Statistica Sinica **14**(2004), 231-248

GOODNESS-OF-FIT TESTS AND MINIMUM POWER DIVERGENCE ESTIMATORS FOR SURVIVAL DATA

Huann-Sheng Chen, Kunjung Lai and Zhiliang Ying

Michigan Technological University, Feng-Chia University and Columbia University

Abstract: Power-divergence statistics are proposed for grouped survival data. They are analogous to the power-divergence family of statistics proposed and studied in detail by Read and Cressie (1988) and Cressie and Read (1984) for contingency tables. The proposed statistics are useful for testing validity of parametric model assumptions in analyses of survival data. It is shown that these statistics have approximately chi-squared distribution under the null hypothesis. They can be used to construct parameter estimates that are consistent and asymptotically normal under usual regularity conditions. Simulation studies indicate that, with a suitable choice of the tuning parameter, the chi-squared approximation performs quite well even with small to moderate sample sizes. The approach is illustrated with a data set from the reaction control system of the Space Shuttle.

Key words and phrases: Chi-squared distribution, goodness-of-fit test, life table, power-divergence statistics.

1. Introduction

The power-divergence family of statistics was proposed by Cressie and Read (1984) for dealing with discrete data, especially data of counts. See Read and Cressie (1988) for a comprehensive coverage of the subject. Let $\mathbf{x} = (x_1, \ldots, x_k)$ denote a random vector of counts having multinomial $(m, \mathbf{p})$ distribution, where $\mathbf{p} = (p_1, \ldots, p_k)$ is the vector of cell probabilities. Then $\sum_{i=1}^k x_i = m$ and $\sum_{i=1}^k p_i = 1$ and, for any k-vector $\mathbf{u} = (u_1, \ldots, u_k)$ with $\sum_{i=1}^k u_i = m$,

$$P(\mathbf{x} = \mathbf{u}) = m! \prod_{i=1}^{k} \frac{p_i^{u_i}}{x_i!}. \tag{1.1}$$

Consider $H_0 : \mathbf{p} \in \mathcal{P}_0$, where $\mathcal{P}_0$ represents a set of values hypothesized for $\mathbf{p}$. Denote by $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_k)$, the maximum likelihood estimator (MLE) of $\mathbf{p}$ under H_0. Then the power-divergence statistics can be written as

$$CR(\lambda) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{k} x_i \left[\left(\frac{x_i}{n\hat{p}_i} \right)^{\lambda} - 1 \right], \tag{1.2}$$

where $\lambda \in (-\infty, \infty)$.

An array of choices of statistics for testing H_0 are provided by the power-divergence family. In particular, we find that

$$
\begin{aligned}
CR(1) &= \sum_{i=1}^{k} \frac{(X_i - n\hat{p}_i)^2}{np_i} && (\text{ Pearson's } \chi^2),\\
CR(0) &= 2\sum_{i=1}^{k} X_i \log \frac{X_i}{n\hat{p}_i} && (G^2),\\
CR(-\tfrac{1}{2}) &= 4\sum_{i=1}^{k} (\sqrt{X_i} - \sqrt{n\hat{p}_i})^2 && (\text{Freeman-Tukey's } F^2), \qquad (1.3)\\
CR(-1) &= 2\sum_{i=1}^{k} n\hat{p}_i \log \frac{n\hat{p}_i}{X_i} && \Big(\text{Neyman's modified } \chi^2\Big),\\
CR(-2) &= \sum_{i=1}^{k} \frac{(X_i - n\hat{p}_i)^2}{X_i} && (\text{ modified } G^2).
\end{aligned}
$$

Note that $CR(0)$ and $CR(-1)$ are defined as limits of $CR(\lambda)$ as $\lambda \to 0$ and -1. Thus, the power-divergence family unifies commonly used goodness-of-fit statistics.

The main purpose of the present paper is to develop a similar class of statistics that are useful in survival analysis. Survival data arise frequently in medical follow-up studies, actuarial calculation and industrial life testing. The data are often modeled semi-parametrically, the Cox model, for example, or parametrically. Although semi-parametric models are popular, parametric models often provide viable alternatives, there are many examples in censored data analysis and reliability (Nelson(1990)). In reliability theory one finds the exponential and Weibull distributions and subsequently IFR (increasing failure rate) and DFR (decreasing failure rate) distributions. Bringing in an understanding of aging (IFR) or of objects whose reliability properties improve over time (DFR) can sometimes be more suitable than employing semi-parametric models (Barlow and Proschan (1981)). See also Kalbfleisch and Prentice (1981), Lawless (1982) and Cox and Oakes (1984) for some other well-known parametric models that can be used to fit survival data gathered from medical research and actuarial sciences. An important aspect then is how to check validity of a specific model assumption. Some research have been done in this regard. For example, Gail and Ware (1979) studied grouped censored survival data by comparing with a known survival distribution, while Akritas (1988) constructed a Pearson-type goodness-of-fit measure for one-sample data that allows for random censorship.

The rest of the paper is organized as follows. The family of power-divergence statistics for grouped survival data is defined in Section 2, where basic notation is also introduced. In Section 3, power-divergence measures are applied to obtain a class of estimators under parametric assumptions. It is shown that, when the model is correctly specified, the resultant estimators are asymptotically equivalent to the maximum likelihood estimator. In consequence, they are asymptotically normal and approximate $(1-\alpha)\times 100\%$ confidence regions can be easily constructed. Use of the proposed power-divergence statistics for checking parametric model assumptions is discussed in Section 4. Extensive simulation studies are presented in Section 5. Section 6 illustrates the new approach with a data example. Concluding remarks are in Section 7. Proofs are relegated to an appendix.

2. Power-divergence Statistics for Grouped Survival Data

We introduce some notation. Suppose the follow-up period is the interval between 0 and τ, partitioned into k subintervals $(\tau_{i-1}, \tau_i]$, $i = 1, \ldots, k$, where $0 = \tau_0 < \tau_1 < \ldots < \tau_k = \tau$. Let n_i be the number of subjects at risk at the beginning of the ith interval and d_i be the number of failures during the interval. To avoid complication, assume that censoring occurs only at τ_i. For each i, let $\mathcal{F}_i = \sigma\{d_1, \ldots, d_{i-1}, n_1, \ldots, n_i\}$ be the σ-field generated by $d_1, \ldots, d_{i-1}, n_1, \ldots, n_i$. Thus, conditional on $\mathcal{F}_i$, d_i has a binomial distribution,

$$\mathcal{P}(d_i = l \mid \mathcal{F}_i) = \binom{n_i}{l} h_{0,i}^l (1 - h_{0,i})^{n_i - l}, \qquad l = 0, \ldots, n_i, \tag{2.1}$$

where, for each i, $h_{0,i}$ is a positive constant between 0 and 1 and may be regarded as the discrete hazard rate at the ith interval.

The grouped survival data as just described can be constructed from continuous survival data. Suppose there are n study subjects whose failure and censoring times are denoted by T_j and C_j, $j = 1, \ldots, n$, so that observations consist of $\tilde{T}_j = \min(T_j, C_j)$ and $\delta_j = I(T_j \leq C_j)$, $j = 1, \ldots, n$. The T_j are independent with a common distribution function F_0. Let $n_j = \#\{j : \tilde{T}_j > \tau_{j-1}\}$ and $d_j = \#\{j : \delta_j = 1, \tilde{T}_j \in (\tau_{i-1}, \tau_i]\}$. If censoring occurs only at τ_i, it is easily verified that (2.1) holds with $h_{0,i} = [F_0(\tau_i) - F_0(\tau_{i-1})]/[1 - F_0(\tau_{i-1})]$.

Mimicking (1.2), we propose the power-divergence family of statistics for grouped survival data

$$D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{k} \left\{ d_i \left[\left(\frac{d_i}{n_i h_i} \right)^\lambda - 1 \right] + d_i^c \left[\left(\frac{d_i^c}{n_i h_i^c} \right)^\lambda - 1 \right] \right\}, \tag{2.2}$$

where $\mathbf{d} = (d_1, \ldots, d_k)$, $\mathbf{n} = (n_1, \ldots, n_k)$, $\mathbf{h} = (h_1, \ldots, h_k)$, $d_i^c = n_i - d_i$ and $h_i^c = 1 - h_i$. Because of (2.1), each d_i can be thought of as a binomial. In this connection, (2.2) is simply a sum of Cressie and Read's (1984) power-divergence statistics applied to the k intervals. In particular,

$$D_1(\mathbf{d},\mathbf{n};\mathbf{h}) = \sum_{i=1}^{k}\left\{\frac{(d_i - n_i h_{0i})^2}{n_i h_{0i}} + \frac{(d_i^c - n_i h_{0i}^c)^2}{n_i h_{0i}^c}\right\}$$

$$D_{-\frac{1}{2}}(\mathbf{d},\mathbf{n};\mathbf{h}) = 4\sum_{i=1}^{k}\left\{\left(\sqrt{d_i} - \sqrt{n_i h_{0i}}\right)^2 + \left(\sqrt{d_i^c} - \sqrt{n_i h_{0i}^c}\right)^2\right\}$$

$$D_0(\mathbf{d},\mathbf{n};\mathbf{h}) = 2\sum_{i=1}^{k}\left\{d_i \log\left(\frac{d_i}{n_i h_{0i}}\right) + d_i^c \log\left(\frac{d_i^c}{n_i h_{0i}^c}\right)\right\}$$

$$D_{-1}(\mathbf{d},\mathbf{n};\mathbf{h}) = 2\sum_{i=1}^{k}\left\{n_i h_{0i} \log\left(\frac{n_i h_{0i}}{d_i}\right) + n_i h_{0i}^c \log\left(\frac{n_i h_{0i}^c}{d_i^c}\right)\right\}$$

$$D_{-2}(\mathbf{d},\mathbf{n};\mathbf{h}) = \sum_{i=1}^{k}\left\{\frac{(d_i - n_i h_{0i})^2}{d_i} + \frac{(d_i^c - n_i h_{0i}^c)^2}{d_i^c}\right\}$$

are analogues of Pearson's χ^2, Freeman-Tukey's F^2, G^2, Neyman's modified χ^2 and modified G^2 statistics.

Theorem 1. *Suppose* (2.1) *holds and* $n_i \to \infty$ *for* $i = 1, \ldots k$. *Then, for each* λ, $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}_0)$ *converges in distribution to* χ_k^2, *the (central) chi-squared distribution with k degrees of freedom. If* $n_i/n \to r_i \geq 0$, $i = 1, \ldots, k$, *then when* h_{0i} *in* (2.1) *is replaced by* $h_i = h_{0i} + c_i/\sqrt{n_i}$, *where* c_i, $i = 1, \ldots, k$, *are constants,* $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}_0)$ *converges to* $\chi_k^2(\eta)$, *a non-central chi-squared distribution with non-centrality parameter* $\eta = \sum_{i=1}^{k} c_i^2/(h_{0i}(1-h_{0i}))$.

The proof of the preceding theorem is straightforward and is outlined in the Appendix. The result is directly applicable to testing the hypothesis that survival times follow a specific distribution. The alternatives $h_i = h_{0i} + c_i/\sqrt{n_i}$ are contiguous to the null hypothesis $h_i = h_{0i}$. Theorem 1 shows that the tests for $\mathbf{h} = \mathbf{h}_0$ derived from this family are asymptotically equivalent in the sense that both type I and type II errors under contiguous alternatives are asymptotically the same. The results also provide a way of obtaining the power of tests with contiguous alternatives, but the type II error under non-contiguous alternatives may be quite different with different choices of λ.

We have assumed that censoring occurs instantaneously at the end of each interval. The assumption is commonly used in life-table estimators, and in most discrete time-to-event or grouped survival data analysis where the censoring information is available only up to intervals. For example, see Guo and Lin (1994), Prentice and Gloeckler (1978) and Cox (1975). If full information about censoring is absent but censoring is not heavy, Cox (1975) considered an approximation to the partial likelihood function, using the assumption that the failures and censoring in any interval are independent Poisson processes. It is shown that under such conditions, the estimates using the partial likelihood function have negligible bias correction when the number of intervals becomes large. The scheme discussed in Cox (1975) can likewise be applied to the power divergence family for grouped survival data while using (2.2).

3. Minimum Power-divergence Estimators

An important subject in parametric survival analysis is the estimation of unknown parameters under the model assumption. This section intends to deal with this subject by using the proposed power-divergence measures.

Consider a parametric family indexed by $\theta \in \Theta \subset \mathcal{R}^p$. Let $\mathcal{H} = \{\mathbf{h}(\theta) = (h_1(\theta), \ldots, h_k(\theta)) : \theta \in \Theta\}$, where h_i are twice continuously differentiable in Θ. We use θ_0 to denote the true parameter value and $\mathbf{h}_0 = \mathbf{h}(\theta_0)$. As before, $h_i^c(\theta) = 1 - h_i(\theta)$. For each λ, we can define a minimum dispersion estimator $\hat{\theta}_\lambda$ as the minimizer of

$$D_\lambda(\theta) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{k} \left\{ d_i \left[\left(\frac{d_i}{n_i h_i(\theta)} \right)^{\lambda} - 1 \right] + d_i^c \left[\left(\frac{d_i^c}{n_i h_i^c(\theta)} \right)^{\lambda} - 1 \right] \right\}. \quad (3.1)$$

Under the assumption that the censoring occurs only at the end of each interval, (2.1) will hold and, because $\lambda = 0$ corresponds to G^2, it follows that $\hat{\theta}_0$ coincides with the maximum likelihood estimator. Likewise, $\hat{\theta}_1$ may be regarded as a minimum χ^2 distance estimator and $\hat{\theta}_{-1/2}$ as a minimum Hellinger distance estimator.

Theorem 2. *Suppose $n_i/n \to \gamma_i \geq 0, i = 1, \ldots, k$. Assume the $p \times p$ matrix*

$$V = \sum_{i=1}^{k} \gamma_i \nabla h_i(\theta_0) \nabla h_i(\theta_0)^T / (h_i(\theta_0) h_i^c(\theta_0))$$

is strictly positive definite. Then, in a neighborhood of θ_0, $\hat{\theta}_\lambda$ is uniquely defined and $\sqrt{n}(\hat{\theta}_\lambda - \theta_0)$ is asymptotically normal with mean zero and variance-covariance matrix V^{-1}. In addition, $\hat{V}_n^{1/2}(\hat{\theta}_\lambda - \theta_0)$ converges to the p-variate standard normal distribution, where

$$\hat{V}_n = \sum_{i=1}^{k} [n_i \nabla h_i(\hat{\theta}_\lambda) \nabla h_i(\hat{\theta}_\lambda)^T / (h_i(\hat{\theta}_\lambda) h_i^c(\hat{\theta}_\lambda))]. \quad (3.2)$$

The proof of Theorem 2 is given in the appendix. The assumption that $n_i/n \to \gamma_i \geq 0$ is to ensure stability of the sample; $\hat{V}_n$ in (3.2) approximates V

and depends on λ. The asymptotic normality of $\hat{\theta}_\lambda$ shows that an approximate $(1-\alpha)\times 100\%$ confidence region for θ_0 is $\{\theta : (\hat{\theta}_\lambda - \theta)\hat{V}_n(\hat{\theta}_\lambda)(\hat{\theta}_\lambda - \theta) \leq \chi^2_{\alpha;p}\}$.

4. Goodness-of-fit Tests

Another important subject in parametric survival analysis is the goodness of fit of the parametric model assumption. We demonstrate in this section that the proposed power-divergence family can be used for model checking.

Let $\hat{\theta}$ be any member in $\{\hat{\theta}_\lambda\}$. Substituting θ in (3.1) by $\hat{\theta}$ we get $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}))$, which can be used to test the null hypothesis that the parametric family contains the true distributions, i.e., $\mathbf{h}_0 \in \{\mathbf{h}(\theta), \theta \in \Theta\}$.

Theorem 3. *Under the same assumptions as those of Theorem* 2*,* $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}))$ *converges to* χ^2_{k-p}.

It is often of interest to test a parametric hypothesis of form $H_0 : \theta_0 \in \Theta_0$ versus $H_A : \theta_0 \in \Theta \setminus \Theta_0$, where Θ_0 is a $q(<p)$ dimensional sunset of Θ. We apply the power-divergence statistics by first finding $\hat{\theta}_\lambda = \arg\min_{\theta\in\Theta} D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}))$ and $\hat{\theta}^{(0)}_\lambda = \arg\min_{\theta\in\Theta_0} D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}))$. We omit λ and use $\hat{\theta}$ and $\hat{\theta}^{(0)}$ to denote the two parameters. Now, for any λ, not necessarily the same as that in the definition of $\hat{\theta}$, we can use $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}))$–$D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}^{(0)}))$ to test H_0 against H_A.

Corollary 4. *Suppose that the true parameter* θ_0 *lies in the interior of* Θ_0*. Then for any* λ*,* $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta})) - D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}(\hat{\theta}^{(0)}))$ *converges to* $\chi^2_{(p-q)}$.

5. Simulation Study

We have conducted extensive simulations to evaluate the performance of the the proposed methods under various circumstances. The main goals of these simulations are (1) to compare the performance of the power-divergence statistics and the traditional Pearson statistics in the special case of no censoring; (2) to compare the performance of the power-divergence statistics family when censoring occurs; and (3) to demonstrate numerically that when parameters are estimated, the power-divergence statistics are approximately χ^2 under random censoring.

5.1 Power-divergence and Pearson statistics

Some classical statistics, such as the Pearson statistics, are commonly used in goodness-of-fit tests when there is no censoring. Our proposed power-divergence statistics family provides an alternative way for goodness-of-fit testing to both censored and uncensored data. It is of interest to compare the proposed method with the classical goodness-of-fit statistics under the condition that no censoring had occurred.

We test

$$H_0 : F(t) = 1 - \exp\{-t\} \text{ versus } H_1 : F(t) = 1 - \exp\{-t^\gamma\}. \tag{5.1}$$

The alternative is the Weibull distribution with shape parameter γ. Equivalently, we can test the hazard function

$$H_0 : h(t) = 1 \text{ versus } H_1 : h(t) = \gamma t^{\gamma-1}. \tag{5.2}$$

Assuming no censoring, since (5.1) and (5.2) are equivalent, the power-divergence statistics D_λ derived for (5.2) can be compared to the classical chi-squared tests $CR(\lambda)$ derived for (5.1). A notable difference between the two tests is that, when the number of interval is k, the tests $CR(\lambda)$ applied to (5.1) have $k-1$ degrees of freedom and the tests D_λ applied to (5.2) have k degrees of freedom. In our simulation, we chose the Pearson's chi-squared test $CR(1)$ to compare with the proposed tests D_λ.

The comparison can be based on the power and the achieved α levels of the tests. Following Akritas (1988), we chose $\gamma = 1/(1 + b/\sqrt{N})$, with $b = 0, -4, -2, 2, 4$, $N = 120$ with $k = 7$, and $N = 50$ with $k = 3$, where the value k was the number of intervals. For each combination of the alternative, we generated 1,000 samples. We grouped the data so that the intervals had the same probability under the null hypothesis. The simulation results are summarized in Table 1. When $b = 0$, the data were generated from the null distribution, hence the values in the table correspond to the achieved level. From the simulation results, it appears that almost all the power-divergence tests achieve higher α values than does the Pearson chi-squared test. Among the power-divergence tests, the achieved powers are similar to each other. When the sample size increases from 50 to 120, the powers of the tests increase.

Additionally, we compared the power-divergence tests and the Pearson chi-squared test with

$$H_0 : F(t) = 1 - \exp\{-t\} \text{ versus } H_1 : F(t) = 1 - \exp\{-\frac{t}{\beta}\}, \tag{5.3}$$

$$H_0 : h(t) = 1 \text{ versus } H_1 : h(t) = 1/\beta. \tag{5.4}$$

Table 1. Achieved power at 5% for $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta_0))$ and the Pearson chi-squared test $CR(1)$ when $h(t) = \gamma t^{\gamma-1}$, $\gamma = 1/(1 + b/\sqrt{N})$.

	$N = 120$, $k = 7$						$N = 50$, $k = 3$					
		λ						λ				
b	$CR(1)$	-2	-1	-0.5	0	1	$CR(1)$	-2	-1	-0.5	0	1
0	0.029	0.069	0.046	0.035	0.033	0.030	0.023	0.040	0.027	0.021	0.019	0.017
-4	0.986	0.989	0.988	0.987	0.990	0.986	0.945	0.981	0.964	0.958	0.956	0.931
-2	0.269	0.406	0.358	0.330	0.314	0.258	0.170	0.266	0.206	0.184	0.174	0.139
2	0.176	0.211	0.186	0.177	0.174	0.185	0.083	0.101	0.096	0.096	0.084	0.091
4	0.591	0.596	0.577	0.584	0.591	0.595	0.224	0.248	0.243	0.253	0.243	0.245

We let $\beta = 1/(1 + b/\sqrt{N})$ and kept the other aspects of the foregoing simulation unchanged. To save space, we do not include the results obtained from (5.3) and (5.4), but similar conclusion can be drawn. The simulation results indicate that the proposed power-divergence statistics provide a viable analysis to the goodness-of-fit test when there is no censoring.

5.2. Achieved power and levels for the power-divergence family when data are censored

We conducted simulations for the cases that data are censored under random censorship model. We considered the power-divergence statistics family under the composite hypotheses of Theorem 3. The studies were to compare the achieved nominal levels as the number of intervals, the degree of censoring and the sample size are varied, and to compare the powers of the tests for different alternative hypotheses. We examined the tests D_λ for $\lambda = -2, -1, -1/2, 0, 1/2, 1$ and 2.

Achieved levels of tests.

The exponential and the Weibull distributions were used as the null distributions in the simulations. We generated right censored data by assuming that the censoring hazard function was proportional to the failure time hazard function. We considered sample sizes $n = 30$, $n = 50$ and $n = 100$; the levels of the tests were set to 1% and 5%; the number of intervals were chosen as $k = 5$ and $k = 7$; the shape and scale parameters of the Weibull distribution were specified as 2 and 1, respectively. For each combination, 1000 replications were performed. For each replication, there were approximately $40\% - 45\%$ uncensored values.

Table 2. Achieved level at 5% for power-divergence tests $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$ using the null distribution Weibull(γ, β) with $(\gamma, \beta) = (2, 1)$. The sample size is n and the number of intervals is k.

		λ							non-censored	
n	k	-2	-1	-0.5	0	0.5	1	2	mean	sd
30	5	0.050	0.010	0.025	0.055	0.025	0.025	0.045	0.452	0.101
30	7	0.050	0.065	0.100	0.060	0.070	0.100	0.055	0.408	0.085
50	5	0.020	0.015	0.040	0.030	0.010	0.025	0.055	0.448	0.078
50	7	0.050	0.060	0.050	0.035	0.060	0.060	0.065	0.410	0.076
100	5	0.010	0.010	0.005	0.020	0.010	0.010	0.015	0.449	0.051
100	7	0.050	0.035	0.040	0.045	0.045	0.030	0.030	0.420	0.048

Table 2 summarizes the achieved levels of the tests for the Weibull distribution. The mean and standard deviation of the uncensored percentages are also reported; to conserve space, only those associated with nominal levels of 5% are presented. From the table, we find that the achieved levels are consistent with the specified asymptotic level. The results for exponential null distribution are not presented here but they show a similar pattern. Overall it can be concluded that the asymptotic approximations are acceptable for small to moderate sample sizes.

Achieved power of tests.

We conducted simulations to examine the achieved powers of the tests against specific alternatives. We tested the exponential when the failure times were generated by a Weibull distribution. Also, we tested the Weibull when the underlying failure times were generated by a log-normal distribution. The sample sizes utilized were 30 and 100, the number of interval was 7, and the number of replications was 1000. We only report the achieved powers of the asymptotic 5% level tests here, although the achieved powers of 1% level asymptotic tests were also obtained. Table 3 shows the results for testing the Weibull when the true failure time distribution is log-normal with parameters $\mu = 0.8, 1.4, 2.0$ and $\sigma = 1$. Examining the simulation results, one sees that power increases with sample size. The powers of power-divergence statistics in the family are comparable to each other, and hence the choice of λ can not be decided by the powers of the tests. For testing exponentiality, the true failure time distribution was the Weibull distribution with shape parameters $\alpha = 0.2, 0.4, \ldots, 2.0$, and scale parameter 1. The results are not presented here, but they also show that the power increases with sample size, and the powers for different λ are similar.

Table 3. Achieved power at 5% for power-divergence tests $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$ using the Weibull null distribution against an alternative log-normal(μ, σ) with $\sigma = 1$. The number of interval is 7.

		λ							Uncensored	
	μ	-2	-1	-0.5	0	0.5	1	2	mean	sd
$n = 30$	0.8	0.125	0.115	0.125	0.120	0.115	0.100	0.120	0.564	0.089
	1.4	0.370	0.345	0.335	0.335	0.320	0.330	0.335	0.561	0.095
	2.0	0.440	0.455	0.470	0.470	0.460	0.485	0.450	0.565	0.100
$n = 100$	0.8	0.140	0.130	0.155	0.120	0.135	0.120	0.135	0.572	0.050
	1.4	0.400	0.340	0.355	0.385	0.380	0.315	0.410	0.569	0.048
	2.0	0.465	0.455	0.470	0.475	0.460	0.440	0.470	0.577	0.045

5.3. Sampling distribution of $D_\lambda(\hat{\theta})$.

We also studied moderate sample properties of $D_\lambda(\hat{\theta})$. We simulated failure times from the Weibull with shape parameter 0.6 and scale parameter 1. Each simulation generated 100 failure times, which were divided into 5 groups. About 20% of the samples were censored. The parameter λ was chosen as 1, 0, and $-1/2$. The simulation was repeated 1000 times for each λ. Figure 1 gives the Q-Q plots of the $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$ versus the theoretical distribution, a chi-squared distribution with three degrees of freedom. The plots shows that for all three λ chosen, the power-divergence statistics performed well in the case of moderate censoring as well as in the case of no censoring.

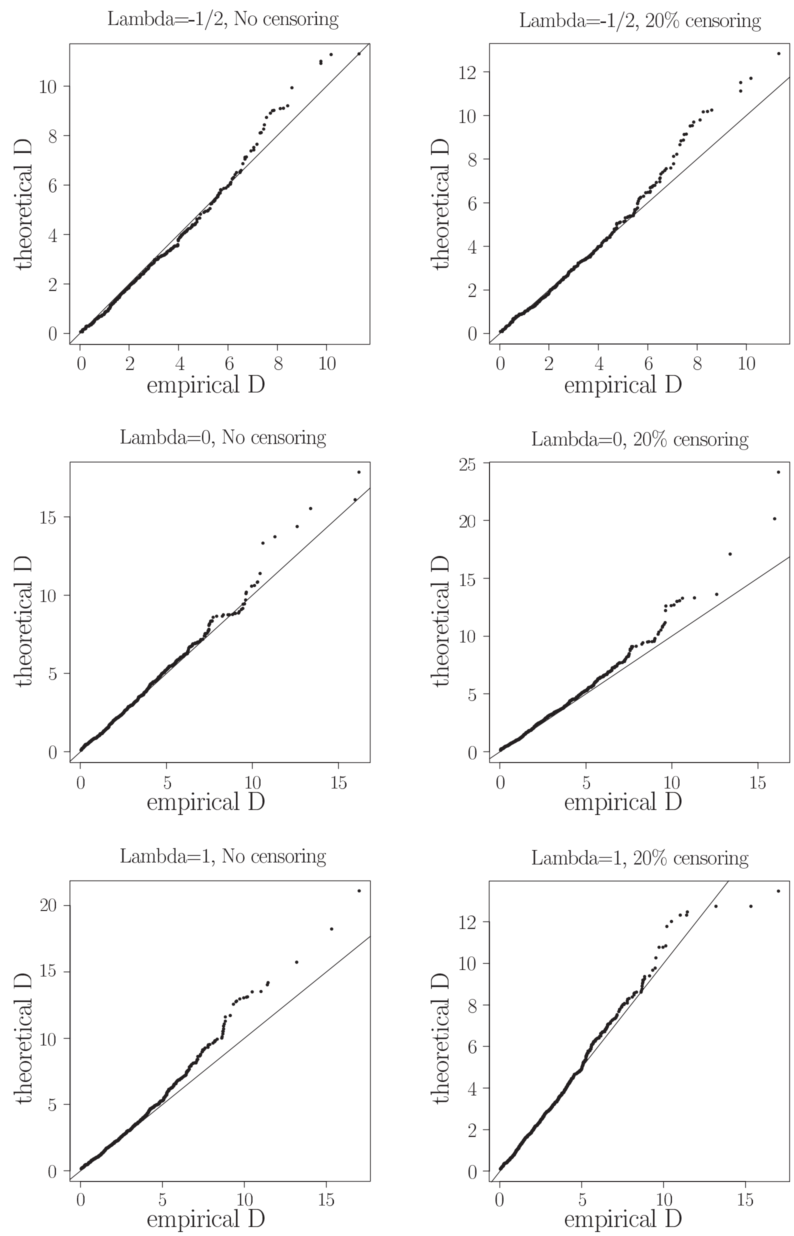

Figure 1. Plots of empirical $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$ versus χ^2 with 3 degrees of freedom. The diagonal lines are 45°. Three parameter values $\lambda = -1/2, 0, 1$ are chosen and both no censoring and 20% censoring are considered.

6. An Example

We investigated the hazard rate of the oxidizer in the reaction control system (RCS) of the Space Shuttle. The data were collected by engineers in the Johnson Space Center. The variable "soak time" was a surrogate for time-to-failure. It referred to the amount of time an oxidizer valve in the RCS was under pressure in a N_2O_4 environment. The data consisted of 258 observations, each a record of a valve used in the RCS. The design of the study was based on random right censoring. Valves entered the study at different times. After a certain period, the condition of each valve was determined. During the operation period, valves might be destroyed or fail due to inadequate maintenance or manufacturing failure. Censoring occurred when the valve failed for reasons other than Nitrate failure, or termination of study. In the data set, 153 out of 258 valves were censored.

The range of observed soak times was 2 to 9792. We took logs and grouped them into non-overlapping intervals $(\tau_{k-1}, \tau_k], k = 1, \ldots, 10$. The partition $\tau_0 < \tau_1 < \cdots < \tau_{10}$ was chosen so that the number of failures was approximately evenly distributed among the ten intervals. We first considered the Weibull. For different choices of the tuning parameter λ, ranging from -2 to 2, we calculated the corresponding power-divergence statistics and $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h})$. The values of $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h})$ ranged from 22.19 to 46.70, depending on the choice of λ. With eight degrees of freedom, the corresponding p-values were far less than 0.05. Clearly soak time was not distributed as a Weibull. Using a partition with six intervals, we found similar results.

To find a suitable parametric family, we considered $H_0^{(1)} : h \in \{h = (h_1, \ldots, h_k) | h_i = h(t_i) = \exp(\beta_0 + \beta_1 t_i + \beta_2 t_i^2)/(1 + \exp(\beta_0 + \beta_1 t_i + \beta_2 t_i^2))\}$, a model considered by Efron (1988). The hypothesis says that the discrete hazard rate h_i is quadratic after a logit transformation, i.e., $\log[h_i/(1 - h_i)] = \beta_0 + \beta_1 t_i + \beta_2 t_i^2$. The second column of Table 4 shows the power-divergence statistics under $H_0^{(1)}$. The tests reject the hypothesis at level 0.05 for all λ. With the same intervals, we tested the hypothesis $H_0^{(2)} : h \in \{h = (h_1, \ldots, h_k) | h_i = h(t_i) = \exp(\beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 t_i^3)/(1 + \exp(\beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 t_i^3))\}$, that the hazard rate was cubic logistic regression in the logarithm of soak time. Table 4 also shows the power-divergence tests under $H_0^{(2)}$. The test statistics had acceptable significance levels for all $-2.0 \leq \lambda \leq 5.0$, and minimum values with λ between 2 and 3. Note that when $|\lambda|$ becomes too large or too small, the chi-squared rejection region tends to be misleading. As the results suggest, we can accept the hypothesis that the data follow a cubic logistic model. The test statistic had minimum value when $\lambda = 2.5$. In this case, the minimum power-divergence estimates for the $\beta' s$ were $\hat\beta_0 = -2.083, \hat\beta_1 = 2.148, \hat\beta_2 = 2.266$ and $\hat\beta_3 = 1.145$, respectively.

Table 4. The power-divergence statistics $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$ under quadratic and cubic logistic models.

	Quadratic		Cubic	
λ	$D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$	*p-value*	$D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$	*p-value*
4.0	17.983	0.012	5.513	0.480
3.0	16.331	0.022	5.396	0.494
2.5	15.793	0.027	5.381	0.496
2.0	15.438	0.031	5.398	0.494
1.5	15.265	0.033	5.448	0.488
1.0	15.284	0.033	5.535	0.477
0.5	15.512	0.030	5.665	0.462
0.0	15.980	0.025	5.844	0.441

Table 4 shows the evidence that the cubic logistic model fits better than the quadratic logistic model. Note that the difference between $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta}_\lambda))$ under the hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ is also significant. This is compared to a χ_1^2 distribution since the cubic model has one more parameter than the quadratic model, and is significant at the 0.001 level. A second partition with six intervals again gave the same conclusion.

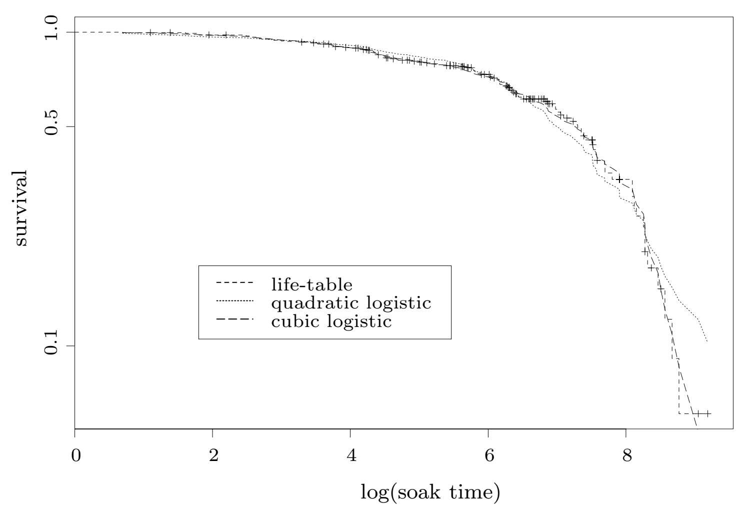

Figure 2. The survival function of soak time using life-table estimate, quadratic logistic model and cubic logistic model.

The life-table survival estimate is $\hat{S}_i = \prod_{1 \le j < i}(1 - \hat{h}_j)$ with $\hat{h}_i = d_i/n_i$. Figure 2 compares the life-table estimate versus the survival function based on the hypotheses $H_0^{(1)}$ and $H_0^{(2)}$, where the $\hat{h}_j$ are substituted for the maximum likelihood estimates obtained under the null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$, respectively. It suggests that the cubic logistic regression is a better fit for the data, consistent with our finding using power-divergence statistics.

7. Discussion and Conclusion

Since there are many power-divergence statistics, it is of interest to find a most desirable one. The selection of λ, however, is a complex issue. As pointed out by Read and Cressie (1988), there is no clear-cut conclusion to this problem and the choice depends on two criteria: efficiency and robustness. There is some discussion of these criteria for discrete multivariate data, for example, see Simpson (1987, 1989) and Lindsay (1994). In general, λ controls the trade-off between efficiencies and robustness. Read and Cressie (1988) suggested that $\lambda = 2/3$ is an excellent compromise between $X^2(\lambda = 1)$ and $G^2(\lambda = 0)$, and they recommend it for use in practice. Generally, λ with $|\lambda| > 5$ is not recommended.

There are other issues that need to be examined. In a study of the Hellinger estimator and test for categorical data, Simpson (1987, 1989) showed that the maximum likelihood estimator for discrete data is sensitive to outliers and the Hellinger distance estimator has breakdown point at 50%. The Hellinger test leads to similar results. How to extend these findings from discrete multivariate data to survival data needs to be explored. Very few results are available for robustness of tests for survival data. Robustness of the Hellinger estimator for continuous survival data was studied in Yang (1992). Further study of parallel robustness results for survival data would be valuable, especially in situations where the data is continuous or grouped subject to censoring. Investigation of this topic is still in progress.

Appendix

Lemma 1. *Suppose that* (2.1) *holds and* $n_i \to \infty$ *for* $i = 1, \ldots k$. *Let* $X_i = (d_i - n_i h_{0i})/\sqrt{n_i h_{0i}(1 - h_{0i})}$, *then* $(X_1, \ldots, X_k) \xrightarrow{D} N(O, I_k)$. *If further,* $n_i/n \to r_i \geq 0$, $i = 1, \ldots, k$, *and the* h_{0i} *in* (2.1) *are replaced by* $h_i = h_{0i} + c_i/\sqrt{n_i}$, *where* c_i *are constants, then* $(X_1, \ldots, X_k) \xrightarrow{D} N(\psi, I_k)$ *with* $\psi = (c_1/\sqrt{(h_{01}(1 - h_{01}))}, \ldots, c_k/\sqrt{(h_{0k}(1 - h_{0k}))}\ )$.

Proof. By induction, we show that the characteristic function ϕ of $(X_1, \ldots, X_K)$ satisfies

$$\phi(t_1, \ldots, t_k) = E\Big[\exp\Big(\sum_{i=1}^{k} i t_i X_i\Big)\Big] \to \exp\Big(-\sum_{i=1}^{k} t_i^2/2\Big). \tag{A.1}$$

Thus, by the Lévy Continuity Theorem, the first part of Lemma 1 holds. By the DeMoivre-Laplace Central Limit Theorem, (A.1) holds if $k = 1$. Suppose it holds in the case of $k-1$. Notice that

$$\phi(t_1,\ldots,t_k) = E\Big\{\exp\Big(\sum_{i=1}^{k-1} it_iX_i\Big)E\Big[\exp(it_kX_k)|X_1,\ldots,X_{k-1}\Big]\Big\}.$$

By the assumption that conditioning on $\mathcal{F}_k$, d_k is binomial $b(n_k,h_{0k})$, and the DeMoivre-Laplace Theorem, $E\left[\exp(it_kX_k)|X_1,\ldots,X_{k-1}\right] = \exp(-t_k^2/2)+o(1)$. Thus by the Lebesgue Dominated Convergence Theorem,

$$\begin{aligned}\phi(t_1,\ldots,t_k) &= \exp(-t_k^2/2)E\Big[\exp\Big(\sum_{i=1}^{k-1} it_iX_i\Big)\Big] + o(1)\\ &= \exp\Big(-\frac{1}{2}\sum_{i=1}^{k} t_i^2\Big)+o(1).\end{aligned}$$

Hence (A.1) holds. The second part of Lemma 1 can be shown by observing that

$$\frac{d_i-n_ih_{0i}}{\sqrt{n_ih_{0i}(1-h_{0i})}} = \frac{d_i-n_ih_i}{\sqrt{n_ih_i(1-h_i)}}\frac{\sqrt{n_ih_i(1-h_i)}}{\sqrt{n_ih_{0i}(1-h_{0i})}} + \frac{\sqrt{n_i}c_i}{\sqrt{n_ih_{0i}(1-h_{0i})}}.$$

Since $h_i\to h_{0i}$, the asymptotic normality of $(X_1,\ldots,X_k)$ follows.

Proof of Theorem 1. We first show that for $\lambda=1$,

$$\begin{aligned}D_1(\mathbf{d},\mathbf{n};\mathbf{h}_0) &= \sum_{i=1}^{k}\Big\{\frac{(d_i-n_ih_{0i})^2}{n_ih_{0i}}+\frac{(d_i^c-n_ih_{0i}^c)^2}{n_ih_{0i}^c}\Big\}\\ &= \sum_{i=1}^{k}\frac{d_i^2h_{0i}^c-2d_in_ih_{0i}h_{0i}^c+n_i^2h_{0i}^2h_{0i}^c}{n_ih_{0i}h_{0i}^c}\\ &\quad+\sum_{i=1}^{k}\frac{h_{0i}(d_i^c)^2-2n_ih_{0i}h_{0i}^cd_i^c+n_i^2h_{0i}(h_{0i}^c)^2}{n_ih_{0i}h_{0i}^c}\\ &= \sum_{i=1}^{k}\frac{\Big(d_i-n_ih_{0i}\Big)^2}{n_ih_{0i}h_{0i}^c},\end{aligned}$$

which converges to χ_k^2 by Lemma 1. Next, assume that $\lambda\notin\{0,-1\}$. By definition of $D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}_0)$, we obtain

$$\begin{aligned}D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}_0) &= \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{d_i\Big[\Big(\frac{d_i}{n_ih_{0i}}\Big)^\lambda-1\Big]+d_i^c\Big[\Big(\frac{d_i^c}{n_ih_{0i}^c}\Big)^\lambda-1\Big]\Big\}\\ &= \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{n_ih_{0i}\Big[\Big(\frac{d_i}{n_ih_{0i}}\Big)^{\lambda+1}-1\Big]\Big\}\end{aligned}$$

$$
\begin{aligned}
&+\frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{n_i h_{0i}^c\Big[\Big(\frac{d_i^c}{n_i h_{0i}^c}\Big)^{\lambda+1}-1\Big]\Big\}\\
&=\frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{n_i h_{0i}\Big[\Big(1+\frac{d_i-n_i h_{0i}}{n_i h_{0i}}\Big)^{\lambda+1}-1\Big]\Big\}\\
&\quad+\frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{n_i h_{0i}^c\Big[\Big(1+\frac{d_i^c-n_i h_{0i}^c}{n_i h_{0i}^c}\Big)^{\lambda+1}-1\Big]\Big\}.
\end{aligned}
$$

Let $u_i = (d_i - n_i h_{0i})/n_i h_{0i}, v_i = (d_i^c - n_i h_{0i}^c)/n_i h_{0i}^c$ and expand in a Taylor series to get

$$
\begin{aligned}
D_\lambda(\mathbf{d},\mathbf{n};\mathbf{h}_0) &= \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{n_i h_{0i}\Big[(\lambda+1)u_i+\frac{\lambda(\lambda+1)}{2}u_i^2+O_p(u_i^3)\Big]\Big\}\\
&\quad+\frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{k}\Big\{n_i h_{0i}^c\Big[(\lambda+1)v_i+\frac{\lambda(\lambda+1)}{2}v_i^2+O_p(v_i^3)\Big]\Big\}\\
&=\sum_{i=1}^{k} n_i\Big\{h_{0i}u_i^2+h_{0i}^c v_i^2+o_p(\frac{1}{n_i})\Big\}\\
&=\sum_{i=1}^{k}\Big\{\frac{(d_i-n_i h_{0i})^2}{n_i h_{0i}}+\frac{(d_i^c-n_i h_{0i}^c)^2}{n_i h_{0i}^c}\Big\}+o_p(1)\\
&=D_1(\mathbf{d},\mathbf{n};\mathbf{h}_0)+o_p(1),
\end{aligned}
$$

which converges to χ_k^2.

The cases of $\lambda = 0$ and $\lambda = 1$ can be proved by slightly modifying the Taylor expansion, hence the details are omitted here. Furthermore, the arguments can be directly applied to the case that h_{0j} is replaced by $h_j = h_{0j} + c/\sqrt{n_j}$ using the second part of Lemma 1.

Proof of Theorem 2. Using an expansion similar to the proof in Theorem 1, it can be shown that

$$
D_\lambda(\theta)=\sum_{i=1}^{k}\Big\{n_i\Big(\frac{h_i(\theta_0)-h_i(\theta)}{\sqrt{h_i(\theta)h_i^c(\theta)}}\Big)^2+o(n_i)\Big\}.
$$

Furthermore from some elementary probability arguments, for some constant M,

$$
\sup_{||\theta-\theta_0||\le M}\Big|\frac{D_\lambda(\theta)}{n}-\sum_{i=1}^{k}\gamma_i\Big(\frac{h_i(\theta_0)-h_i(\theta)}{\sqrt{h_i(\theta)h_i^c(\theta)}}\Big)^2\Big|=o(1).
$$

Since $h(\theta)$ is continuous, for any ϵ,

$$
\inf_{||\theta-\theta_0||>\epsilon}\sum_{i=1}^{k}\gamma_i\Big(\frac{h_i(\theta_0)-h_i(\theta)}{\sqrt{h_i(\theta)h_i^c(\theta)}}\Big)^2>0.
$$

Hence by the definition of $\hat{\theta}_\lambda$, in a neighborhood of θ_0, $\hat{\theta}_\lambda$ is uniquely defined and $\hat{\theta}_\lambda \to \theta_0$ almost surely.

By definition, the minimum dispersion estimator $\hat{\theta}_\lambda$ satisfies $\nabla D_\lambda(\theta)|_{\theta=\hat{\theta}_\lambda} = 0$. It follows $\nabla D_\lambda(\theta)|_{\theta=\hat{\theta}_\lambda} = \nabla D_\lambda(\theta)|_{\theta=\theta_0} + \nabla^2 D_\lambda(\theta)|_{\theta=\theta^*}(\hat{\theta}_\lambda - \theta_0)$, for some θ^* between θ_0 and $\hat{\theta}_\lambda$. Then the score

$$\begin{aligned}
\nabla D_\lambda(\theta)|_{\theta=\theta_0} &= \frac{2}{\lambda+1}\sum_{i=1}^{k} n_i \nabla h_i(\theta_0)\Big\{\Big[\frac{d_i^c}{n_i h_i^c(\theta_0)}\Big]^{\lambda+1} - \Big[\frac{d_i}{n_i h_i(\theta)}\Big]^{\lambda+1}\Big\} \\
&= 2\sum_{i=1}^{k} n_i \nabla h_i(\theta_0)\Big\{\frac{d_i^c}{n_i h_i^c(\theta_0)} - \frac{d_i}{n_i h_i(\theta)} + O_p(n_i^{-1})\Big\} \\
&= 2\sum_{i=1}^{k} \frac{\nabla h_i(\theta_0)}{h_i(\theta_0)(1-h_i(\theta_0))}\Big\{h_i(\theta_0) - \frac{d_i}{n_i}\Big\} + O_p(1).
\end{aligned}$$

One can further show that

$$\nabla^2 D_\lambda(\theta)|_{\theta=\theta^*} = 2\sum_{i=1}^{k} \frac{n_i \nabla h_i(\theta_0)\nabla h_i(\theta_0)^T}{h_i(\theta_0)(1-h_i(\theta_0))} + o_p(n). \tag{A.2}$$

Hence, by the assumption that $h_i(\theta)$ are twice differentiable in Θ, it is not difficult to show that

$$\begin{aligned}
&\hat{\theta}_\lambda - \theta_0 \\
&= \Big\{\sum_{i=1}^{k} \frac{n_i \nabla h_i(\theta_0)\nabla h_i(\theta_0)^T}{h_i(\theta_0)(1-h_i(\theta_0))} + o_p(n)\Big\}^{-1}\Big\{\sum_{i=1}^{k} \frac{n_i \nabla h_i(\theta_0)\Big(d_i/n_i - h_i(\theta_0)\Big)}{h_i(\theta_0)(1-h_i(\theta_0))} + O_p(1)\Big\} \\
&= \Big\{\sum_{i=1}^{k} \frac{n_i \nabla h_i(\theta_0)\nabla h_i(\theta_0)^T}{n h_i(\theta_0)(1-h_i(\theta_0))}\Big\}^{-1}\Big\{\sum_{i=1}^{k} \frac{n_i \nabla h_i(\theta_0)\Big(d_i/n_i - h_i(\theta_0)\Big)}{n h_i(\theta_0)(1-h_i(\theta_0))}\Big\} + o_p(n^{-1/2}).
\end{aligned}$$

Since conditioning on $\mathcal{F}_i, d_i$ has binomial distribution $b(n_i, h_i(\theta_0))$, $\sqrt{n}(\hat{\theta}_\lambda - \theta_0)$ converges in distribution to $N(0, V^{-1})$. The asymptotical normality of $\hat{V}_n^{1/2}(\hat{\theta}_\lambda - \theta_0)$ thus follows.

Proofs of Theorem 3 and Corollary 4. By a Taylor expansion of $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta_0))$ at $\hat{\theta}$,

$$\begin{aligned}
D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta_0)) = D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta})) &+ \nabla D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta))|_{\theta=\hat{\theta}}(\theta_0 - \hat{\theta}) \\
&+ \frac{1}{2}(\theta_0 - \hat{\theta})^T \nabla^2 D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta))|_{\theta=\theta^*}(\theta_0 - \hat{\theta}).
\end{aligned}$$

Note that by Theorem 2 and the definition of $\hat{\theta}$, $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta_0)) = \chi_k^2 + o_p(n^{-1/2})$ and $\nabla D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta))|_{\theta=\hat{\theta}} = 0$. Also from (A.2) and the Central Limit Theorem,

$(\theta_0 - \hat{\theta})^T \nabla^2 D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\theta))|_{\theta=\theta^*}(\theta_0 - \hat{\theta})/2 = \chi_p^2 + o_p(n^{-1/2})$. By a quadratic form decomposition theorem (see, e.g., Rao(1973, p.187)), $D_\lambda(\mathbf{d}, \mathbf{n}; \mathbf{h}(\hat{\theta})) = \chi_{k-p}^2 + o_p(n^{-1/2})$, hence Theorem 3 follows. The Corollary can be easily shown by the same arguments.

Acknowledgements

The authors would like to express their appreciation to the editors, an associate editor and the referees for their constructive comments that improved the presentation of the article, and to Dr. Richard Heydorn of the Johnson Space Center, NASA, for providing the Space Shuttle data.

References

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

Barlow, R. E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing.* 2nd edition. Holt, Rinehart and Winston, New York.

Cox, D. R. (1975) Partial likelihood. *Biometrika* **62,** 269-276.

Cox, D. R. and Oaks, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.

Cressie, N. A. C. and Read, T. R. C.(1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B.* **46,** 440-464.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *J. Amer. Statist. Assoc.* **83,** 414-425.

Gail, M. H. and Ware, J. H. (1979). Comparing observed life table data with a known survival curve in the presence of random censorship. *Biometrics* **35**, 385-391.

Guo, S. W. and Lin, D. Y. (1994). Regression analysis of multivariate grouped survival data. *Biometrics* **50**, 632-639.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Lai, K. J. (1993). On logistic regression approach to survival data and power divergence statistics for life tables. Ph.D. Thesis, University of Illinois, Department of Statistics.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York.

Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Statist.* **22**, 1081-1114.

Nelson, W. (1990). *Accelerated Testing : Statistical Models, Test Plans and Data Analysis.* Wiley, New York

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to Breast Cancer Data. *J. Biometrics* **34,** 57-67.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* 2nd edition. Wiley, New York.

Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer, New York.

Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *J. Amer. Statist. Assoc.* **82**, 802-807.

Simpson, D. G. (1989). Hellinger deviance tests: efficiency, breakdown points, and examples. *J. Amer. Statist. Assoc.* **84**, 107-113.

Yang, S. (1991). Minimum Hellinger distance estimation of parameter in the random censorship model. *Ann. Statist.* **19,** 579-602.

Department of Mathematical Sciences, Michigan Technological University, 1400 Townsend Dr. Houghton, MI 49931, U.S.A.

E-mail: hschen@mtu.edu

Department of Statistics, Feng-Chia University, Taichung 407, Taiwan.

E-mail: kjlai@fcu.edu.tw

Department of Statistics, Columbia University, MC4403, 2990 Broadway, NY, NY 10027, U.S.A.

E-mail: zying@stat.columbia.edu

(Received September 2001; accepted May 2003)