HODGES-LEHMANN INVERSE LIKELIHOOD ESTIMATES (HLE'S)

KJELL DOKSUM

DEPT. OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
COLUMBIA UNIVERSITY

4TH LEHMANN SYMPOSIUM
RICE UNIV.

MAY 11, 2011

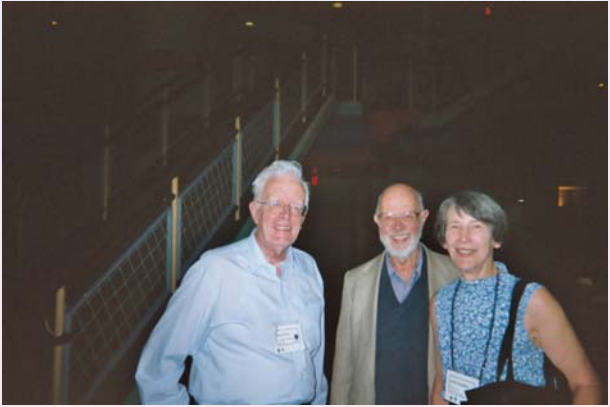

Figure: Javier Rojo

ACKNOWLEDGEMENTS

AKI OZEKI
UNIV. OF WISCONSIN

OUTLINE

1. SOME LIKELIHOODS
2. ASYMPTOTIC DISTRIBUTIONS OF HLE's
3. MINIMAX RESULTS
4. ONE STEP ESTIMATORS

WHY HL-ESTIMATORS?

1. IN LINEAR REGRESSION MODELS WITH ERROR$\sim F$, THE HL NORMAL SCORES ESTIMATE IS ASYMPTOTICALLY MORE EFFICIENT THAN THE LEAST SQUARES ESTIMATE, UNIFORMLY IN F.
2. SCHOLZ'S THEOREM. FOR EACH ONE SAMPLE ESTIMATE THAT CAN BE WRITTEN AS A LINEAR COMBINATION OF ORDER STATISTICS, THERE IS A HL-ESTIMATE THAT IS ASYMPTOTICALLY MORE EFFICIENT.

SOME LIKELIHOODS

$$X = DATA = (Y, Z),\ Y \in R,\ Z \in R^p.$$

$$\theta = (\beta \in R^p, \Lambda \in \mathcal{F}) = \text{PARAMETER}$$

$$\text{LIKELIHOOD} = \prod_i p(x_i; \theta)$$

$$\text{COX LIK} = \prod_i \frac{\lambda(y_i; \beta | z_i)}{\sum_{j \geq i} \lambda(y_i; \beta | z_j)}$$

$$\text{EMPIRICAL LIK} = \prod_i p(x_i; \theta),\ \sum p(x_i; \theta) = 1.$$

SOME LIKELIHOODS

PROFILE EMPIRICAL (PE) LIKELIHOOD

$$L_{PE}(\beta) = \sup\{\prod p(x_i;\theta);\Lambda\} \tag{1}$$

$\hat{\beta}_{PE} = \arg\max L_{PE}(\beta)$

THIS ESTIMATE $\hat{\beta}_{PE}$ IS A FUNCTION OF THE RANK $R_1, \cdots, R_n$ OF $Y_1, \cdots, Y_n$

SOME LIKELIHOODS

HOEFFDING LIK $\equiv L_H(r(y);\beta) = P(R=r) =$

$$\frac{1}{n!}E_0\left[\prod_i \frac{p(V^{(r_i)};\theta|z_i)}{p(V^{(r_i)};\theta_0|z_i)}\right],$$

WHERE $r_i \equiv r(y_i) \equiv RANK(y_i)$.
$V^{(1)} < \cdots < V^{(n)}$ ARE $p(v;\theta_0|z)$ ORDER STATISTICS.

RANK LIKELIHOOD ESTIMATOR

EXAMPLE: $Y_i = z_i^T\beta + \epsilon_i, \quad \epsilon_i \sim F$, IID.

FORWARD RANK MLE $=$ ARG MAX $L_H(r(y);\beta)$
$=$ KP MLE

KP$=$ KALBFLEISCH-PRENTICE (1973)

$\hat{\beta}_{KP}$ SOLVES $\nabla_\beta L_H(r(y);\beta) = 0$

RANK LIKELIHOOD ESTIMATOR

BECAUSE $RANK(\Lambda(y_i)) = RANK(y_i)$. FOR $\Lambda \nearrow$, $\hat{\beta}_{KP}$ APPLIES TO SEMIPARA. TRANS. MODEL. $\hat{\beta}_{KP}$ IS A FUNCTION OF THE RANKS OF Y_i, AS IS THE COX ESTIMATE.

HODGES-LEHMANN INVERSE MLE (1963)

DEFINITION:
IN THE LINEAR MODEL, $\hat{\beta}_{HL}$ SOLVES

$$\nabla_\beta L_H(r(y - z^T\beta^*); \beta)|_{\beta=0} = 0$$

1ST COMPUTE $\nabla_\beta L_H(r(y;\beta))|_{\beta=0}$, THEN CONSTRUCT AN ESTIMATING EQUATION IN β^* BY REPLACING y WITH $y - z^T\beta^*$. HERE $y - z^T\beta$ IS THE "INVERSE" OF $y = z^t\beta + \epsilon$.

HODGES-LEHMANN INVERSE MLE (1963)

HL INVERSE LIK. EST: ALIGN RANK OF RESIDUALS WITH THE "BASELINE" RANKS USING HOEFFDING LIKELIHOOD.

EXAMPLE: TWO SAMPLE CASE, LOGISTIC SHIFT MODEL, $F_2(x) = F_1(x - \Delta)$,
$\nabla_\Delta L_H|_{\Delta=0} =$ WILCOXON STAT
$\hat{\Delta}_{HL} = med(X_{2j} - X_{1i})$,

GENERAL HODGES-LEHMANN INVERSE MLE

MODEL: $Y = h(\epsilon, z, \beta)$,
LET $g(y; z, \beta)$ BE THE SOLUTION (INVERSE)
FOR ϵ OF $h(\epsilon, z, \beta) = y$.
$\hat{\beta}_{HL}$ SOLVES

$$\nabla_\beta L_H[g(y; z, \beta^*); \beta]|_{\beta=0} = 0$$

GENERAL HODGES-LEHMANN INVERSE MLE

IN THE EXAMPLE $Y_i = z_i^T \beta + \epsilon$,

$$\nabla_\beta L_H[r(y - z^T \beta^*); \beta]|_{\beta=0}$$

ARE p LINEAR RANK STATISTICS

$$T_{nj}(\beta^*) = \sum_{i=1}^{n} (z_{ij} - \bar{z}_j) a_n(r_i(\beta^*)), \quad j = 1, \cdots, p$$

WHERE $r_i(\beta^*) = RANK(y_i - z_i^T \beta^*)$, AND
$a_n(r) = a\left(\frac{r}{n+1}\right)$,
$a(u) = -\frac{f'}{f}(F^{-1}(u))$

SCALE MODEL:

$Y = \epsilon \exp[z^T\beta]$, $\epsilon \sim F$

$\epsilon = Y/\exp[z^T\beta]$,

$a_n(RANK(y_i/\exp[z_i^T\beta]))$,

$a_n(r) = a_1\left(\frac{r}{n+1}\right)$,

$a_1(u) = -F^{-1}(u)\frac{f'}{f}(F^{-1}(u)) - 1$

HERE $\hat{\beta}_{HL}$ SOLVES

$$\nabla_\beta L_H[r(y_i/\exp[z_i^T\beta^*]);\beta]|_{\beta=0} = 0$$

WHICH IS EQUIVALENT TO

$$\sum_{i=1}^n (z_{ij} - \bar{z}_j)a_n(r_i(\beta^*)) = 0, \quad j = 1, \cdots, p$$

LINEAR MODELS:

EX1: $\epsilon \sim$ LOGISTIC $\Longrightarrow a_n(r) = \frac{r}{n+1}$

EX2: $\epsilon \sim$ NORMAL $\Longrightarrow a_n(r) = \Phi^{-1}\left(\frac{r}{n+1}\right)$,
NORMAL SCORES

SCALE MODEL:

EX3: $\epsilon \sim EXP \Longrightarrow a_n(r) = -\log\left(1 - \frac{r}{n+1}\right)$,
$\hat{\beta}_{HL}$ IS THE LOGISTIC SCORES ESTIMATOR

THEOREM (JAECKEL 1972):
IN THE $Y_i = z_i^T\beta + \epsilon$ MODEL,

1. THE HL ESTIMATE $\hat{\beta}_{HL}$ IS A MAXIMIZER OF

$$S(\beta) = \prod_{i=1}^{n} \exp[-(y_i - z_i^T\beta) \cdot a_n(RANK(y_i - z_i^T\beta))]$$

2. HERE $\log[S(\beta)]$ is NONNEGATIVE, CONTINUOUS AND CONCAVE.

ASYMPTOTICS

IN THE LINEAR MODEL, LET $\varphi(u, f_0) = -\frac{f_0'}{f_0}(F_0^{-1}(u))$. THEN,

$$\sqrt{n}(\hat{\beta}_{HL} - \beta) \to N(0, \left[\frac{\int_0^1 [\varphi(u, f_0) - \bar{\phi}]^2 du}{\int_0^1 \varphi(u, f_0)\varphi(u, f)du}\right]^2 \Sigma^{-1})$$

WHERE $\bar{\phi} = \int_0^1 \varphi(u, f_0)du$,$\Sigma = LIM_{n\to\infty}\ n^{-1}Z^TZ$, $Z =$ CENTERED DESIGN MATRIX, $\epsilon_i \sim F$, $f = F'$

HERE $f_0(\cdot)$ GENERATES $L_H(\cdot)$ AND $\hat{\beta}_{HL}$.
$f(\cdot)$ IS THE TRUE DENSITY of ϵ.

LINEAR MODEL EXAMPLES:

1. $F_0()$=LOGISTIC

$$\sqrt{n}(\hat{\beta}_{HL} - \beta) \rightarrow N(0, \left[\frac{1}{12(\int_0^1 f^2(u)du)^2}\right] \Sigma^{-1})$$

2. $F_0()=NORMAL(0, \sigma^2)$

$$\sqrt{n}(\hat{\beta}_{HL} - \beta) \rightarrow N(0, \left[\frac{\Sigma^{-1}}{(\int_0^1 \Phi^{-1}(u)\phi(u, f)du)^2}\right])$$

ASYMPTOTIC INEQUALITY

HODGES-LEHMANN (56) CONJECTURE.
CHERNOFF-SAVAGE (58) THEOREM.

IF $\hat{\beta}_{HL}$ IS BASED ON SCORES DERIVED BY TAKING $f_0 = N(0,1)$, AND IF $\hat{\beta}_{MLE}$ IS THE MLE FOR THE MODEL WITH $\epsilon \sim N(0,\sigma^2)$, THEN

ASYMPTOTIC $VARIANCE_F(\hat{\beta}_{HL}) \leq$
ASYMPTOTIC $VARIANCE_F(\hat{\beta}_{MLE})$

WHERE $F =$ TRUE DIST. OF ϵ.
EQUALITY ONLY WHEN $F = N(0,\sigma^2)$.

IN THE AFT MODEL WITH $\epsilon \sim F$, THE HL EXPONENTIAL SCORES STATISTIC SATISFIES

$$\sqrt{n}(\hat{\beta}_{HL} - \beta) \rightarrow N(0, \left[\frac{1}{\int_0^1 t\lambda(t)dF(t)}\right]^2 \Sigma^{-1})$$

WHERE $\lambda(t) = f(t)/[1 - F(t)]$.

RESULT: THE COX ESTIMATE IS ASYMPTOTICALLY MINIMAX FOR THE PROPORTIONAL HAZARD (PH) MODEL: $\lambda(y; z) = \lambda_0(y) e^{z^T \beta}$

PROOF:
STEP A: THE COX ESTIMATE IS OPTIMAL FOR THE EXPONENTIAL MODEL,

$$INF_{\hat{\beta}} R_E(\beta, \hat{\beta}) = R_E(\beta, \hat{\beta}_C) \tag{2}$$

STEP B: THE PH MODEL CAN BE WRITTEN AS $\Lambda_0(Y) \sim$ EXP-DISTR$(z^T \beta)$ WHERE $\Lambda_0 \nearrow$ IS THE BASELINE HAZARD FUNCTION.

NAIVE MINIMAX THEORY

THE COX ESTIMATE $\hat{\beta}_C$ IS INVARIANT, $\hat{\beta}_C(y) = \hat{\beta}_C(\Lambda_0(y))$, SO IT HAS CONSTANT RISK,

$$\sup_{F(y|z) \in PH} R_F(\beta, \hat{\beta}_C) = R_E(\beta, \hat{\beta}_C), \tag{3}$$

STEP C: SINCE THE EXP MODEL IS PH,

$$\sup_{F(y|z) \in PH} R_F(\beta, \hat{\beta}) \geq R_E(\beta, \hat{\beta}), \tag{4}$$

STEP D: (2),(3),(4) $\Rightarrow$

$$\sup_F R_F(\beta, \hat{\beta}_C) = \inf_{\hat{\beta}} \sup_F R_F(\beta, \hat{\beta}). \quad QED.$$

NON-NAIVE PROOF: PAGE 332 of BICKEL, KLAASSEN, RITOV, WELLNER.

NAIVE MINIMAX THEORY

RESULT: THE HL EXP SCORES EST IS A MINIMAX FOR THE IHR ACCELERATED FAILURE TIME MODEL (IHRAFT)

$$Y = Y_0 \exp(z^T \beta), \qquad Y_0 \sim F,$$

WITH $F \in IHR =$ INCR. HAZARD RATE

PROOF:

STEP A: THE HL EXP. SC. ESTIMATE IS OPTIMAL FOR THE EXPONENTIAL MODEL,

$$INF_{\hat{\beta}} R_E(\beta, \hat{\beta}) = R_E(\beta, \hat{\beta}_{HL}) \tag{5}$$

STEP B: THE EXP MODEL IS LEAST FAVORABLE FOR $\hat{\beta}_{HL}$

$$\sup_F R_F(\beta, \hat{\beta}_{HL}) = R_E(\beta, \hat{\beta}_{HL}), \qquad (6)$$

$F(y|z) \in IHRAFT$

NAIVE MINIMAX THEORY

STEP C: SINCE THE EXP MODEL IS IHRAFT,

$$\sup_F R_F(\beta, \hat{\beta}) \geq R_E(\beta, \hat{\beta}), \qquad (7)$$

$F(y|z) \in IHRAFT$
STEP D: (5),(6),(7) $\Rightarrow$

$$\sup_F R_F(\beta, \hat{\beta}_{HL}) = \inf_{\hat{\beta}} \sup_F R_F(\beta, \hat{\beta}). \quad QED.$$

TO PROVE STEP B, USE DOKSUM (1967); ARGUMENT BASED ON VAN ZWET ORDERINGS.

ASYMPTOTIC miniMAX RESULT

CONSIDER $f_0 =$ LOGISTIC, SO

$$a_n(r_i) = \frac{r_i}{n+1} \tag{8}$$

THEN $\hat{\beta}_{HL}$ IS ASYMPTOTICALLY miniMAX OVER THE CLASS OF DISTRIBUTIONS WITH (VAN ZWET TYPE) LIGHTER TAILS THAN THE LOGISTIC DISTRIBUTION.

ONE STEP ESTIMATIORS

LET $\hat{\tau}$ BE A CONSISTENT ESTIMATOR OF

$$\tau = \frac{1}{\int_0^1 \phi(u, f_0)\phi(u, f)du} \tag{9}$$

AND LET $\hat{\beta}_{LSE}$ BE THE LSE OF β.

DEFINE

$$\hat{\beta}_{HL} = \hat{\beta}_{LSE} + \hat{\tau} \cdot (Z^T Z)^{-1} \cdot T_n(RANK(Y - Z^T \hat{\beta}_{LSE})) \tag{10}$$

THEN,

$$\sqrt{n}(\hat{\beta}_{HL} - \beta) \to N(0, \Gamma) \tag{11}$$

JURECKOVA(69), KRAFT AND VAN EEDEN(72), HETTMANSPERGER, MCKEAN, TSIATIS, ETC.

GMT MODEL

IN THE MODEL $Y_i = z_i^T \beta + \epsilon_i$
THERE EXISTS $G : R \to R$, INCREASING, SUCH THAT

$$G(y_i - z_i^T \beta), \quad i = 1, \cdots, n \tag{12}$$

ARE IID. HERE G is UNKNOWN.

SUMMARY

1. SOME LIKELIHOODS
2. ASYMPTOTIC DISTRIBUTIONS OF HLE's
3. MINIMAX RESULTS
4. ONE STEP ESTIMATORS