ASYMPTOTIC EFFICIENCY OF RELATIVE RISK ESTIMATES
Janet M. Begun and Jon A. Wellner
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

and

Department of Statistics
University of Rochester
Rochester, N.Y. 14627

A representation theorem, similar in spirit to those of Hájek (1970) and Beran (1977), is established for nonparametric regular estimates of the relative risk parameter \( \theta \) in the special two-sample case of Cox's (1972) proportional hazards model \( (1-G) = (1-F) \theta \). This representation theorem makes precise the heuristic nonparametric efficiency calculations of Efron (1977) in this special case, and lends credence to Efron's more general conclusions. From the point of view of adaptive estimation, the theorem implies that it is impossible to "adapt" to the shape of \( F \) in the proportional hazards problem.

AMS Subject Classification: 62E20, 62G05, 62G20

KEY WORDS AND PHRASES: Adaptive inference, censored data, Cox's model, efficient estimates, proportional hazards, regular estimates, relative risk, representation theorem.

1. INTRODUCTION

Let \( X_1, \ldots, X_n \) be i.i.d. random variables with continuous df \( F \), and let \( Y_1, \ldots, Y_n \) be i.i.d. random variables with df \( G \) where \( 1-G = (1-F) \theta \) and \( 0 < \theta < \infty \). The two parameters \( \theta \) and \( F \) are, in general, both unknown. This model is a particular Lehmann alternative; see e.g. Savage (1956). It is easily seen that, for absolutely continuous \( F \) with density \( f \),

\[ r_G = \theta r_F \]

where \( r_F = f/(1-F) \) and \( r_G = g/(1-G) \) are the respective hazard functions and \( g \) is the density of \( G \). The parameter \( \theta \) is therefore called the "relative risk".

The above model is simply a very special case of Cox's (1972) well-known "proportional hazards model", with \( \theta = e^\beta \) where \( -\infty < \beta < \infty \) in Cox's notation. Cox's model, with censoring and covariates, has been studied by several authors including Cox (1972) and (1975), Efron (1977), and Oakes (1977); also see Chapters 4 and 5 of Kalbfleisch and Prentice (1980). Efron (1977), in particular, carried out parametric efficiency calculations which "roughly" suggest that inferences based on Cox's partial likelihood are asymptotically fully efficient in the nonparametric problem; see Efron (1977), page 560.

Our purpose here is to use the methods of Hájek (1970) and Beran (1977a) and (1977b) to make the "admittedly rough" results of Efron (1977) precise in the above two-sample case of Cox's model. The main result, given in Section 2, is a representation theorem in the spirit of Hájek (1970) and Beran (1977). It asserts that all suitably regular estimates of \( \theta \) in the nonparametric model must...
be asymptotically more dispersed than \( N(0, \sigma^2 \omega^2) \), in the precise sense of Theorem 1, where \( \sigma^2 \omega^2 = 1/\text{asymptotic information from the Cox partial likelihood in the two-sample problem} \), and \( \omega^2 = \sigma^2(\omega) \geq 1 \) always (equality holds only if \( \omega = 1 \)).

In the two-sample case, Cox's partial likelihood reduces to the likelihood of the ranks, (see Kalbfleisch and Prentice (1973) or Begun (1982)) which depends only on the relative risk \( \theta \) and not on the continuous df \( F \). Consequently, Cox's estimate of \( \theta \) is, in the two-sample case, the Maximum Likelihood Estimate (MLE) based on the ranks. The representation theorem proved by Begun (1982) shows that every regular rank estimator of \( \theta \) has a limiting distribution more dispersed than \( N(0, \theta^2 \omega^2) \). Theorem 1 here shows that the same is true for all regular estimates of relative risk, including those that depend on the magnitudes of the \( X \)'s and \( Y \)'s such as

\[
\hat{\theta}_{m,n} = \int_{-\infty}^{\mu} -\log(1-G_n)/-\log(1-F_m) d\mu
\]

where \( F_n \) and \( G_n \) denote the empirical df's of the two samples and \( \mu \) is a fixed probability measure on \( \mathbb{R} \), or

\[
\hat{\theta}_{m,n} = \int_{-\infty}^{\infty} -\log(1-G_n) d\nu/\int_{-\infty}^{\infty} -\log(1-F_m) d\nu
\]

where \( \nu \) is a fixed probability measure with \( \int_{-\infty}^{\infty} [-\log(1-F)] d\nu < \infty \), or the estimates of Steck and Zimmer (1972), and Steck, Zimmer, and Williams (1974).

If \( 1-F = (1-F_n)^\gamma \) in the above two-sample model where \( F_n \) is a fixed known df and \( 0 < \gamma < \infty \) is unknown, then it is easily seen: (a) that the MLE of \( \theta \) is \( \hat{\theta}_0 = \int_{-\infty}^{\infty} \frac{\log(1-F_n)}{\log(1-F_n)} dG_n/\int_{-\infty}^{\infty} \log(1-F_n) dG_n \); and (b) that \( \hat{\theta}_0 \) has asymptotic variance \( \sigma^2 \leq \theta^2 \omega^2 \). More precisely, with \( N \equiv m n \),

\[
(mn/N)^{1/2} (\hat{\theta}_0 - \theta) \xrightarrow{d} N(0, \sigma^2)
\]

as \( mn \to \infty \). Thus our main result may also be interpreted from the point of view of "adaptive estimation" as follows: In the two-sample proportional hazards model it is not possible to estimate the relative risk parameter "adaptively", i.e. as if \( F_n \) (or \( F \)) were known. The information contained in the \( X \)'s and \( Y \)'s about \( F \) cannot be used to "estimate away" the nuisance function \( F_n \) and thereby estimate \( \theta \) as if it were known. The optimal rank estimates are asymptotically best among all regular estimates of \( \theta \).

2. THE MAIN RESULT

Suppose that \( \omega = \mathbb{R}^d \) where \( 1 \leq \omega < \infty \) (without loss of generality; if \( 0 < \omega < 1 \), write \( F = F_1/\omega \) and relabel the two populations), \( F \) is absolutely continuous with density \( f \), and \( F \equiv 1-F \) for any df \( F \). Let \( N \equiv m n \), \( \lambda_n \equiv m/n, \lambda_N = 1/\lambda_n \). We will suppose throughout that \( \lambda_n \to \lambda \in (0,1) \) as \( mn \to \infty \), write \( X = 1-\lambda \), and use \( I \) to denote both the identity function and Lebesgue measure on \( (-\infty, \infty) \).
Let \( \| \cdot \| \) denote the usual norm in \( L^2 = L^2(\mathbb{R}, 1) \). Let \( \mathcal{C}(f, \alpha) \) denote the set of all sequences of densities \( \{ f_N \} \) such that

\[
\lim_{m,n} \alpha \| (mn/N)^{1/2}(f_N^m - f_N^n) - \alpha \| = 0
\]

(2.1)

where \( \alpha \in L^2 \) and \( f \in D \equiv \{ \text{all density functions on } (-\infty, \infty) \} \). Note that (2.1) implies that \( \alpha \) is orthogonal to \( f^2 \) since \( \| f_N^2 \| = 1 = \| f^2 \| \) for all \( N \geq 1 \). Let \( \mathcal{C}(f) \) denote the union over \( \alpha \) of all sets \( \{ \mathcal{C}(f, \alpha) : \alpha \in L^2, \alpha \perp f^2 \} \). When \( \text{support}(\alpha) \) is contained in \( \text{support}(f) \) we denote \( \mathcal{C}(f, \alpha) \) by \( \mathcal{C}_0(f, \alpha) \) and let \( \mathcal{C}_0(f) = \bigcup \mathcal{C}_0(f, \alpha) : \alpha \in L^2, \alpha \perp f^2, \text{support}(\alpha) \subset \text{support}(f) \). The sequences \( \{ f_N \} \) in \( \mathcal{C}_0(f) \) consist essentially of densities \( f_N \) with support contained in that of \( f \); if \( f \) is everywhere positive, \( \mathcal{C}_0(f) = \mathcal{C}(f) \). Finally, let \( \Theta(s, h) \) denote the set of all sequences \( \{ \theta_N \} \) such that

\[
\lim_{m,n} \Theta(s, h) \| (mn/N)^{1/2}(\theta_N - \theta) \| = 0
\]

where \( \theta = 1, h \in \mathbb{R} \), and let \( \Theta(s) = \bigcup \Theta(s, h) : h \in \mathbb{R} \). Given \( \{ f_N \} \in \mathcal{C}_0(f) \) and \( \{ \theta_N \} \in \Theta(s) \), define the corresponding sequence of densities \( \{ g_N \} \) by

\[
g_N = \theta_N^T f_N.
\]

(2.3)

We say that an estimator \( \hat{\theta}_{m,n} \) of \( \theta \) is regular at \( (f, \alpha) \) (or \( \mathcal{C}_0 \)-regular at \( (f, \alpha) \)) if for every \( \{ f_N \}, \{ \theta_N \} \in \mathcal{C}_0(f) \times \Theta(s) \) and for \( X_m = (X_1, \ldots, X_m) \) iid \( f_N \), \( Y_n = (Y_1, \ldots, Y_n) \) iid \( g_N \), the distribution of \( (mn/N)^{1/2}(\theta_{m,n} - \theta) \) converges weakly to a law \( L = L(f, \alpha) \) that depends only on \( f \) and \( \theta \) and not upon the particular sequences \( \{ f_N \} \) and \( \{ \theta_N \} \). Thus \( L = L(f, \alpha) \) does not depend on \( s \) or \( h \) (although it will, of course, depend on the particular estimator \( \hat{\theta}_{m,n} \)). This is a very desirable stability property which is precisely in the spirit of Hájek (1970), and Beran (1977a), (1977b). It is easily seen to be implied by uniform weak convergence in neighborhoods of \( f \) and \( \alpha \).

Finally, before stating our theorem, we introduce an important function \( J_* \) and a related quantity \( \sigma_*^2 \). Define \( J_* : [0,1] \to \mathbb{R}^+ \) by

\[
J_*(s,t) = 1/(\lambda(1-s) + \theta \lambda(1-t)),
\]

(2.4)

and define \( \sigma_*^2 = \sigma_*^2(\theta) \) by

\[
1/\sigma_*^2 = \int J_*(f, \theta)(1-F)dG
\]

\[
= \int_0^1 [1 + \theta \alpha_x/(\theta - 1)/\gamma_x]dx
\]

(2.5)

by the change of variables \( 1 - \theta x = \gamma \). Thus \( \sigma_*^2 \) is a function of only \( \theta \) and \( \lambda \); it has appeared before in Efron (1977, page 561) and Begun (1980).
THEOREM 1. Suppose that \( \hat{\theta}_{m,n} \) is a \( C_0 \)-regular estimator of \( \theta \) at \((f, \theta)\) in the proportional hazards model with limit law \( L = L(f, \theta) \). Then \( L \) may be represented as the convolution of a \( N(0, \sigma^2 \cdot F) \) distribution with \( L_1 = L_1(f, \theta) \), a distribution depending only on \((f, \theta)\). Thus if \( S \sim L \) and \( W \sim L_1 \) independent of \( N(0, \sigma^2 \cdot F) \), then \( S \overset{d}{=} N(0, \sigma^2 \cdot F) + W \).

Theorem 1 can be extended to allow for random censorship as follows: Suppose that \( U_1, \ldots, U_n \) are i.i.d. censoring times with density \( c \) and that \( V_1, \ldots, V_n \) are i.i.d. censoring times with density \( d \) and that the pairs \((Y_i^0, \xi_i)\), \( i = 1, \ldots, n \) and \((V_i, \zeta_i)\), \( i = 1, \ldots, n \) are observed where \( Y_i^0 = \min(X_i, U_i) \), \( \xi_i = 1[Y_i^0 = X_i] \), \( V_i = \min(Y_i, V_i) \), \( \zeta_i = 1[Y_i^0 = Y_i] \).

Let
\[
F^0 = \frac{d}{dF}, \quad F^u = \int_{-\infty}^\cdot \frac{d}{dF} \, dF,
\]
and
\[
\sigma^2 = \frac{d}{d\sigma}, \quad \sigma^u = \int_{-\infty}^\cdot \frac{d}{d\sigma} \, d\sigma.
\]

and define \( \sigma^2 \) by
\[
1/\sigma^2 = \int_{-\infty}^\theta \left( J_*(F^0, \sigma^0) (1 - F^0) \right) d\sigma^u.
\]

Then Theorem 1 continues to hold with \( \sigma^2 \) defined by (2.8) rather than (2.5). The proof of this extension of Theorem 1 involves somewhat more complicated versions of Lemmas 1 - 3 below. In particular, in the censored cases \( \alpha^* \) and \( \beta^* \) of Lemma 3 must be replaced by
\[
\alpha^*(\text{censored}) = -\frac{1}{2} \chi^2(\int J_*(F^0, \sigma^0) \sigma^0 \left[ \int J_*(F^0, \sigma^0) \sigma^0 \right] \frac{dF}{1-F})
\]
and
\[
\beta^*(\text{censored}) = \frac{1}{2} \frac{1}{\beta_0} \left[ 1 + \log \frac{1}{\beta_0} \right] - \frac{1}{\beta_0} \left[ \int J_*(F^0, \sigma^0) \sigma^0 \right] \frac{d\sigma^0}{1-\sigma^0}
\]

and \( \chi^2 \) of Lemma 2 below must be replaced by
\[
\chi^2(\text{censored}) = 4(1 - \frac{\beta}{\beta_0})^2 \left( \frac{\|\sigma^0\|^2}{\|\sigma^0\|^2} + \frac{\|\sigma^0 - \beta^0\|}{\|\beta^0\|^2} \right)
\]
\[
+ 4(1 - \frac{\beta}{\beta_0}) \left( \frac{\|\sigma^0\|^2}{\|\sigma^0\|^2} + \frac{\|\sigma^0 - \beta^0\|}{\|\beta^0\|^2} \right).
\]

where \( A = \int_{-\infty}^\cdot \frac{d}{d\sigma} \, dI \) and \( B = \int_{-\infty}^\cdot \frac{d}{d\sigma} \, dI \).

3. PROOF OF THE MAIN RESULT

We begin with three key lemmas. Proofs of Lemmas 1 and 3 are given in Section 4.

**LEMMA 1.** If \( (f, h) \in C_0(f, \alpha) \), \( (\theta, h) \in \mathcal{B}(\theta, h) \) with \( \theta \geq 1 \), and \( (g_i) \) is defined by
(2.3), then
\[
\lim_{m,n} = m (mn/N)^b \left[ \frac{N}{N} g^b - b \right] = 0
\]
where
\[
\beta = g^b \left( \frac{h}{2b} [1 + \log B] + a \right) - (a - 1) \int_0^\infty \gamma^b \gamma^2 \, d\gamma
\]
(3.1)
It is easily checked that \( \beta \cong g^b \). Now define
\[
\eta_{m,n} = \log \left\{ \prod_{i=1}^m f_N(X_i) \prod_{j=1}^n g_N(Y_j) \right\}
\]
(3.2)
\[
= 2 \log \left( \prod_{i=1}^m \frac{g_N(X_i)}{f_N(X_i)} \prod_{j=1}^n \frac{g_N(Y_j)}{g(Y_j)} \right)
\]
**Lemma 2.** If \( (f_N) \in C_o(f, \alpha), (g_N) \in C_o(g, \beta) \) with \( 2 \geq 1 \) where \( \alpha \in L_2, \beta \in R \), then, for every \( \epsilon > 0 \),
\[
\lim_{m,n} = \prod_{i=1}^m f_N(X_i) \prod_{j=1}^n g_N(Y_j) \left\{ \left[ -2 \gamma^b \sum_{i=1}^m a(X_i) f^{-b}(X_i) \right] \right. \\
- \left. 2 \gamma^b \sum_{j=1}^n b(Y_j) g^{-b}(Y_j) + b \gamma^2 \right\} > \epsilon
\]
where \( \gamma^2 = 4x^{-1} \| \alpha \|^2 + 4x^{-1} \| \beta \|^2 \), and \( \beta \) is given in Lemma 1. In particular
it follows that \( \eta_{m,n} \rightarrow N(-mN^2, \gamma^2) \) under \( (f, \alpha) \) as \( m \rightarrow \infty, \lambda_N \rightarrow \lambda \).

Lemma 2 is an easy two-sample extension of the Lemma given on page 439 of
Beran (1977b). It is easily deduced from Lemma 1 and LeCam's second lemma; see
LeCam (1969) or Hájek and Šidák (1967).

Note that \( \gamma^2 \) in Lemma 2 depends upon \( \alpha \in L^2 \) and \( h \in R \); the explicit dependence is given by way of Lemma 1. In the following lemma we compute the value of
\( \gamma^2 \) for a particular choice of \( \alpha \in L^2 \). Define \( \alpha_\ast \in L^2 \) by
\[
\alpha_\ast = -\frac{1}{2} \int f^2 \left( J_\ast(F,G) \right) - \int J_\ast(F,G) \arctan \left( \frac{dF}{1-dF} \right) \frac{dF}{1-F}
\]
(3.3)
where \( J_\ast \) is given in (2.4); denote the \( \beta \) given in (3.1) corresponding to \( h_{\alpha_\ast} \) by \( h\alpha_\ast \). Note that support(\( \alpha_\ast \)) \( \subset \) support(\( f \)).

**Lemma 3.** If \( \alpha = h\alpha_\ast \), then
\[
\beta = h\alpha_\ast = h h_{\alpha_\ast} \left( \frac{1}{2} \left[ 1 + \log B \right] - \tau \left( J_\ast(F,G) \right) \arctan \left( \frac{dF}{1-dF} \right) \right)
\]
(3.4)
\[ y^2 = \gamma_0^2 = h^2u^2 \quad (3.5) \]

where \( \gamma_0^2 \) is given by (2.5).

**Proof of Theorem 1.** By (2.1), (2.2), and Lemma 2 the characteristic function of 
\[(mN)^2(\hat{\alpha}_{m,n} - \theta_N) \quad \text{under } (f, \theta) \]
is
\[
E_{f_N, \theta} \exp \left[ iu(mN)^2(\hat{\alpha}_{m,n} - \theta_N) \right] = E_{f_N, \theta} \exp \left[ iu(mN)^2(\hat{\alpha}_{m,n} - 0) - iuh \right] + o(1)
\]
\[
= E_{f_N, \theta} \exp \left[ iu(mN)^2(\hat{\alpha}_{m,n} - 0) - iuh \right] + L_{m,n} - iuh \quad (3.6)
\]
the latter being true for all \( \alpha \in L^2, \alpha \in \mathcal{V}, \) support(\( \alpha \)) = support(\( f \)) and \( h \in \mathbb{R} \).

We choose \( \alpha = \nu \alpha, \) where \( \alpha \) is given in (3.3). By considering only a subsequence if necessary, we may assume that under \((f, \theta)\) the random vectors \(((mN)^2(\hat{\alpha}_{m,n} - 0), \ldots, \lambda_N^{-1}n^{-1/2} \alpha(X_i) - \lambda_N^{-1}n^{-1/2} \alpha(Y_j))\) converge weakly to a random vector \((S, Z)\) such that \( Z \sim N(0, \nu^2) \).

It then follows from Lemma 2 that the random vectors \(((mN)^2(\hat{\alpha}_{m,n} - 0), L_1)\) converge weakly under \((f, \theta)\) to the random vector \((S, 2hZ - \nu^2)\).

Hence, by regularity of \( \hat{\alpha}_{m,n} \) the characteristic function (3.6) converges to 
\[ E \exp \left[ iuS + 2hZ - \nu^2 - iuh \right] \]
while (3.7) converges, by Vitali's theorem and an almost surely convergent construction as in Beran (1977b), to

\[ E \exp \left[ iuS + 2hZ - \nu^2 - iuh \right] \]

Hence, by letting \( \varphi(u, v) \equiv E \exp [iuS + ivZ] \) denote the characteristic function of \((S, Z)\), we have from (3.6), (3.7), and the preceding that
\[ \varphi(u, 0) = E \exp \left[ iuS + 2hZ \right] \exp [-iuh - \nu^2] \quad (3.8) \]
for all real \( h \). The right side of (3.8) is analytic in \( h \), constant for all real \( h \), hence constant for all complex \( h \). In particular, the choice \( h = iu/2 \) yields, for all real \( u,v \)
\[ \varphi(u, 0) = \frac{1}{2} \left( \gamma_0^2 \right)^{-1} \exp \left[ \frac{uv}{2} \right] \quad (3.9) \]

The choice \( v = -2(\alpha_0^2)u \) gives
\[ \varphi(u, 0) = \frac{1}{2} \left( \gamma_0^2 \right)^{-1} \exp \left[ -\frac{uv}{2} \right] \]
for all real \( u \). Since \( \varphi(u, 0) \) is the characteristic function of \( S \), or equiva-
lently of the law \( L = L(f, \theta) \), and the first factor on the right side of (3.10) is the characteristic function of \( \frac{1}{2} (\tilde{\theta})^2 \) while the second factor is the characteristic function of a \( N(0, \tilde{\theta}^2) \) rv, Theorem 1 follows.

The preceding proof used the apparently arbitrary choice \( \alpha = \alpha_0 \) with \( \alpha_0 \) given in (3.3). This choice of \( \alpha \), however, yields the strongest possible theorem. It may be of help at this point to explain briefly how the function \( \alpha \) was obtained.

Efron (1977) parameterized the proportional hazards model using the "average hazard distribution" \( F_0 \) given by
\[
F_0 = F^{\lambda} G^{1-\lambda} = \frac{F^\lambda}{G}\ .
\]
Note that \( r_{F_0} = (\lambda + \bar{\lambda}) r_F \) is the average hazard rate for \( F \) and \( G \), where \( r_0 \equiv r_{F_0} = f_0/F_0 \). He then assumed that \( r_0(t) = \exp(\gamma \cdot w(t)) \) where \( w = w(t) \) is a \( K \)-vector of \( (\text{known}) \) functions, and \( \gamma \) is a \( K \)-vector of \( (\text{unknown}) \) parameters.

This implies that \( f_0(t) = r_0(t) \exp (\int_0^t r_0(s) ds) = \exp (\gamma \cdot w(t) - \int_0^t \exp (\gamma \cdot w(s)) ds) \), and an easy computation then shows that scores for \( \gamma \),
\[
S_k(t) = \int_0^t \frac{1}{\gamma_k} \log f_0(t) = 2 \int_0^t \log f_0(t) = 2 \int_0^t \frac{1}{\gamma_k} f_0(t) f_0^{1/2}(t) \ , \quad (3.11)
\]

\( k = 1, \ldots, K \), are given by
\[
S_k(t) = w_k(t) - \int_0^t w_k(s) F_0(s)^{-1} dF_0(s) \ .
\]

Hence for any \( K \)-vector \( g \),
\[
g \cdot S(t) = g \cdot w(t) - \int_0^t g \cdot w(s) F_0(s)^{-1} dF_0(s) \ , \quad (3.12)
\]
or, by letting \( \nabla \gamma \) denote the gradient operator \( \left( \frac{\partial}{\partial \gamma_1}, \ldots, \frac{\partial}{\partial \gamma_K} \right) \) and using the expression on the right side of (3.11) for the scores,
\[
2 (g \cdot \nabla \gamma) f_0^{1/2} = g \cdot w - \int_0^t g \cdot w F_0^{-1} dF_0 \ . \quad (3.13)
\]

Finally, inspection of equation (3.13) of Efron's (1977) Lemma 1 shows that the information for \( \bar{\theta} = \log \theta \), and hence for \( \theta \), is smallest when the second term inside the expectation is zero, and a relatively straightforward calculation shows that, in the two-sample case, Efron's
\[
E_{\tilde{\theta}}[z | R(t)] - E_{\tilde{\theta}}[z] \approx P \begin{bmatrix} J_\theta(F(t), G(t)) \tilde{\theta}(t) \end{bmatrix} \Delta \quad (3.14)
\]
as \( m, n \to \infty, \lambda \to \lambda, \) where \( \Delta \equiv 1/(\lambda + \bar{\lambda}) \) and the notation on the left side is Efron's, that on the right ours. Thus we "guess" that
\[
g \cdot w(t) = \bar{\theta} J_\theta(F(t), G(t)) \tilde{\theta}(t) \Delta \quad (3.15)
\]
is a "worst possible case" with minimum information. By substituting (3.15) into the right side of (3.13), noting that the "infinite-dimensional analogue" of the left side of (3.13) is \( \int_0^\infty f^2 \), and computing the corresponding \( \alpha \), for \( f \), (by a result similar to Lemma 1), we obtain the function \( \alpha_* \) given in (3.3), up to a constant multiple.

A more general approach which justifies the present \( \alpha_* \) as yielding an "effective score" for \( \theta \) which is orthogonal to all nuisance parameter scores is given by Begun, Hall, Huang, and Wellner (1983).

4. PROOFS OF THE LEMMAS

We will denote the function \( J_\theta(F,G) \) by simply \( J_* \), write \( H \equiv \lambda F + \xi G \), and make frequent use of the identities

\[
\theta \triangle F = \overline{F} \theta G 
\]

and

\[
J_* dH = F^{-1} dF,
\]

which follow easily from \( G = F^\theta \) and the definition of \( H \).

Proof of Lemma 3. The formula (3.4) for \( \Phi_* \) follows easily from (3.1) and (3.3) after noting that

\[
\int \alpha_* F^2 dI = \int \alpha_* F^2 dI
\]

(since \( \alpha_\theta \equiv F^\theta \))

\[
= \frac{1}{\theta} \left( \int J_\theta \theta dF - \int \frac{t}{\theta} \int J_\theta \theta \frac{dF}{F} dF(t) \right)
\]

\[
= \frac{1}{\theta} \overline{F}(x) \int J_\theta \theta \frac{dF}{F}
\]

by use of Fubini's theorem on the second term, and that \( \theta F^{-1} dF = \overline{G}^{-1} dG \) by (4.1).

The next task is to compute \( \| \alpha_* \|^2 \) and \( \| \Phi_* \|^2 \). Note that for any \( \alpha \in L^2 \),

\[
\| \alpha \|^2 = \int \alpha^2 dI = (\int f^2)^2 dF,
\]

and that \( X \sim F \) (continuous) implies that

\(-\log \overline{F}(X) \sim \text{exponential} (1)\). Hence by routine computation it is easily seen that

\[
\int \overline{[1+\log \theta]}^2 dG = 1,
\]

\[
\int [J_\theta \theta - J_\theta \theta F^{-1} dF]^2 dF = \int J_\theta ^2 \theta^2 dF,
\]

and

\[
\int [J_\theta \theta - J_* \theta dG]^2 dG = \int J_* ^2 \theta^2 dG.
\]
Also,
\[ \int_{-\infty}^{\infty} \left[ 1 + \log \overline{G} \right] J_d \overline{G} \, dG - \int_{-\infty}^{\infty} J_d \overline{G} \, dG = \frac{[1 - \lambda \sigma_*^2]}{\overline{\theta}} \quad (4.6) \]
since, by Fubini applied to the second term,
\[ \int_{-\infty}^{\infty} [J_d \overline{G} - \int_{-\infty}^{\infty} J_d \overline{G} \, dG] \, dG = 0 \]
and, again by Fubini applied to the second term together with the identity
\[ \int_{-\infty}^{\infty} \log \overline{G} \, dG = \overline{G} \log \overline{G} - \overline{G} , \]
\[ \int_{-\infty}^{\infty} J_d \log \overline{G} \, dG - \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} J_d \overline{G} \right] \log \overline{G} \, dG \]
\[ = \int_{-\infty}^{\infty} J_d \overline{G} \, dG \]
\[ = \int_{-\infty}^{\infty} \left[ 1 - \lambda J_d \overline{F} \right] \, dG / \overline{\theta} \quad \text{by the definition of } J_d \]
\[ = \left( 1 - \lambda \sigma_*^{-2} \right) / \overline{\theta} \quad \text{by (2.5)} . \]

It now follows from (3.3) and (4.4) that
\[ \| \alpha_* \|^2 = \lambda \sigma_*^2 \int_{-\infty}^{\infty} J_d^2 \overline{G}^2 \, dF ; \quad (4.8) \]
and from (3.4), (4.3), (4.5), and (4.6) that
\[ \| \beta_* \|^2 = \lambda (\beta - 2 \sigma_*^{-2})^2 + \lambda^2 \int_{-\infty}^{\infty} J_d^2 \overline{G}^2 \, dG \]
\[ = \lambda \left( 2 \lambda (\beta \sigma_*^{-2}) - \beta^2 + \lambda^2 \int_{-\infty}^{\infty} J_d^2 \overline{G}^2 \, dG \right) . \]

Before combining (4.8) and (4.9) to compute \( \gamma^2 \) we note that
\[ \int_{-\infty}^{\infty} J_d^2 \overline{G}^2 \, dH = \int_{-\infty}^{\infty} J_d^2 \overline{G}^2 \, F^{-1} \, dF \quad \text{by (4.2)} \]
\[ = \theta^{-1} \int_{-\infty}^{\infty} J_d \overline{G} \, dG \quad \text{by (4.1)} \]
\[ = \theta^{-2} \chi^{-1} \left[ 1 - \lambda \sigma_*^{-2} \right] \quad \text{by (4.7)} . \]

Hence from (4.8), (4.9), and (4.10) it follows that corresponding to the particular choice of \( \alpha = h \alpha_* \) and the corresponding \( \beta = h \beta_* \) we have
\[ \gamma_*^2 = 4 \chi^{-1} \| h \alpha_* \|^2 + 4 \chi^{-1} \| h \beta_* \|^2 \]
\[ \|
\]
\[
\sum_{\lambda} h^2 \int_{\alpha} \varphi^2 d\gamma + h^2 \left[ 2 \left( \frac{\psi_0}{\alpha} \right)^2 - \lambda^{-1} \alpha^{-2} \right] \\
= h^2 (\varphi^2 \alpha^{-1})^{-1},
\]
and this completes the proof of Lemma 3.

\textbf{Proof of Lemma 1.} We begin by noting several basic facts: Since \( \gamma_N \to \gamma < (0,1) \), (2.1) is equivalent to (2.1) with \((\alpha N)^{\beta} \) replaced by \( N^{\beta} \) and \( \alpha \) replaced by \( \alpha \equiv (\alpha N)^{\beta} \), and similarly in (2.2) and the statement of Lemma 1; we shall use the normalization \( N^{\beta} \) here to simplify notation. Also, \( (f_N)^{\beta} \leq C_0(f, \alpha) \) implies that (dropping the tilde on \( \alpha \))
\[
\| N^{\beta}(F_N - F) \|_\infty \to 0
\]
where \( F_N, F \) are the df's corresponding to \( f_N, f \) and \( \| \cdot \|_\infty \) denotes the supremum norm. Thus \( g_N \to g \) in measure and since \( \| g_N^{\beta} \| = 1 = \| g^{\beta} \| \) it follows by Vitali's theorem that \( \| g_N^{\beta} - g^{\beta} \| \to 0 \).

Now let \( e_N \equiv (\alpha - 1)/2 - (\alpha - 1)/2 \equiv n \geq 0 \) since \( \alpha \geq 1 \), and use Minkowski's inequality and (3.1) to write
\[
\| N^{\beta}(g_N^{\beta} - g_N^{\beta}) - \beta \| = \| N^{\beta}(g_N^{\beta} - g_N^{\beta}) \|_{\infty} + \| g_N^{\beta} \|_{\infty} = \| g_N^{\beta} \|_{\infty} = 1 = \| g^{\beta} \|_{\infty}
\]
where the four terms \( A_N \sim D_N \) correspond to the four preceding lines. We shall show that each of these terms converges to zero as \( N \to \infty \).

First \( A_N \): by easy algebra and Minkowski's inequality
\[
A_N \leq \| [N^{\beta}(\alpha_0 - \alpha)^{\beta} - 1](\alpha_0^{\beta} - h/2\alpha)g_0^{\beta} + \| \alpha \|_\infty \| g_N^{\beta} - g^{\beta} \| \to 0
\]
since \( N^{\beta}(\alpha_0 - \alpha) = h \) and \( g_0^{\beta} \to g^{\beta} \) in \( L^2 \).
To handle $B_N$, first note that $\| F_N - F \|_{\infty} \to 0$, $r^3_N \to r^3$ in $L^2$ and $N^3(\tau_N - \tau) \to h N$
implies that
$$\sum_{N}^{\infty} \| F_N - F \|_{\infty} r^3_N \to \log F \|_{\infty} r^3$$
in measure, and by direct calculation,
$$\| N^3(\tau_N - \tau) \|_{\infty}^2 = N(\tau_N - \tau)^2 + \tau_N \|_{\infty}^2 = \| \log F \|_{\infty}^2 r^3.$$

Therefore, by Vitali's theorem it follows that
$$B_N \equiv \| N^3(\tau_N - \tau) \|_{\infty} r^3 N - \| \log F \|_{\infty} r^3 = 0.$$

That $C_N \to 0$ is easily shown by using Minkowski's inequality, (2.1), and
$$\| F_N - F \|_{\infty} \to 0,$$
$$C_N \leq \| F_N^3 - \| \| F_N^3 - F^3 \| \| F_N - F \|_{\infty} \| \alpha \| \to 0$$
$$\| F_N^3 \|_{\infty} \leq 1 \text{ and } \| F_N - F \|_{\infty} \to 0 \text{ implies } \| F_N^3 - F^3 \|_{\infty} \to 0 \text{ for } n \to 0.$$}

Finally, the (more difficult) term $D_N$ is
$$D_N \equiv \| N^3(\tau_N - \tau) - 2 N^3 \| \| \alpha \| \| \alpha \| \| \alpha \| \to 0.$$

To show that $D_N \to 0$, first note that the result is trivial when $n = 0(\alpha = 1)$, so we assume $n > 0$. Let
$$e_N = N^3(\tau_N - \tau) - \alpha \text{, } (4.11)$$
and note that $\| e_N \| \to 0$ implies that there exists a finite constant $M$ such that
$$\| \alpha + e_N \| \leq M \text{ (4.12)}$$
Thus we have both
$$\| F_N(x) - F(x) \| = \int_{X}^{\infty} (F_N^3 - F^3)(\alpha) d\tau$$
$$= \int_{X}^{\infty} N^3(\alpha + e_N)(\tau_N^3 - \tau^3) d\tau$$

(4.13)
\[ \leq N^{-b_N} [F_N(x)^b + F(x)^b]. \]

which yields
\[ |F_N(x)^b - F(x)^b| \leq N^{-b_N}, \quad (4.14) \]

and
\[
F_N(x) - F(x) = \int_0^\infty \left( \frac{r_N^b - r^b}{x} \right) (2t^b + r_N^b - r^b) dt - 2N^{-b_N} \int_0^\infty \left( \alpha + c_N \right) r_N^b dt + N^{-1} \int_0^\infty (\alpha + c_N)^2 dt. \quad (4.15)
\]

Now let \( a_N = F^{-1}(1-2M N^{-b_N}) \), where \( F^{-1} \) denotes the left-continuous inverse of \( F \). Then, by (4.11) - (4.15) and
\[ y^n - x^n = (y-x) \cdot x^{n-1} + b(y-x)^2 \cdot (n-1)x^{n-2} \]

where \( |x - x| \leq |y-x| \), we find that
\[
\frac{a_N}{2} \int \left( N^{b_N} [F_N(x) - F(x)] - \frac{2}{N} \int_0^\infty \left( \alpha + c_N \right) r_N^b dt \right)^2 \, dx
\]

\[
= \frac{a_N}{2} \int \left[ (\alpha + c_N)^2 + \left( F_N(x)^b - F(x)^b \right) \right] r_N^b \, dx
\]

\[
\leq 2N^{-1} \int \left( \left[ (\alpha + c_N)^2 + 2N^{-b_N} F(x)^b \right] \right)^{2n-2} f(x) \, dx
\]

\[
= \left( \left[ (\alpha + c_N)^2 + 2N^{-b_N} F(x)^b \right] \right)^{2n-2} f(x) \, dx
\]

where \( |F_N(x) - F(x)| \leq |F_N(x) - F(x)| \)

\[
\leq 2N^{-1} \int \left( \left[ (\alpha + c_N)^2 + 2N^{-b_N} F(x)^b \right] \right)^{2n-2} f(x) \, dx
\]

\[
= \left( \left[ (\alpha + c_N)^2 + 2N^{-b_N} F(x)^b \right] \right)^{2n-2} f(x) \, dx
\]

where \( |R(x)| \leq 1 \) and by using \( (a+b)^2 \leq 2(a^2+b^2) \) repeatedly

\[
\leq 16n^2 \left\| c_N \right\|^2 \cdot \int \frac{a_N}{2} \int F(x)^{2n-1} f(x) \, dx
\]

\[
\leq 16n^2 \left\| c_N \right\|^2 \cdot \int F(x)^{2n-1} f(x) \, dx
\]

for \( N \) sufficiently large
\[ \leq B_n \left( \| c_n \|^2 + [4 + 9(n-1)^2] n^2 N^{-1} (2n-1)^{-1} [1 - \mathbb{P}(a_n)] 2n^{-1} \right)_{[n \neq \frac{1}{2}]} \]
\[ + [4 + 9(n-1)^2] n^2 N^{-1} \log (1/\mathbb{P}(a_n)) \right)_{[n = \frac{1}{2}]} \]
\[ \rightarrow 0. \]

We also have, by Cauchy-Schwarz,
\[ \int_{a_n}^{\infty} [2n \mathbb{P}(x)^{n-1}] f(x) \, dx \]
\[ \leq 4n^2 \| \alpha \|^2 \int_{a_n}^{\infty} f(x) \, dI = 2n \| \alpha \|^2 \mathbb{P}(a_n) \rightarrow 0, \]  
\[ (4.17) \]

and hence it remains only to show that
\[ \int_{a_n}^{\infty} N [F_n - F]^2 \, dI = o(1). \]  
\[ (4.18) \]

When \( n \geq 1 \) it follows from (4.13) that
\[ N \int_{a_n}^{\infty} [F_n - F]^2 \, dI = N n^2 \int_{a_n}^{\infty} (1 - \mathbb{P})^{2n-2} [F_n - F]^2 \, dI \]
\[ \leq N n^2 \left[ 2n^{-b_n^2} \right] \int_{a_n}^{\infty} f \, dI = 4n^2 M \mathbb{P}(a_n) = o(1) \]  
\[ (4.19) \]

and therefore it remains only to prove (4.18) for \( 0 < n < 1 \). Thus suppose that
\( 0 < n < 1 \), and set
\[ d_n = \left( \int_{a_n}^{\infty} (\alpha + c_n)^2 \, dI \right)^{\frac{1}{2}} \]  
\[ (4.20) \]

and
\[ b_n = \mathbb{P}^{-1} (1 - \mathbb{P})^{\frac{1}{2}}. \]  
\[ (4.21) \]

Note that \( \text{support}(\alpha) \subset \text{support}(f) \) implies that \( \int_{a_n}^{\infty} \alpha^2 \, dI \rightarrow 0 \) and hence
\[ \int_{a_n}^{\infty} \]  
\[ \int_{a_n}^{\infty} \]  
\[ \| c_n \|^2 \int_{a_n}^{\infty} \alpha^2 \, dI + \| c_n \| \rightarrow 0. \]  
\[ (4.22) \]

Then, as in (4.13), we have for \( x \geq a_n \).
\[ |\tilde{F}_N(x) - F(x)| \leq N^{-1} \int_{a_N}^{b_N} \frac{(x + \varepsilon_N)^2}{\omega_N} \left[ (\tilde{F}_N(x))^2 + F(x)^2 \right] \]

which implies
\[ |\tilde{F}_N(x)^2 - F(x)^2| \leq N^{-1} d_N^2 \quad \text{for } x \geq a_N. \]  

(4.23)

Hence, from (4.22) and (4.23) it follows that for \( a_N \leq x \leq b_N \)
\[ \tilde{F}_N(x)^2 \geq F(x)^2 \left( 1 - \frac{N^{-1} d_N^2}{F(x)^2} \right) \geq F(x)^2 (1 - d_N), \]

and similarly
\[ \tilde{F}_N(x)^2 \leq F(x)^2 (1 + d_N). \]

(4.24)

(4.25)

Thus it follows, for \( 0 < n < 1 \), that
\[ N \int_{a_N}^{b_N} \left[ (\tilde{F}_N^n - F^n)^2 \right] f \, dI = N n^2 \int_{a_N}^{b_N} \left[ (1 - \tilde{F}_N^n)^2 \right] f \, dI \]
\[ \leq N n^2 \int_{a_N}^{b_N} F^{2n-2} (1 - d_N)^{4n-4} N^{-1} d_N^2 F^2 (a_N)^{2n} \, dI \]
\[ \leq n^2 (1 - d_N)^{4n-4} \int_{a_N}^{b_N} F^{2n-1} (2 + d_N)^2 f \, dI \]
\[ \leq b_n (1 - d_N)^{4n-4} (2 + d_N)^2 \int_{a_N}^{b_N} F(a_N)^{2n} = o(1). \]

(4.26)

and that, from the definition of \( b_n \) and \(|\tilde{F}_N^n - F^n| \leq 1\),
\[ N \int_{b_n}^{a_n} \left[ (\tilde{F}_N^n - F^n)^2 \right] f \, dI \leq N \int_{b_n}^{a_n} F(a_n)^{2n} = o(1). \]

(4.27)

Combining (4.19), (4.26), and (4.27) completes the proof of (4.18): Hence \( D_N \to 0 \), and this completes the proof of Lemma 1.
ACKNOWLEDGMENTS: We owe thanks to Chris Klaassen of the Mathematisch Centrum, Amsterdam, for help with the proof of Lemma 1. We also owe thanks to W. J. Hall for many helpful discussions.

REFERENCES


