ the square root of a generalized inverse of $M(\cdot,\cdot) = J\Sigma^{(1)}(\cdot,\cdot)J'$, for example the $D(\cdot,\cdot)$ given by (2.17). This brings up a computational problem: To minimize $\boldsymbol{\zeta}_n(\theta)'D^2(\theta,\boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\theta)$ we need to find a formula for $D^2(\theta,\boldsymbol{a}^{(n)})$. This forces us to do a symbolic inversion of a matrix, which requires a symbolic manipulations program. An alternative is to replace the matrix $D(\theta,\boldsymbol{a}^{(n)})$ given by (2.17) with a consistent estimate $\hat{D}$ of $D(\theta_0,\boldsymbol{a})$. We then need only to do the matrix inversion numerically, i.e. do it just once. Remark A.1 in the Appendix establishes the validity of this procedure.*

Remark 3.2 *To carry out the test one must decide on a value for ϵ. Intuitively, the smaller the value of ϵ, the smaller the information loss in the left tail $(0,\epsilon]$. On the other hand, if condition (3.3) is not satisfied, a smaller ϵ requires a larger n for the asymptotics to set in. In practice, one selects ϵ such that only a small proportion of the X_l^*'s fall in the tail $(0,\epsilon]$, and this choice is made subjectively.*

3.3 Chi-Square Tests for Right-Censored Data

We first review the random censorship model of survival analysis. The pairs of positive random variables (X_l, Y_l), $l = 1, \ldots, n$, are independent and identically distributed, with distribution functions $F(t) = P(X_l \leq t)$ and $G(t) = P(Y_l \leq t)$ and the Y's are independent of the X's. We observe only (Z_l, δ_l), $l = 1, \ldots, n$, where $Z_l = \min(X_l, Y_l)$ and $\delta_l = I(X_l \leq Y_l)$. The X's represent survival times, the Y's represent censoring times, and the problem is to estimate F. The most commonly used estimate of F is the Kaplan-Meier estimator defined by

$$\hat{F}(t) = 1 - \prod_{Z_{(l)} \leq t} \left(\frac{n-l}{n-l+1} \right)^{\delta_{(l)}} \tag{3.5}$$

where $Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)}$ denote the ordered values of $Z_1, Z_2, \ldots, Z_n$ and $\delta_{(l)}$ is the δ corresponding to $Z_{(l)}$. Weak convergence of $n^{\frac{1}{2}}(\hat{F} - F)$ is well known. The following proposition is a special case of Theorem 4.2.2 of Gill (1980).

Theorem 3.2 *Assume that F is continuous and let τ be such that $F(\tau) < 1$ and $G(\tau) < 1$. Then,*

$$n^{\frac{1}{2}}(\hat{F} - F) \xrightarrow{d} \bar{F} \cdot V \text{ in } D[0, \tau],$$

where V is a zero-mean Gaussian martingale with variance function

$$v(t) = \int_0^t \frac{dF(s)}{\bar{F}(s-)\bar{F}(s)\bar{G}(s-)},$$

for which a consistent estimator is

$$\hat{v}(t) = \int_0^t \frac{d\hat{F}(s)}{\hat{\bar{F}}(s-)\hat{\bar{F}}(s)\hat{\bar{G}}(s-)}.$$

Here, $\hat{G}$ is the Kaplan-Meier estimate of G, i.e. the right side of (3.5), except that $\delta_{(l)}$ is replaced by $1 - \delta_{(l)}$; also, for a function g, $g(s-) = \lim_{u \uparrow s} g(u)$.

We construct our test statistic as prescribed in Section 2.

Chi-square goodness-of-fit tests for the random censorship model were investigated by Habib and Thomas (1986), who estimated θ_0 by the maximum likelihood estimate. In a paper that provided the impetus for the present work, Hollander and Peña (1990) developed a chi-square goodness-of-fit test for the case of a simple null hypothesis. If we use fixed cells and take D to be the diagonal matrix (2.13), then our procedure is identical to theirs. Actually, it is not very difficult to see that the test statistic is invariant under the choice of the matrix D, so that nothing can be gained by using the matrix (2.17). This is true only in the case of a simple null hypothesis, for which the parameter θ_0 does not need to be estimated.

Actually Theorem 4.2.2 of Gill (1980) gives a weak convergence result for $n^{\frac{1}{2}}(\hat{F} - F)$ for a wide class of censoring mechanisms which includes the random censoring mechanism discussed above but also fixed censoring and Type II censoring. Therefore, our procedure is applicable to data subject to these censoring mechanisms as well. For Type II censoring, we refer to the paper by Mihalko and Moore (1980). We mention briefly that they studied a number of chi-square tests and that our procedure reproduces one of their tests.

The special structure $W \stackrel{d}{=} \bar{F} \cdot V$ (cf. Proposition 3.1) is by no means necessary for (2.4) to be satisfied. For example, in fitting a Cox model to survival data, one is sometimes interested in investigating a parametric model for the baseline survival function. If $\hat{S}$ is the "Nelson-Aalen estimator" of the baseline survival function S_0, then $n^{\frac{1}{2}}(\hat{S} - S_0)$ converges to a limiting Gaussian process W which does *not* have the form $W \stackrel{d}{=} S_0 \cdot V$ for some martingale V (see Theorem 3.4 of Andersen and Gill (1982)). It is easy to show, however, that (2.4) is satisfied. Goodness-of-fit tests for the parametric Cox model have also been studied by Hjort (1990, Section 6).

4 Analysis of Transfusion-Related AIDS Infection Data

An important problem in studies of acquired immune deficiency syndrome (AIDS) is to determine the distribution of the length of the "incubation period" (i.e. the time from the human immunodeficiency virus (HIV) infection to the diagnosis of AIDS). This problem is difficult because one generally does not have accurate information on the date of HIV infection. Nevertheless the date of infection can be ascertained for patients who are thought to be infected with HIV by blood or blood product transfusion. Tables 2–4 on pages 745–746 of Wang (1989) give transfusion-related AIDS data reported by the Centers for Disease Control (CDC) in Atlanta, Georgia. These data consist of 295 cases diagnosed with AIDS prior to July 1, 1986, and for which infection could be attributed to a single transfusion or short series of transfusions. The tables report the incubation time X (in months), the time Y (in months) from the HIV infection to the end of the study (July 1, 1986), and the individual's age at the time of transfusion. Because disease resistance depends on age, the data are divided into three groups: 34 "children" aged 1-4, 120 "adults" aged 5-59, and 141 "elderly patients" aged 60 and older. Obviously the data for the incubation time X are right truncated by Y since patients who had HIV infection prior to July 1, 1986 but developed AIDS after July 1, 1986 were not included in the data. That is, we observe (Y, X) only if $X < Y$.

Let F and G be the distributions of X and Y respectively. Then, F may be estimated by (3.2). Recall that in the random censorship model of survival analysis one can show that the Kaplan-Meier estimator (3.5) is the nonparametric maximum likelihood estimator of F whether or not we have any knowledge of G. The situation is different for the random truncation model: The product-limit estimator (3.2) is the nonparametric maximum likelihood estimator of F only in the model where G is completely unspecified. If G is completely or partially specified, (3.2) is no longer the maximum likelihood estimate, and Wang (1989) has shown that knowledge that G belongs to a parametric family can be exploited to obtain a more efficient estimate of F. Thus it is important to be able to determine if a given parametric model holds. The CDC's AIDS data have been studied through various parametric models by several authors; see Kalbfleisch and Lawless (1989) and the references therein. However, there has not yet been a formal test of fit to determine whether the parametric assumptions are appropriate. The chi-square test developed in this paper provides a straightforward way of checking the parametric assumptions on both F and G.

Here we shall be concerned only with the parametric assumptions on G. Note that one can regard Y as being left truncated by X. Therefore the chi-square test developed in Section 3.2 can be applied directly. Table 2 below shows the analysis we have done on the AIDS data. For each age group we tested the null hypothesis that G is a Weibull distribution with fixed shape parameter ν and unknown scale parameter θ, i.e. $G(t) = 1 - \exp(-\theta t^{\nu})$ for some θ. The values of ν were taken to be the 10 values indicated in Table 2, and three cells were used for each of the tests. We used the test statistic described in Section 3.2, and we took the value of ϵ to be the .07, .06, and .03 quantile of the product-limit estimator for the children, adults, and elderly patients, respectively (larger values of ϵ are required by the asymptotic theory if the sample size is smaller). The cell boundaries for the children and adults groups were taken to be ϵ, 30, 60, and ∞, and for the elderly patients these were taken to be ϵ, 30, 40, and ∞. The values of the chi-square test statistic $\tilde{Q}$ and corresponding p-values are reported for each combination of the null hypothesis and the age group.

Table 2 *Test Statistics $\tilde{Q}$ and p-values for the AIDS data. Null hypothesis is that G is a Weibull distribution with an arbitrary scale parameter and fixed shape parameter ν.*

	ν	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.50	2.75	2.85
Children	$\tilde{Q}$	10.16	6.85	4.34	2.50	1.22	0.42	0.041	0.42	1.13	1.52
	p-value	<0.01	0.01	0.04	0.11	0.27	0.52	0.84	0.52	0.28	0.22
Adults	$\tilde{Q}$	2.07	0.97	0.27	0.003	0.18	0.81	1.92	5.69	8.42	9.69
	p-value	0.15	0.33	0.60	0.95	0.68	0.37	0.17	0.02	<0.01	<0.01
Elderly Patients	$\tilde{Q}$	4.33	3.23	2.30	1.52	0.91	0.45	0.16	0.03	0.19	0.29
	p-value	0.04	0.07	0.13	0.22	0.34	0.50	0.69	0.87	0.67	0.59

In her analysis of this data set, Wang (1989) made the assumption that G is exponential (this is Weibull with with $\nu = 1$), citing an informal analysis to support this. For the children group, Table 2 shows that our test provides evidence against this assumption (p-value = .04) and instead suggests the Weibull family with shape parameter $\nu = 2$ as a reasonable model. For the adults group the table shows that the exponential distribution is adequate, while for the elderly patients the table indicates that the exponential assumption is suspect and that the Weibull family with shape parameter $\nu = 2.5$ provides a better fit.

5 Proofs of Main Results

Proof of Theorem 1

We first show that under H_0

$$\boldsymbol{\xi}_n(\theta_0) \xrightarrow{d} \mathcal{N}_k(0, D(\theta_0)J\Sigma^{(1)}J'D(\theta_0)) \quad \text{as } n \to \infty, \tag{5.1}$$

where J and $\Sigma^{(1)}$ are given by (2.12) and (2.3) and (2.4), respectively. Define

$$W_n(\cdot) = n^{\frac{1}{2}} \left(\hat{F}(\cdot) - F_{\theta_0}(\cdot)\right). \tag{5.2}$$

Then

$$\boldsymbol{\zeta}_n(\theta_0) = n^{\frac{1}{2}}(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta_0)) = J\boldsymbol{\eta}_n, \tag{5.3}$$

where $\boldsymbol{\eta}_n = (W_n(a_1^{(n)}), \ldots, W_n(a_{k-1}^{(n)}))'$. Write $\boldsymbol{\eta}_n = \boldsymbol{\eta}_n^{(1)} + \boldsymbol{\eta}_n^{(2)}$, where $\boldsymbol{\eta}_n^{(1)} = (W_n(a_1), \ldots, W_n(a_{k-1}))'$ and $\boldsymbol{\eta}_n^{(2)} = (W_n(a_1^{(n)}) - W_n(a_1), \ldots, W_n(a_{k-1}^{(n)}) - W_n(a_{k-1}))'$. We shall show that $\boldsymbol{\eta}_n^{(1)} \xrightarrow{d} \mathcal{N}_{k-1}(0, \Sigma^{(1)})$ and $\boldsymbol{\eta}_n^{(2)} \xrightarrow{P} 0$.

From (2.1), $W_n \xrightarrow{d} W$, where W is a continuous Gaussian process with zero mean. The weak convergence result for $\boldsymbol{\eta}_n^{(1)}$ follows immediately since the finite dimensional distributions of $W_n(\cdot)$ converge to a multinormal distribution. To prove that $\boldsymbol{\eta}_n^{(2)} \xrightarrow{P} 0$, we use a standard Skorohod construction (see Item 3.1.1 in Skorohod, 1956) to obtain random elements $\tilde{W}_n$ and $\tilde{W}$ on a new probability space, such that $\tilde{W}_n \stackrel{d}{=} W_n$, $\tilde{W} \stackrel{d}{=} W$, and $\tilde{W}_n \xrightarrow{a.s.} \tilde{W}$ in $D[\epsilon, M]$. Since $\tilde{W}$ has continuous sample paths, this implies

$$\sup_{\epsilon \le t \le M} |\tilde{W}_n(t) - \tilde{W}(t)| \xrightarrow{a.s.} 0. \tag{5.4}$$

For $i = 1, \ldots, k-1$, define $\tilde{a}_i^{(n)} = a_i(F + n^{-\frac{1}{2}}\tilde{W}_n)$. Because $\tilde{a}_i^{(n)} \stackrel{d}{=} a_i^{(n)}$ and $a_i^{(n)} \xrightarrow{P} a_i$, we have $\tilde{a}_i^{(n)} \xrightarrow{P} a_i$. So for large n,

$$\begin{aligned}
|\tilde{W}_n(\tilde{a}_i^{(n)}) - \tilde{W}_n(a_i)| &\le |\tilde{W}_n(\tilde{a}_i^{(n)}) - \tilde{W}(\tilde{a}_i^{(n)})| + |\tilde{W}(\tilde{a}_i^{(n)}) - \tilde{W}(a_i)| + |\tilde{W}(a_i) - \tilde{W}_n(a_i)| \\
&\le 2 \sup_{\epsilon \le t \le M} |\tilde{W}_n(t) - \tilde{W}(t)| + |\tilde{W}(\tilde{a}_i^{(n)}) - \tilde{W}(a_i)| \xrightarrow{P} 0
\end{aligned}$$

by (5.4) and the fact that $\tilde{W}$ has continuous paths. Since $\tilde{W}_n(\tilde{a}_i^{(n)}) - \tilde{W}_n(a_i) \stackrel{d}{=} W_n(a_i^{(n)}) - W_n(a_i)$, we conclude that $\boldsymbol{\eta}_n^{(2)} \xrightarrow{P} 0$. Thus $\boldsymbol{\eta}_n \xrightarrow{d} \mathcal{N}_{k-1}(0, \Sigma^{(1)})$, and (5.3) and (2.7) give

$$\boldsymbol{\zeta}_n(\theta_0) \xrightarrow{d} \mathcal{N}_k(0, J\Sigma^{(1)}J') \quad \text{and} \quad \boldsymbol{\xi}_n(\theta_0) \xrightarrow{d} \mathcal{N}_k(0, D(\theta_0)J\Sigma^{(1)}J'D(\theta_0)). \tag{5.5}$$

Now we are ready to obtain the weak convergence result for $\boldsymbol{\xi}_n(\hat{\theta})$. Note that

$$\begin{aligned}
\boldsymbol{\xi}_n(\hat{\theta}) &= D_n(\hat{\theta})n^{\frac{1}{2}}(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\hat{\theta})) \\
&= D_n(\hat{\theta})n^{\frac{1}{2}}(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta_0)) - D_n(\hat{\theta})n^{\frac{1}{2}}(\boldsymbol{p}^{(n)}(\hat{\theta})) - \boldsymbol{p}^{(n)}(\theta_0)) \\
&= (D_n(\theta_0) + o_p(1))\boldsymbol{\zeta}_n(\theta_0) - (D(\theta_0) + o_p(1))n^{\frac{1}{2}}\left(\frac{\partial \boldsymbol{p}(\theta_0)}{\partial \theta'} + o_p(1)\right)(\hat{\theta} - \theta_0) \\
&= D_n(\theta_0)\boldsymbol{\zeta}_n(\theta_0) - Cn^{\frac{1}{2}}(\hat{\theta} - \theta_0) + o_p(1) \\
&= \boldsymbol{\xi}_n(\theta_0) - C\left((C'C)^{-1}C'\boldsymbol{\xi}_n(\theta_0) + o_p(1)\right) + o_p(1) \\
&= P\boldsymbol{\xi}_n(\theta_0) + o_p(1),
\end{aligned}$$

where in the fourth equality we have used the fact that $\boldsymbol{\zeta}_n(\theta_0) = D^{-1}(\theta_0)\boldsymbol{\xi}_n(\theta_0)$ and $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ are bounded in probability, by (2.8). Therefore

$$\boldsymbol{\xi}_n(\hat{\theta}) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where $\Sigma = PD(\theta_0)J\Sigma^{(1)}J'D(\theta_0)P$.

Proof of Theorem 2

We first prove Part 1 of the theorem. Because $\Sigma^{(1)}$ is assumed to be positive definite, there exists a nonsingular matrix T such that $\Sigma^{(1)} = TT'$, so that

$$\Sigma = (PDJT)(PDJT)'.$$

Thus

$$\text{rank}(\Sigma) = \text{rank}(PDJT) = \text{rank}(PDJ),$$

and we wish to show that

$$\text{rank}(PDJ) = k - q - 1.$$

If A is a matrix, we will let $\mathcal{M}(A)$ denote the space spanned by the column vectors of A and let $\mathcal{M}^+(A)$ be the space of all vectors orthogonal to $\mathcal{M}(A)$. Let $\boldsymbol{d} = D^{-1}1$. Then

$$C'\boldsymbol{d} = \left(\frac{\partial \boldsymbol{p}}{\partial \theta'}\right)' DD^{-1}1 = \left(1'\frac{\partial \boldsymbol{p}}{\partial \theta'}\right)' = \left(\frac{\partial(\sum p_i)}{\partial \theta'}\right)' = 0.$$

So $\boldsymbol{d}$ is orthogonal to $\mathcal{M}(C)$, and the space $\mathcal{M}([C, \boldsymbol{d}])$ spanned by the columns of C together with $\boldsymbol{d}$ has dimension $q+1$. We will show that $\mathcal{M}(PDJ) = \mathcal{M}^+([C, \boldsymbol{d}])$. From this we conclude that

$$\text{rank}(PDJ) = k - \text{dimension of } \mathcal{M}([C, \boldsymbol{d}]) = k - q - 1.$$

This is done in two steps: first we show that $\mathcal{M}(PDJ)$ is a subspace of $\mathcal{M}^+([C, \boldsymbol{d}])$, then we prove that $\mathcal{M}(PDJ)$ is identical to $\mathcal{M}^+([C, \boldsymbol{d}])$.

Using the fact that $P = I_k - C(C'C)^{-1}C'$ is a projection onto $\mathcal{M}^+(C)$, we have

$$(PDJ)'C = J'DPC = J'D0 = 0$$

and

$$(PDJ)'\boldsymbol{d} = J'DP\boldsymbol{d} = J'D\boldsymbol{d} = J'DD^{-1}1 = J'1 = 0.$$

So $\mathcal{M}(PDJ)$ is orthogonal to $\mathcal{M}([C, \boldsymbol{d}])$. Therefore $\mathcal{M}(PDJ)$ is a subspace of $\mathcal{M}^+([C, \boldsymbol{d}])$.

On the other hand, for every $\boldsymbol{x} \in \mathcal{M}^+([C, \boldsymbol{d}])$

$$(PDJ)'\boldsymbol{x} = J'DP\boldsymbol{x} = J'D\boldsymbol{x} \neq 0.$$

Here the nonequality in the last step comes from the fact that $J'D$ has exactly one zero eigenvalue and $\boldsymbol{d}$ is the corresponding eigenvector, which is orthogonal to $\boldsymbol{x}$. Therefore $\mathcal{M}(PDJ) = \mathcal{M}^+([C, \boldsymbol{d}])$. (Otherwise, if $\mathcal{M}(PDJ)$ is a strictly smaller subspace of $\mathcal{M}^+([C, \boldsymbol{d}])$, then there is an orthogonal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{k-q-1}$ of $\mathcal{M}^+([C, \boldsymbol{d}])$ such that $\mathcal{M}(PDJ) = \mathcal{M}([\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r])$ for some $r < k-q-1$. But this implies $(PDJ)'\boldsymbol{e}_{k-q-1} = 0$ which is a contradiction.)

To prove Part 2, note that if A is a square matrix whose determinant is not 0, and if A_n is a sequence of matrices such that $A_n \to A$, then $\det(A_n) \to \det(A)$, so that $\text{rank}(A_n) \to \text{rank}(A)$. We observe that $\hat{C} \xrightarrow{P} C$ (see 2.10), and that C has full rank q for all θ, so that $\text{rank}(\hat{C}) \xrightarrow{P} q$. The same argument gives $\text{rank}(D_n(\hat{\theta})) \xrightarrow{P} k$ and

$\text{rank}(\hat{\Sigma}^{(1)}) \xrightarrow{P} k-1$. Note that if $\text{rank}(\hat{C}) = q$, $\text{rank}(D_n(\hat{\theta})) = k$ and $\text{rank}(\hat{\Sigma}^{(1)}) = k-1$, then $\text{rank}(\hat{\Sigma}) = k-q-1$. The proof is identical to the proof of Part 1. Therefore, $P(\text{rank}(\hat{\Sigma}) = k-q-1) \to 1$, as desired.

Proof of Theorem 3

It is a fact that if A_n, $n = 1, 2, \ldots$ and A are matrices such that

$$A_n \to A \tag{5.6}$$

and

$$\text{rank}(A_n) = \text{rank}(A) \quad \text{for all large } n, \tag{5.7}$$

then the Moore-Penrose inverses satisfy

$$A_n^\dagger \to A^\dagger \tag{5.8}$$

(see e.g. Theorem 10.4.1 of Campbell and Meyer (1979)). Thus, by Parts 1 and 2 of Theorem 2 we see that

$$\hat{\Sigma}^\dagger \xrightarrow{P} \Sigma^\dagger. \tag{5.9}$$

Theorem 1 and Part 1 of Theorem 2 imply that

$$\boldsymbol{\xi}_n'(\hat{\theta})\Sigma^\dagger\boldsymbol{\xi}_n(\hat{\theta}) \xrightarrow{d} \chi^2_{k-q-1} \tag{5.10}$$

and now Theorem 3 follows from (5.9) and (5.10).

Remark 5.1 *Condition (5.6) by itself is not enough to guarantee (5.8), and it is necessary to also have (5.7) (unless, of course, A is nonsingular). There are errors on this point in some papers, e.g. Hjort (1990, pp. 1233–1234) and McKeague and Utikal (1991, last paragraph of Section 3, and Section 5); also, Akritas (1988, last paragraph of Section 4.1) makes a statement that ignores this point and so is misleading. These authors base chi-square tests on asymptotic normality results of the form*

$$\boldsymbol{w}_n \xrightarrow{d} \mathcal{N}(0, R) \tag{5.11}$$

where R is a possibly singular matrix. They provide a consistent estimator R_n of R, i.e. one that satisfies

$$R_n \xrightarrow{P} R, \tag{5.12}$$

take a generalized inverse R_n^- of R_n and claim that the test statistic $\boldsymbol{w}_n' R_n^- \boldsymbol{w}_n$, satisfies

$$\boldsymbol{w}_n' R_n^- \boldsymbol{w}_n \xrightarrow{d} \chi^2_{rank(R)}. \tag{5.13}$$

However, (5.11) and (5.12) by themselves do not imply (5.13) and so additional work is needed.

Consider the following simple counterexample. Take the two-dimensional vectors $\boldsymbol{w}_n$ to be normally distributed with mean $\boldsymbol{0}$ and covariance matrix

$$R_n = \begin{pmatrix} 1 & 0 \\ 0 & n^{-1} \end{pmatrix}.$$

Let

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

for which $R^\dagger = R$. *Then (5.11) and (5.12) hold (but note that for all* n, *rank*$(R_n) \neq$ *rank*(R)). *Since* $\boldsymbol{w}_n' R_n^- \boldsymbol{w}_n \sim \chi_2^2$ *for each* n *and rank*$(R) = 1$, *we see that (5.13) does not hold.*

In our setup, it is only because we chose $\hat{\Sigma}$ *to have the same structure as* Σ *that we were able to establish that rank*$(\hat{\Sigma}) \xrightarrow{P}$ *rank*(Σ).

Before proving Proposition 2.1, we recall the following well-known fact.

Lemma 5.1 *Let* A *be a symmetric matrix and let* $\mathcal{M}(A)$ *denote the space spanned by the column vectors of* A. *Then for any pair of* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{M}(A)$, *the quadratic form* $\boldsymbol{x}' A^- \boldsymbol{y}$ *is invariant under the choice of generalized inverse* A^-.

Proof If $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{M}(A)$, then $\boldsymbol{x} = A\boldsymbol{c}$ and $\boldsymbol{y} = A\boldsymbol{d}$ for some vectors $\boldsymbol{c}$ and $\boldsymbol{d}$. Thus

$$\boldsymbol{x}' A^- \boldsymbol{y} = \boldsymbol{c}' A A^- A \boldsymbol{d} = \boldsymbol{c}' A \boldsymbol{d},$$

which does not depend on A^-.

Proof of Proposition 2.1

We follow the notation used in Example 3 of Section 2.3. We first note that for the matrix $D(\theta, \boldsymbol{t})$ given by (2.17), R1–R3 follow immediately from the assumptions on $\Sigma^{(1)}(\cdot,\cdot)$. Moreover, minimizing $\boldsymbol{\zeta}_n(\theta)' M^-(\theta, \boldsymbol{a}^{(n)}) \boldsymbol{\zeta}_n(\theta)$ is the same as minimizing $\boldsymbol{\xi}_n(\theta)'\boldsymbol{\xi}_n(\theta)$ where $\boldsymbol{\xi}_n(\theta) = D(\theta, \boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\theta)$. So, by Lemma A.1 of the Appendix, the resulting estimator $\hat{\theta}$ satisfies A3. Recall that $Q = \boldsymbol{\xi}_n'(\hat{\theta})\ \hat{\Sigma}^\dagger\ \boldsymbol{\xi}_n(\hat{\theta})$ and $\tilde{Q} = \boldsymbol{\zeta}_n'(\hat{\theta}) M^-(\hat{\theta}, \boldsymbol{a}^{(n)}) \boldsymbol{\zeta}_n(\hat{\theta})$. Theorem 3 applies and we have $Q \xrightarrow{d} \chi^2_{k-q-1}$. Let $\Omega_n = \{\omega \in \Omega : \text{rank}(\hat{C}) = q;\ \text{rank}(D_n(\hat{\theta})) = k;\ \text{rank}(\hat{\Sigma}^{(1)}) = k-1\}$. We saw in the proof of Part 2 of Theorem 2 that $P(\Omega_n) \to 1$. We shall show that on Ω_n

$$Q = \tilde{Q} - \left(\boldsymbol{\zeta}_n'(\hat{\theta})\hat{M}^- \frac{\partial \boldsymbol{p}^{(n)}(\theta)}{\partial \theta'}\right)\left(\frac{\partial \boldsymbol{p}^{(n)'}(\hat{\theta})}{\partial \theta}\hat{M}^- \frac{\partial \boldsymbol{p}^{(n)}(\hat{\theta})}{\partial \theta'}\right)^{-1}\left(\frac{\partial \boldsymbol{p}^{(n)'}(\hat{\theta})}{\partial \theta}\hat{M}^- \boldsymbol{\zeta}_n(\hat{\theta})\right), \quad (5.14)$$

where $\hat{M}$ denotes $M(\hat{\theta}, \boldsymbol{a}^{(n)})$. Note that since on Ω_n we have $\hat{M}'1 = 0$ and $\text{rank}(\hat{M}) = k-1$, the equations $1'\partial \boldsymbol{p}^{(n)}(\hat{\theta})/\partial\theta' = \partial(1)/\partial\theta = 0$ and $1'\boldsymbol{\zeta}_n(\hat{\theta}) = 0$ imply that $\boldsymbol{\zeta}_n(\hat{\theta})$ and the column vectors of $\partial \boldsymbol{p}^{(n)}(\hat{\theta})/\partial\theta'$ are in the range space $\mathcal{M}(\hat{M})$ of $\hat{M}$. Therefore, by Lemma 5.1 the three quadratic forms in parentheses in (5.14) are invariant under the choice of generalized inverse of $\hat{M}$. Now in the proof of Lemma A.1 we show that

$$\frac{\partial \boldsymbol{p}^{(n)'}(\hat{\theta})}{\partial \theta}\hat{M}^- \boldsymbol{\zeta}_n(\hat{\theta}) = \frac{\partial \boldsymbol{p}^{(n)'}(\hat{\theta})}{\partial \theta} D^2(\hat{\theta}, \boldsymbol{a}^{(n)})\boldsymbol{\zeta}_n(\hat{\theta}) = \frac{\partial \boldsymbol{p}^{(n)'}(\hat{\theta})}{\partial \theta} D(\hat{\theta}, \boldsymbol{a}^{(n)})\boldsymbol{\xi}_n(\hat{\theta}) = O_p(n^{-\frac{1}{2}})$$

under H_0 (see (A.7)). Therefore, (5.14) implies

$$Q = \tilde{Q} + O_p(n^{-1}) \xrightarrow{d} \chi^2_{k-q-1} \quad \text{under } H_0.$$

The rest of the proof is devoted to the verification of (5.14). By (2.16 and (2.17),

$$D(\cdot,\cdot)M(\cdot,\cdot)D(\cdot,\cdot) = H\begin{pmatrix} I_{k-1} & 0 \\ 0' & 0 \end{pmatrix}H' = HH' - H\begin{pmatrix} 0_{(k-1)\times(k-1)} & 0 \\ 0' & 1 \end{pmatrix}H' = I_k - 11'/k$$

and

$$D(\cdot,\cdot)1 = H\begin{pmatrix} A^{-\frac{1}{2}}(\cdot,\cdot) & 0 \\ 0' & 1 \end{pmatrix}H'1 = \left(GA^{-\frac{1}{2}}(\cdot,\cdot)G' + 11'/k\right)1 = 0 + 1 = 1, \qquad (5.15)$$

where we recall that $G = J(J'J)^{-1}$. Thus,

$$C'1 = \frac{\partial \boldsymbol{p}'(\theta_0)}{\partial \theta} D(\theta_0)1 = \frac{\partial \boldsymbol{p}'(\theta_0)}{\partial \theta} 1 = 0.$$

Therefore,

$$\begin{aligned} \Sigma &= PD(\theta_0)J\Sigma^{(1)}J'D(\theta_0)P \\ &= P[D(\theta_0)M(\theta_0,\boldsymbol{a})D(\theta_0)]P \\ &= \left(I_k - C(C'C)^{-1}C'\right)\left(I_k - 11'/k\right)\left(I_k - C(C'C)^{-1}C'\right) \\ &= I_k - 11'/k - C(C'C)^{-1}C', \end{aligned} \qquad (5.16)$$

which is idempotent (i.e. $\Sigma^2 = \Sigma$). Similarly, $\hat{\Sigma} = I_k - 11'/k - \hat{C}(\hat{C}'\hat{C})^{-1}\hat{C}'$ is idempotent on Ω_n. Recall that an idempotent matrix is its own Moore-Penrose inverse. Thus, on Ω_n we have $\hat{\Sigma}^\dagger = \hat{\Sigma}$ and

$$\begin{aligned} Q &= \boldsymbol{\xi}_n'(\hat{\theta})\, \hat{\Sigma}^\dagger\, \boldsymbol{\xi}_n(\hat{\theta}) \\ &= \boldsymbol{\zeta}_n'(\hat{\theta})\, D_n(\hat{\theta})\left(I_k - 11'/k - \hat{C}(\hat{C}'\hat{C})^{-1}\hat{C}'\right)D_n(\hat{\theta})\boldsymbol{\zeta}_n(\hat{\theta}) \\ &= \boldsymbol{\zeta}_n'(\hat{\theta})\, D_n(\hat{\theta})^2\boldsymbol{\zeta}_n(\hat{\theta}) - (\boldsymbol{\zeta}_n'(\hat{\theta})D_n(\hat{\theta})1)^2/k - \boldsymbol{\zeta}_n'(\hat{\theta})\, D_n(\hat{\theta})\hat{C}(\hat{C}'\hat{C})^{-1}\hat{C}'D_n(\hat{\theta})\boldsymbol{\zeta}_n(\hat{\theta}) \\ &= \tilde{Q} - \boldsymbol{\zeta}_n'(\hat{\theta})\, D_n^2(\hat{\theta})\frac{\partial \boldsymbol{p}^{(n)}(\hat{\theta})}{\partial \theta'}\left(\frac{\partial \boldsymbol{p}^{(n)\prime}(\hat{\theta})}{\partial \theta}D_n^2(\hat{\theta})\frac{\partial \boldsymbol{p}^{(n)}(\hat{\theta})}{\partial \theta'}\right)^{-1}\frac{\partial \boldsymbol{p}^{(n)\prime}(\hat{\theta})}{\partial \theta}D_n^2(\hat{\theta})\boldsymbol{\zeta}_n(\hat{\theta}) \end{aligned}$$

where in the last equality we have used (5.15) and the fact that $\boldsymbol{\zeta}_n'(\hat{\theta})1 = 0$. This proves (5.14).

Appendix

A.1 The Asymptotic Distribution of the Minimum Chi-Square Estimator

Here we prove that the minimum chi-square estimator defined by (2.9) satisfies (2.8). We assume that the parametric model holds and that $\theta = \theta_0$.

Lemma A.1 *Let $\{\hat{\theta}_n\}_{n=1}^{\infty}$ be an infinite sequence of statistics such that $\hat{\theta} = \hat{\theta}_n$ satisfies (2.9) for every n. Assume that (2.1) and Assumptions A1 and A2 hold, and that $D(\cdot,\cdot)$ satisfies R1–R3. Then*

1 $\hat{\theta} \xrightarrow{P} \theta_0$

2 Assumption A3 is satisfied.

Note that in the statement of Lemma A.1 we assume that for each n, there *exists* a number $\hat{\theta}_n$ that minimizes $\boldsymbol{\xi}_n'(\theta)\boldsymbol{\xi}_n(\theta)$, i.e. that satisfies (2.9) (actually, we need only assume that with probability tending to one as $n \to \infty$ such a $\hat{\theta}_n$ exists). No assumption of uniqueness is made.

Proof The proof of Part 1 is based on an application of the Implicit Function Theorem. In the first part of the proof of Theorem 1 we showed (without using Assumption A3) that $\boldsymbol{\zeta}_n(\theta_0) = n^{\frac{1}{2}}(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta_0)) \xrightarrow{d} \mathcal{N}_k(0, J\Sigma^{(1)}J')$ and $\boldsymbol{\xi}_n(\theta_0) \xrightarrow{d} \mathcal{N}_k(0, D(\theta_0)J\Sigma^{(1)}J'D(\theta_0))$ as $n \to \infty$ (see (5.5)). This implies

$$\hat{\boldsymbol{p}} - \boldsymbol{p}(\theta_0) = \left(\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta_0)\right) + \left(\boldsymbol{p}^{(n)}(\theta_0) - \boldsymbol{p}(\theta_0)\right) \xrightarrow{P} 0.$$

We also have

$$\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\hat{\theta}) = n^{-\frac{1}{2}}\boldsymbol{\zeta}_n(\hat{\theta}) \xrightarrow{P} 0,$$

since

$$||\boldsymbol{\zeta}_n(\hat{\theta})|| \leq ||D_n^{-1}(\hat{\theta})||\ ||\boldsymbol{\xi}_n(\hat{\theta})|| \leq ||D_n^{-1}(\hat{\theta})||\ ||\boldsymbol{\xi}_n(\theta_0)|| = O_p(1) \tag{A.1}$$

where the last step follows since $D_n^{-1}(\hat{\theta})$ is bounded in probability (by regularity condition R2 and the fact that $\boldsymbol{a}^{(n)} \xrightarrow{P} \boldsymbol{a}$) and $\boldsymbol{\xi}_n(\theta_0)$ converges in distribution. Here $||\cdot||$ denotes Euclidean distance in $\mathbb{R}^k$. Therefore

$$\boldsymbol{p}^{(n)}(\hat{\theta}) \xrightarrow{P} \boldsymbol{p}(\theta_0). \tag{A.2}$$

Without loss of generality, assume that the first q row vectors of $\frac{\partial \boldsymbol{p}(\theta)}{\partial \theta'}$ are linearly independent. Let $\boldsymbol{f} = (f_1, \ldots, f_q)$ be the vector-valued function defined on the open set $S = \Theta \times (\epsilon, M)^q \times (0,1)^q \subset R^{q+2q}$ by

$$\begin{cases} f_1(\theta; \boldsymbol{x}, \boldsymbol{y}) &=& y_1 - F_\theta(x_1) \\ f_2(\theta; \boldsymbol{x}, \boldsymbol{y}) &=& y_2 - (F_\theta(x_2) - F_\theta(x_1)) \\ &\vdots& \\ f_q(\theta; \boldsymbol{x}, \boldsymbol{y}) &=& y_q - (F_\theta(x_q) - F_\theta(x_{q-1})) \end{cases}$$

where $\boldsymbol{x} = (x_1, \ldots, x_q)$ and $\boldsymbol{y} = (y_1, \ldots, y_q)$. Denote $\boldsymbol{x}_0 = (a_1, \ldots, a_q)$ and $\boldsymbol{y}_0 = (p_1(\theta_0), \ldots, p_q(\theta_0))$. Then

$$\boldsymbol{f}(\theta_0; \boldsymbol{x}_0, \boldsymbol{y}_0) = 0. \tag{A.3}$$

Moreover, by Assumptions A1 and A2, $\boldsymbol{f}$ is continuously differentiable on S and its Jacobian determinant with respect to θ at $(\theta_0, \boldsymbol{x}_0, \boldsymbol{y}_0)$ is not equal to zero. Thus, by the Implicit Function Theorem (see e.g. Theorem 13.7 of Apostol (1974)), there exist a neighborhood T_0 of $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ in R^{2q} and one, and only one, continuously differentiable function $\boldsymbol{g} : T_0 \to R^q$ such that $\boldsymbol{g}(\boldsymbol{x}_0, \boldsymbol{y}_0) = \theta_0$, and

$$\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x}, \boldsymbol{y}) = 0 \ \text{ for all } \ (\boldsymbol{x}, \boldsymbol{y}) \in T_0. \tag{A.4}$$

Note that for $\boldsymbol{x}_n = \left(a_1^{(n)}, \ldots, a_q^{(n)}\right)$ and $\boldsymbol{y}_n = \left(p_1^{(n)}(\hat{\theta}), \ldots, p_q^{(n)}(\hat{\theta})\right)$, we have

$$\boldsymbol{f}(\hat{\theta}; \boldsymbol{x}_n, \boldsymbol{y}_n) = 0 \quad \text{for all } n.$$

This together with (A.4) implies that

$$P(\hat{\theta} \neq g(\boldsymbol{x}_n, \boldsymbol{y}_n)) \leq P((\boldsymbol{x}_n, \boldsymbol{y}_n) \notin T_0). \tag{A.5}$$

Therefore, for every $\epsilon > 0$,

$$\begin{aligned} P(|\hat{\theta} - \theta_0| \geq \epsilon) &\leq P(|\hat{\theta} - \theta_0| \geq \epsilon,\ \hat{\theta} = g(\boldsymbol{x}_n, \boldsymbol{y}_n)) + P(\hat{\theta} \neq g(\boldsymbol{x}_n, \boldsymbol{y}_n)) \\ &\leq P(|g(\boldsymbol{x}_n, \boldsymbol{y}_n) - g(\boldsymbol{x}_0, \boldsymbol{y}_0)| \geq \epsilon) + P((\boldsymbol{x}_n, \boldsymbol{y}_n) \notin T_0), \end{aligned}$$

and this converges to 0 since $(\boldsymbol{x}_n, \boldsymbol{y}_n) \xrightarrow{P} (\boldsymbol{x}_0, \boldsymbol{y}_0)$, T_0 is a neighborhood of $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, and $g(\cdot,\cdot)$ is continuous on T_0. Here the last inequality follows from (A.5). This proves Part 1 of the lemma.

To prove Part 2 of the lemma we write

$$\begin{aligned} \boldsymbol{\xi}_n(\hat{\theta}) &= D_n(\hat{\theta}) n^{\frac{1}{2}} (\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\hat{\theta})) \\ &= D_n(\hat{\theta}) n^{\frac{1}{2}} (\hat{\boldsymbol{p}} - \boldsymbol{p}^{(n)}(\theta_0)) - D_n(\hat{\theta}) n^{\frac{1}{2}} (\boldsymbol{p}^{(n)}(\hat{\theta}) - \boldsymbol{p}^{(n)}(\theta_0)) \\ &= \Big(D_n(\theta_0) + o_p(1)\Big)\boldsymbol{\zeta}_n(\theta_0) - \Big(D(\theta_0) + o_p(1)\Big)\left(\frac{\partial \boldsymbol{p}(\theta_0)}{\partial \theta'} + o_p(1)\right) n^{\frac{1}{2}}(\hat{\theta} - \theta_0) \\ &= \boldsymbol{\xi}_n(\theta_0) + o_p(1)\boldsymbol{\zeta}_n(\theta_0) - (C + o_p(1)) n^{\frac{1}{2}}(\hat{\theta} - \theta_0) \\ &= \boldsymbol{\xi}_n(\theta_0) - (C + o_p(1)) n^{\frac{1}{2}}(\hat{\theta} - \theta_0) + o_p(1) \end{aligned} \tag{A.6}$$

where to obtain the last step we have used the fact that $\boldsymbol{\zeta}_n(\theta_0)$ is bounded in probability since it converges in distribution (by (5.5)).

Now since $\boldsymbol{\xi}_n'(\theta)\boldsymbol{\xi}_n(\theta)$ is continuously differentiable in θ and it has a local minimum at $\hat{\theta}$, we have for $j = 1, \ldots, q$

$$\frac{\partial(\boldsymbol{\xi}_n'(\hat{\theta})\boldsymbol{\xi}_n(\hat{\theta}))}{\partial \theta_j} = -2n^{\frac{1}{2}} \frac{\partial \boldsymbol{p}^{(n)\prime}(\hat{\theta})}{\partial \theta_j} D_n(\hat{\theta})\boldsymbol{\xi}_n(\hat{\theta}) + \boldsymbol{\zeta}_n'(\hat{\theta}) \frac{\partial(D_n^2(\hat{\theta}))}{\partial \theta_j} \boldsymbol{\zeta}_n(\hat{\theta}) = 0$$

(recall that $\hat{\theta}$ is in Θ, which is open by assumption). This implies that for $j = 1, \ldots, q$,

$$\frac{\partial \boldsymbol{p}^{(n)\prime}(\hat{\theta})}{\partial \theta_j} D_n(\hat{\theta})\boldsymbol{\xi}_n(\hat{\theta}) = \frac{1}{2} n^{-\frac{1}{2}} \boldsymbol{\zeta}_n'(\hat{\theta}) \frac{\partial(D_n^2(\hat{\theta}))}{\partial \theta_j} \boldsymbol{\zeta}_n(\hat{\theta}) = O_p(n^{-\frac{1}{2}}), \tag{A.7}$$

where the last step follows from (A.1), the consistency of $\hat{\theta}$, and regularity condition R3. Because (A.7) can be rewritten as

$$\left(\frac{\partial \boldsymbol{p}'(\theta_0)}{\partial \theta} D(\theta_0) + o_p(1)\right)\boldsymbol{\xi}_n(\hat{\theta}) = O(n^{-\frac{1}{2}}),$$

we have

$$C'\boldsymbol{\xi}_n(\hat{\theta}) = o_p(1)\boldsymbol{\xi}_n(\hat{\theta}) + O_p(n^{-\frac{1}{2}}) = o_p(1). \tag{A.8}$$

Here again we have used the fact that $||\boldsymbol{\xi}_n(\hat{\theta})|| \leq ||\boldsymbol{\xi}_n(\theta_0)||$ is bounded in probability. Now, multiplying both sides of (A.6) by C' and using (A.8), we obtain

$$\left(C'C + o_p(1)\right) n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = C'\boldsymbol{\xi}_n(\theta_0) + o_p(1),$$

from which we conclude that

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = \left(C'C + o_p(1)\right)^{-1} C'\boldsymbol{\xi}_n(\theta_0) + o_p(1) = \left(C'C\right)^{-1} C'\boldsymbol{\xi}_n(\theta_0) + o_p(1),$$

as desired.

Remark A.1 *Suppose that $\hat{D}$ is a consistent estimate of $D(\theta_0, \boldsymbol{a})$. If we replace $D_n(\theta)$ by $\hat{D}$ in (2.7) and let $\hat{\theta}$ be the value of θ which minimizes $\boldsymbol{\zeta}_n'(\theta)\hat{D}^2\boldsymbol{\zeta}_n(\theta)$, then the conclusions of Lemma A.1 still hold. The proof of this is identical to the proof of Lemma A.1, except that (A.1), (A.7) and (A.8) are more straightforward. This fact is used in Remark 3.1.*

References

Akritas, M. (1988). Pearson-type goodness-of-fit tests: The univariate case. *J. Amer. Statist. Assoc.* ***83*** 222–230.

Andersen, P. K. and Borgan, O. (1985). Counting process models for life history data: a review (with discussion). *Scand. J. Statist.* ***12*** 97–158.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* ***10*** 1100-1120.

Apostol, T. M. (1975). *Mathematical Analysis.* Addison-Wesley, Reading, Mass.

Campbell, S. L. and Meyer, C. D. Jr. (1979). *Generalized Inverses of Linear Transformations.* Pitman, San Francisco.

Čebyšev, D. M. (1971). Certain chi-square type tests for continuous distributions. *Theor. Probability Appl.* ***16*** 1–22.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum-likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* ***25*** 579–586.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, N. J.

Doss, H. and Gill, R. D. (1990). A method for obtaining weak convergence results for quantile process, with applications to censored survival data. Technical Report No. M-856, Department of Statistics, Florida State University.

Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *J. Roy. Statist. Soc.* ***85*** 87–94.

Fisher, R. A. (1924). The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *J. Roy. Statist. Soc.* ***87*** 442–450.

Gill, R. D. (1980). *Censoring and Stochastic Integrals.* Math. Centre Tracts 124, Mathematical Centre, Amsterdam.

Habib, M. and Thomas, D. (1986). Chi-square goodness-of-fit tests for randomly censored data. *Ann. Statist.* ***14*** 759–765.

Hjort, N. L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann. Statist.* ***18*** 1221–1258.

Hollander, M. and Peña, E. (1990). A chi-square goodness-of-fit test for randomly censored data. *J. Amer. Statist. Assoc.* ***87*** 458–563.

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inferences based on retrospective ascertainment: An analysis of data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* ***84*** 360–372.

Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* ***18*** 582–602.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Not. Astr. Soc.* ***155*** 95–118.

McKeague, I. W. and Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand. J. Statist.* ***18*** 177-195.

Mihalko, D. and Moore, D. (1980). Chi-square tests of fit for Type II censored data. *Ann. Statist.* ***8*** 625–644.

Miller, R. G., Jr. (1983). What price Kaplan-Meier? *Biometrics* ***39*** 1077–1081.

Moore, D. S. (1971). A chi-square statistic with random cell boundaries. *Ann. Math. Statist.* ***42*** 147–156.

Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared test of fit. *Ann. Statist.* ***3*** 599–616.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag. Ser. 5,* ***50*** 157–172.

Pollard, D. (1979). General chi-square goodness-of-fit tests with data-dependent cells. *Z. Wahrsch. Verw. Gebiete* ***50*** 317–331.

Skorohod, A. V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* ***1*** 261–290.

Wang, M-C. (1989). A semiparametric model for randomly truncated data. *J. Amer. Statist. Assoc.* ***84*** 742–748.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* ***13*** 163–177.