

Stat 701 Take-Home Test, Spring 2009

Instructions. Do all three of the following problems, and hand them in by Monday evening, May 11, 2009.

(I) Suppose that a sample of *iid* pairs $\{(Z_i, Y_i), i = 1, \dots, n\}$ follows the ‘true’ probability model

$$Z_i \sim \text{Expon}(\lambda), \quad E(Y_i|Z_i) = \alpha + \beta Z_i, \quad \text{Var}(Y_i|Z_i) = \sigma^2 Z_i \quad (*)$$

(a) Not knowing the true model (*), we analyze the data by least-squares estimates, i.e. by estimates of (α, β) minimizing $\sum_{i=1}^n (Y_i - \alpha - \beta Z_i)^2$. Show that this estimator is \sqrt{n} consistent, and find an expression for the variance-covariance matrix V of the asymptotic distribution of $\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta)$. Note: this variance will involve other (nuisance) parameters λ and σ^2 .

(b) How can V be estimated consistently, if you know only that $Z_i \sim \text{Expon}(\lambda)$ and that (Z_i, Y_i) are *iid* with $E(Y_i - \alpha - \beta Z_i | Z_i) = 0$ (or that $E(Y_i - \alpha - \beta Z_i) = E(Z_i (Y_i - \alpha - \beta Z_i)) = 0$) ?

Note that this part is intended to use theory about misspecified-model or estimating-equation estimators covered in class and in the book. However, you may develop and justify estimators from first principles if you wish.

(II) Now consider two-group survival data in the form of *iid* pairs $\{(X_i, Y_i)\}_{i=1}^n$, with group-label $X_i \sim \text{Binom}(1/2, 1)$ and for $x = 0, 1$,

$$P(Y_i \geq t | X_i = x) = \exp\left(-\int_0^t \lambda e^{\vartheta x a(s)} ds\right)$$

for some bounded nonnegative function $a(s)$.

(a). Find the locally optimal test of $H_0 : \vartheta = 0$ with approximate size .05 in large samples, which is based on the score statistic, constructed under the working (but not necessarily correct) assumption that $a(s) \equiv 1$ for all $s \geq 0$, in terms of the data $\{(X_i, Y_i), i = 1, \dots, n\}$.

(b). Show that this test of $H_0 : \vartheta = 0$ is *consistent* (i.e., has rejection probability tending to 1) against all fixed alternatives with $\vartheta > 0$ and true $a(s)$ function which is everywhere positive.

(c). Find a large-sample formula for the limiting power of your test in part (a) against $\vartheta = 1/\sqrt{n}$ for the true $a(s) = I_{[s \leq 2]} + (1/2)I_{[s > 2]}$. (Do your probability-of-rejection calculation using this as the true $a(s)$ function in the model, even though your test in part (a) is based on a simpler and different $a(s)$ function.)

(III) Suppose that we observe X_i , $i = 1, \dots, n$, *iid* scalar-valued measurements, and we wish to test the hypothesis \mathbf{H}_0 that they are distributed according to the Logistic(μ, σ) distribution with d.f. $F(x, \vartheta) = (1 + e^{-(x-\mu)/\sigma})^{-1}$ for some $\mu \in \mathbf{R}$, $\sigma > 0$. The parameters $\vartheta = (\mu, \sigma)$ are estimated in two distinct ways:

- (i) first, $\hat{\vartheta}$ denotes the MLE's, and
- (ii) second, denote by $\tilde{\vartheta}$ the MLE's for the same parameters based on the reduced data (N_1, N_2, N_3, N_4) where N_j is the number $\sum_{i=1}^n I_{[X_i \in (a_{j-1}, a_j])}$ of observations falling between a_{j-1} and a_j , where the a_j 's are fixed and respectively

$$a_0 = -\infty, \quad a_1 = -1, \quad a_2 = 0, \quad a_3 = 1, \quad a_4 = \infty$$

- (a) Find the asymptotic variance matrices of these two estimators.
- (b) Find the asymptotic null-hypothesis distribution of the test statistic

$$\sum_{j=1}^4 \frac{(N_j - n(F(a_j, \tilde{\vartheta}) - F(a_{j-1}, \tilde{\vartheta})))^2}{n(F(a_j, \tilde{\vartheta}) - F(a_{j-1}, \tilde{\vartheta}))}$$

- (c) Find the asymptotic null-hypothesis distribution of the random variable $(N_1 - nF(-1, \hat{\vartheta}))/\sqrt{n}$.