Statistics 582, Problem Set 1 Solutions

Wellner; 1/17/98

1. This is a continuation of problem 10.4:
   A. For the data in problem 10.4, compute the maximum likelihood estimator of $q(\theta) \equiv F_\theta^{-1}(1/2)$ and a 95% confidence interval.
   B. For the data in problem 10.4, compute the nonparametric estimator of $F_\theta^{-1}(1/2)$ and a 95% confidence interval. Compare with the confidence interval computed in A.
   Solution: A. Since $F_\theta(x) = 1 - \exp(-(x/\alpha)^\beta)$ and $F_\theta^{-1}(t) = \alpha(-\log(1-t))^{1/\beta}$, the MLE of $q(\theta) = F_\theta^{-1}(1/2) = \alpha(\log 2)^{1/\beta}$ is simply $q(\widehat{\theta}_n) = \widehat{\alpha}(\log 2)^{1/\widehat{\beta}} \doteq 7.597....$ Furthermore, since

   $$\sqrt{n}(q(\widehat{\theta}_n) - q(\theta)) \rightarrow_d N(0, \dot{q}^T(\theta)I^{-1}(\theta)\dot{q}(\theta)),$$

   where $\dot{q}^T(\theta) = (q(\theta)/\alpha, (-1/\beta^2)(\log\log 2)q(\theta))$, a 95% confidence (approximate) confidence interval for $q(\theta)$ is given by

   $$q(\widehat{\theta}_n) \pm 1.96 n^{-1/2}\sqrt{\dot{q}^T(\widehat{\theta})I^{-1}(\widehat{\theta})\dot{q}(\widehat{\theta})} \quad = \quad 7.597 \pm 5.203 = (2.393, 12.8007).$$

   B. The natural nonparametric estimator of $F^{-1}(1/2)$ is the sample median $\mathbb{F}_n^{-1}(1/2)$; in this case we find $\mathbb{F}_n^{-1}(1/2) = X_{(10)} = 6.5$. Moreover, since

   $$\begin{aligned} P(X_{(r)} < F^{-1}(1/2) < X_{(s)}) &= P(X_{(r)} < F^{-1}(1/2)) - P(X_{(s)} < F^{-1}(1/2)) \\ &= P(n\mathbb{F}_n(F^{-1}(1/2)) \ge r) - P(n\mathbb{F}_n(F^{-1}(1/2)) \ge s) \\ &= P(Bin(n, 1/2) \ge r) - P(Bin(n, 1/2) \ge s) \\ &= .990 - .032 = .958 \end{aligned}$$

   when $n = 19$, $r = 5$, and $s = 14$. Hence $(X_{(5)}, X_{(14)})$ is a .958 confidence interval for $F^{-1}(1/2)$. For the present data we find that $(X_{(5)}, X_{(14)}) = (2.78, 12.06)$. Note that this nonparametric CI is in excellent agreement with the parametric confidence interval calculated in part A.

2. (Right censored data). Suppose that $X, X_1, \ldots, X_n$ are i.i.d. survival times with unknown distribution function F, that $Y, Y_1, \ldots, Y_n$ are i.i.d. censoring times with unknown distribution function G, assumed to be independent of the X_i's, and that we can observe only the iid pairs $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ where $Z_i \equiv X_i \wedge Y_i$ and $\delta_i \equiv 1_{[X_i \le Y_i]}$; also let $Z \equiv X \wedge Y$ and $\delta = 1_{[X \le Y]}$.
   A. Show that the joint distribution of (Z, δ) is given by

   $$P(Z \le z, \delta = 1) = \int_{(0,z]} (1 - G(x-))dF(x)$$

where $G(x-) \equiv \lim_{y \uparrow x} G(y)$, and

$$P(Z \le z, \delta = 0) = \int_{(0,z]} (1 - F(y)) dG(y).$$

Furthermore, show that the survival function $1 - H(z) = P(Z > z)$ is given by $1 - H(z) = (1 - F(z))(1 - G(z))$.

B. Suppose we assume a parametric family of densities $\{f_\theta : \theta \in \Theta\}$ for the distribution of the survival time X. Find general expressions for the density of the observations (Z, δ) (assuming that the censoring distribution G also has density g which does not depend on θ), and for the scores for θ based on observation of one (Z, δ) pair.

C. Suppose that the parametric family $\{f_\theta : \theta \in \Theta\}$ in part B is the Weibull family with survival functions $\overline{F}_\theta(x) = \exp(-(x/\alpha)^\beta)$ with $\theta = (\alpha, \beta) \in R^{+2}$. Derive an asymptotic likelihood based test of the hypothesis that the X_i are Rayleigh (i.e. $\beta = 2$), versus the alternative that they are Weibull distributed with $\beta \neq 2$.

Solution: A. First,

$$\begin{aligned} P(Z \le z, \delta = 1) &= P(X \le z, X \le Y) = E\{1_{[X \le z]} 1_{[X \le Y]}\} \\ &= E\{1_{[X \le z]} E(1_{[X \le Y]} | X)\} = E\{1_{[X \le z]}(1 - G(X-))\} \\ &= \int_{(0,z]} (1 - G(x-)) dF(x). \end{aligned}$$

Similarly,

$$\begin{aligned} P(Z \le z, \delta = 0) &= P(Y \le z, Y < X) = E\{1_{[Y \le z]} 1_{[Y \le X]}\} \\ &= E\{1_{[Y \le z]} E(1_{[Y < X]} | Y)\} = E\{1_{[Y \le z]}(1 - F(Y))\} \\ &= \int_{(0,z]} (1 - F(y)) dG(y) . \end{aligned}$$

Also note that, using integration by parts,

$$\begin{aligned} H(z) &= P(Z \le z) = \int_{(0,z]} (1 - G(x-)) dF(x) + \int_{(0,z]} (1 - F(y)) dG(y) \\ &= (1 - G) F|_{(0,z]} - \int_{(0,z]} F d(1 - G) + \int_{(0,z]} (1 - F) dG \\ &= (1 - G(z)) F(z) + G(z) - \int_{(0,z]} (1 - F) dG + \int_{(0,z]} (1 - F) dG \\ &= 1 - (1 - F(z))(1 - G(z)). \end{aligned}$$

B. When F and G have densities f and g with respect to the dominating measure μ, it follows from A that the density of (Z, δ) with respect to $\mu \times$counting measure is given by

$$p(z, \delta) = \{\overline{G}(z) f(z)\}^\delta \{\overline{F}(z) g(z)\}^{1-\delta} = f^\delta(z) \overline{F}^{1-\delta}(z) \overline{G}^\delta(z) g^{1-\delta}(z).$$

When $f = f_\theta$ for a parametric family, this yields

$$\begin{aligned} l(\theta|Z,\delta) &= \log p_\theta(z,\delta) \\ &= \delta \log f_\theta(z) + (1-\delta)\log \overline{F}_\theta(z) + \text{constant in } \ \theta\,. \end{aligned}$$

Hence the score(s) for θ based on observation of (Z,δ), $\dot{l}_\theta(Z,\delta)$, are related to the score(s) for θ based on observation of X, $\dot{l}_\theta(X)$, by

$$\begin{aligned} \dot{l}_\theta(Z,\delta) &= \delta \dot{l}_\theta(Z) + (1-\delta)\frac{\int_Z^\infty \dot{l}_\theta(x) f_\theta(x) d\mu(x)}{\overline{F}_\theta(Z)} \\ &= E(\dot{l}_\theta(X)|Z,\delta). \end{aligned}$$

C. I propose to use the Rao (or score statistic); of course the Wald and Likelihood ratio tests are also reasonable possibilities. I will use the parametrization of the Weibull famly given in example 3.2.5: with $\theta = (\alpha,\beta)$,

$$f_\theta(x) = \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1}\exp(-(\frac{x}{\alpha})^\beta)1_{(0,\infty)}(x), \qquad \overline{F}_\theta(x) = \exp(-(\frac{x}{\alpha})^\beta).$$

Thus $p_\theta(z,\delta) = (f_\theta(z)\overline{G}(z))^\delta(\overline{F}_\theta(z)g(z))^{1-\delta}$, and

$$\begin{aligned} \log p_\theta(z,\delta) &= \delta \log f_\theta(z) + (1-\delta)\log \overline{F}_\theta(z) + \text{constant in } \theta \\ &= \delta\{\log(\frac{\beta}{\alpha}) + (\beta-1)\log(\frac{z}{\alpha})\} - (\frac{z}{\alpha})^\beta. \end{aligned}$$

Hence the scores for α and β are

$$\dot{l}_\alpha(z,\delta) = -\frac{\beta\delta}{\alpha} + \beta(\frac{z}{\alpha})^{\beta-1}\frac{z}{\alpha^2} = \frac{\beta}{\alpha}\{(\frac{z}{\alpha})^\beta - \delta\}$$

$$\dot{l}_\beta(z,\delta) = \frac{\delta}{\beta} - \frac{1}{\beta}\log[(\frac{z}{\alpha})^\beta]\{(\frac{z}{\alpha})^\beta - \delta\}\,.$$

Note that these reduce to the scores we derived in section 3.2 when there is no censoring, so $\delta = 1$. When $\beta = 2$ the score for α simplifies to

$$\dot{l}_\alpha(z,\delta;\theta_0) = \frac{2}{\alpha}\{\left(\frac{z}{\alpha}\right)^2 - \delta\}\,,$$

so the score equation for α in the smaller model $\beta = 1$

$$\frac{1}{\alpha}\sum_{i=1}^n\left\{\left(\frac{Z_i}{\alpha}\right)^2 - \delta_i\right\} = 0\,,$$

and this yields $\hat{\alpha}_0 = \{\sum Z_i^2/\sum\delta_i\}^{1/2}$. The score for β is, for $\theta_0 = (\alpha,1)\in\Theta_0$,

$$\dot{l}_\beta(z,\delta;\theta_0) = \frac{\delta}{2} - \frac{1}{2}\log\left((\frac{z}{\alpha})^2\right)\left\{(\frac{z}{\alpha})^2 - \delta\right\}$$

and hence

$$Z_{n2}(\hat{\theta}_n^0) \quad = \quad \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \left\{ \delta_i - \log(\frac{Z_i}{\widehat{\alpha}_0})^2 \left\{ \left(\frac{Z_i}{\widehat{\alpha}_0} \right)^2 - \delta_i \right\} \right\}$$

The next step is to compute an estimator of the information matrix under the null hypothesis $\beta = 2$; I propose to do this via the second derivative matrix (Hessian) $-\sum_{i=1}^n \ddot{l}_{\theta\theta}(X_i; \widehat{\theta}_n^0)$. Then the Rao statistic for testing $H_0 : \beta = 2$ is

$$R_n = \{Z_{n2}(\widehat{\theta}_n^0)\}^2 \widehat{I}^{22} = \{Z_{n2}(\widehat{\theta}_n^0)\}^2 / (\widehat{I}_{\beta\beta} - \widehat{I}_{\alpha\beta}^2 \widehat{I}_{\alpha\alpha}^{-1}),$$

and we reject H_0 if R_n is big relative to a critical point selected from χ_1^2.

3. Suppose that $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. with bivariate normal distribution $N_2(\mu, \Sigma)$ where $\mu \in R2$ and

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma\tau\rho \\ \sigma\tau\rho & \tau^2 \end{pmatrix}$$

where $\sigma^2 > 0$, $\tau^2 > 0$, and $\rho \in (-1, 1)$.
A. If we assume that $\mu_1 = \mu_2 \equiv \theta$ and Σ is known, what is the MLE of θ?
B. If we assume that μ is known and $\sigma^2 = \tau^2 \equiv \theta$, what is the MLE of θ?
C. What is the asymptotic distribution of the estimator you found in B?
D. Under the same assumption as in B, what is the MLE of ρ?
E. What is the asymptotic distribution of the estimator you found in D?

Solution: A. When $\mu_1 = \mu_2 = \theta$ and Σ is known, then the log-likelihood for one observation is (relabelling $\mu_1 = \mu$, $\mu_2 = \nu$),

$$\log p(x; \theta) = -\frac{1}{2(1-\rho^2)} \left\{ \frac{(x-\theta)^2}{\sigma^2} - 2\rho \frac{(x-\theta)}{\sigma}\frac{(y-\theta)}{\tau} + \frac{(y-\theta)^2}{\tau^2} \right\} + constant.$$

Hence the score for θ for one observation is

$$\begin{aligned} \dot{l}_\theta(x, y) &= \frac{1}{1-\rho^2} \left\{ \frac{(x-\theta)}{\sigma^2} + \frac{(y-\theta)}{\tau^2} - \frac{\rho}{\tau}\frac{(x-\theta)}{\sigma} - \frac{\rho}{\sigma}\frac{(y-\theta)}{\tau} \right\} \\ &= \frac{1}{1-\rho^2} \left\{ \frac{(x-\theta)}{\sigma}\left(\frac{1}{\sigma} - \frac{\rho}{\tau}\right) + \frac{(y-\theta)}{\tau}\left(\frac{1}{\tau} - \frac{\rho}{\sigma}\right) \right\}. \end{aligned}$$

Thus the score equation for θ is

$$0 = \dot{l}_{n\theta}(\theta) = \frac{n}{1-\rho^2} \left\{ \frac{(\overline{X}_n - \theta)}{\sigma}\left(\frac{1}{\sigma} - \frac{\rho}{\tau}\right) + \frac{(\overline{Y}_n - \theta)}{\tau}\left(\frac{1}{\tau} - \frac{\rho}{\sigma}\right) \right\},$$

and hence

$$\begin{aligned} \widehat{\theta}_n &= \frac{\overline{X}_n\left(\frac{1}{\sigma^2} - \frac{\rho}{\sigma\tau}\right) + \overline{Y}_n\left(\frac{1}{\tau^2} - \frac{\rho}{\sigma\tau}\right)}{\frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2}} \\ &= a\overline{X}_n + (1-a)\overline{Y}_n \end{aligned}$$

where

$$a = \frac{\frac{1}{\sigma^2} - \frac{\rho}{\sigma\tau}}{\frac{1}{\sigma^2} - \frac{2\rho}{\sigma\tau} + \frac{1}{\tau^2}}.$$

Note that this yields

$$Var(\widehat{\theta}) = \frac{1}{n}\left\{ a^2\sigma^2 + 2a(1-a)\rho\sigma\tau + (1-a)^2\tau^2 \right\}.$$

B. and D. If $\sigma^2 = \tau^2 = \theta$ and μ is known, then the log-likelihood for one observation is (again relabelling $\mu_1 = \mu$, $\mu_2 = \nu$),

$$\log p(x;\theta,\rho) = -\log\theta - \frac{1}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)\theta}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} + \text{constant}.$$

Thus the scores for θ and ρ are given by

$$\dot{l}_\theta(x,y) = -\frac{1}{\theta} + \frac{1}{2(1-\rho^2)\theta^2}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\},$$

$$\dot{l}_\rho(x,y) = \frac{\rho}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)^2\theta}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} + \frac{1}{\theta(1-\rho^2)}(x-\mu)(y-\nu).$$

Hence the score equations for estimation of θ and ρ are given by

$$0 = \dot{l}_{n\theta}(\theta) = \sum_{i=1}^n \dot{l}_\theta(X_i,Y_i) = -\frac{n}{\theta} + \frac{n}{\theta^2 2(1-\rho^2)}\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\},$$

and

$$0 = \dot{l}_{n\rho}(\rho) = \sum_{i=1}^n \dot{l}_\rho(X_i,Y_i) = \frac{n\rho}{1-\rho^2} - \frac{n\rho}{\theta(1-\rho^2)^2}\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\} + \frac{n}{\theta(1-\rho^2)}S_{XY}$$

where

$$S_{XX} \equiv n^{-1}\sum_{i=1}^n (X_i-\mu)^2, \quad S_{XY} \equiv n^{-1}\sum_{i=1}^n (X_i-\mu)(Y_i-\nu), \quad S_{YY} \equiv n^{-1}\sum_{i=1}^n (Y_i-\nu)^2.$$

Solving the first of these for $\widehat{\theta}$ yields

$$\widehat{\theta} = \frac{1}{2(1-\widehat{\rho}^2)}\left\{S_{XX} - 2\widehat{\rho} S_{XY} + S_{YY}\right\};$$

Rewriting the score equation for ρ with a common denominator of $\theta(1-\rho^2)^2$ yields

$$\theta\rho(1-\rho^2) - \rho\left\{S_{XX} - 2\rho S_{XY} + S_{YY}\right\} + (1-\rho^2)S_{XY} = 0;$$

and then plugging in the estimator $\widehat{\theta}$ of θ yields the equation

$$(1-\widehat{\rho}^2)S_{XY} = \frac{1}{2}\widehat{\rho}\left\{S_{XX} - 2\widehat{\rho} S_{XY} + S_{YY}\right\}.$$

This has the solution

$$\widehat{\rho} = \frac{2S_{XY}}{S_{XX} + S_{YY}};$$

plugging this (or more precisely the last form of the equation for $\widehat{\rho}$) into the expression for $\widehat{\theta}$ yields $\widehat{\theta} = (S_{XX} + S_{YY})/2$.
C. and E. To find the asymptotic distributions of $\widehat{\theta}$ and $\widehat{\rho}$ we could either (i) proceed directly from first principles (central limit theorms and the delta method), or (b) use theorem 4.1.5 concerning the asymptotic behavior of maximum likelikhood estimators. I'll take the second route here. The first step in this direction is to compute the information matrix for (θ, ρ). Now

$$\ddot{l}_{\theta\theta}(x,y) = \frac{1}{\theta^2} - \frac{1}{(1-\rho^2)\theta^3}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\},$$

$$\begin{aligned}\ddot{l}_{\theta\rho}(x,y) &= \frac{2\rho}{2\theta^2(1-\rho^2)^2}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad - \frac{1}{2\theta^2(1-\rho^2)^2}2(x-\mu)(y-\nu),\end{aligned}$$

and

$$\begin{aligned}\ddot{l}_{\rho\rho}(x,y) &= \frac{1}{1-\rho^2} + \frac{2\rho^2}{(1-\rho^2)^2} \\ &\quad - \left\{\frac{1}{(1-\rho^2)^2} + \frac{4\rho^2}{(1-\rho^2)^3}\right\}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\quad + \frac{\rho}{\theta(1-\rho^2)^2}2(x-\mu)(y-\nu) + \frac{2\rho}{\theta(1-\rho^2)^2}(x-\mu)(y-\nu) \\ &= \frac{1+\rho^2}{(1-\rho^2)^2} - \frac{1-3\rho^2}{\theta(1-\rho^2)^3}\left\{(x-\mu)^2 - 2\rho(x-\mu)(y-\nu) + (y-\nu)^2\right\} \\ &\qquad + \frac{4\rho}{\theta(1-\rho^2)^2}(x-\mu)(y-\nu).\end{aligned}$$

Here

$$E\left\{(X-\mu)^2 - 2\rho(X-\mu)(Y-\nu) + (Y-\nu)^2\right\} = 2\theta(1-\rho^2)$$

and

$$E(X-\mu)(Y-\nu) = \rho\theta.$$

Thus we find that

$$I_{\theta\theta} = E(-\ddot{l}_{\theta\theta}(X,Y)) = \theta^{-2},$$

$$I_{\theta\rho} = E(-\ddot{l}_{\theta\rho}(X,Y)) = \frac{\rho}{\theta(1-\rho^2)},$$

and

$$I_{\rho\rho} = E(-\ddot{l}_{\rho\rho}(X,Y)) = \frac{1+\rho^2}{(1-\rho^2)^2}\,.$$

This yields

$$I_{\theta\theta\cdot\rho} = I_{\theta\theta} - I_{\theta\rho} I_{\rho\rho}^{-1} I_{\rho\theta} = \frac{1}{\theta^2}\frac{1}{1+\rho^2}$$

and

$$I_{\rho\rho\cdot\theta} = I_{\rho\rho} - I_{\rho\theta} I_{\theta\theta}^{-1} I_{\theta\rho} = (1-\rho^2)^{-2}\,.$$

Hence it follows from theorem 4.1.5 that

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow_d N(0, I_{\theta\theta\cdot\rho}^{-1}) = N(0, \theta^2(1+\rho^2))$$

while

$$\sqrt{n}(\widehat{\rho}_n - \rho) \rightarrow_d N(0, I_{\rho\rho\cdot\theta}^{-1}) = N(0, (1-\rho^2)^2)\,.$$