Statistics 582, Midterm Exam Solution

Wellner; 2/17/2010

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
   (a) The *risk function* $R(\theta, d)$, $\theta \in \Theta$, of a decision rule d.
   (b) The *Bayes risk* $\mathcal{R}(\Lambda, d)$ of a decision rule d with respect to a prior distribution Λ.
   (c) A minimax decision rule d.
   (d) A Bayes decision rule d with respect to the prior distribution Λ.
   (e) An *integrable envelope function* F of a class of functions $\mathcal{F} = \{f : \mathcal{X} \to \mathbb{R}\}$ in the context of a Uniform Strong Law of Large Numbers (or Glivenko-Cantelli theorem).

   Solution: See course notes, Chapters 4 and 5.

2. (24 points) **State and prove** *any two* of the following results:
   (a) A theorem concerning admissibility of Bayes rules (in the context of finite parameter, action, and sample spaces).
   (b) A formula expressing a survival function $1 - F(x) = P(X > x)$ of a non-negative random variable X in terms of the corresponding cumulative hazard function $\Lambda(x) = \int_{[0,x]} (1 - F(y-))^{-1} dF(y)$.
   (c) An inequality satisfied by the Kullback-Leibler information or divergence, $K(P, Q)$.
   (d) A relationship between the score for "incomplete" data Y and "complete" data X when $Y = T(X)$ for some measurable function T and when $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a model for the complete data X.

   Solution: See course notes, Chapters 4 and 5.

Do **either** problem 3 **or** problem 4.

3. (42 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F on $\mathbb{R}^+$ and $Y_1, \ldots, Y_n$ are i.i.d. on $\mathbb{R}^+$ with distribution function G. We observe $(Z_i, \Delta_i) \equiv (X_i \wedge Y_i, 1_{[X_i \le Y_i]})$ for $i = 1, \ldots, n$, and our goal is to estimate the distribution function F (or, equivalently, the survival function $1 - F$.
(a) Describe the distribution of (Z_1, Δ_1) in terms of two sub-distribution functions expressed in terms of F and G, and give $P(Z_1 \ge z)$ in terms of F and G
(b) Relate the cumulative hazard function Λ of F to the two sub-distribution functions and survival function of Z_1 that you found in (a).
(c) Give the resulting natural nonparametric estimators of the three functions involved in (a) in terms of the observed data.
(d) Use the estimators in (c) to give the nonparametric MLE of the cumulative hazard function Λ.
(e) State a general formula expressing an arbitrary survival function $1 - F$ in terms of the corresponding hazard function Λ.
(f) Combine the results of (d) and (e) to give an explicit expression for the Kaplan-Meier (nonparametric maximum likelihood) estimator of $1 - F$.

Solution: (a) $H^{uc}(t) = P(Z_1 \le t, \Delta_1 = 1) = \int_{[0,t]}(1 - G(s-))dF(s)$, $H^c(t) \equiv P(Z_1 \le t, \Delta_1 = 0) = \int_{[0,t]}(1 - F(s))dG(s)$, while $1 - H(z) = P(Z_1 > z) = P(X_1 > z, Y_1 > z) = (1-F(z))(1-G(z))$. This yields $P(Z_1 \ge z) = 1 - H(z-) = (1 - F(z-))(1 - G(z-))$.
(b) The cumulative hazard function Λ corresponding to F is given by

$$
\begin{aligned}
\Lambda(t) &= \int_{[0,t]} \frac{1}{1 - F(s-)} dF(s) = \int_{[0,t]} \frac{1 - G(s-)}{(1 - F(s-))(1 - G(s-))} dF(s) \\
&= \int_{[0,t]} \frac{1}{1 - H(s-)} dH^{uc}(s).
\end{aligned}
$$

(c) The natural estimators of H^{uc} and H^c are simply the functions $\mathbb{H}_n^{uc}(t) = n^{-1}\sum_{i=1}^n 1\{Z_i \le t, \Delta_i = 1\} = n^{-1}\sum_{i=1}^n \Delta_i 1\{Z_i \le t\}$, $\mathbb{H}_n^c(t) = n^{-1}\sum_{i=1}^n 1\{Z_i \le t, \Delta_i = 0\} = n^{-1}\sum_{i=1}^n (1 - \Delta_i) 1\{Z_i \le t\}$, and the natural estimator of $1 - H$ is

$$
1 - \mathbb{H}_n(t) = 1 - n^{-1}\sum_{i=1}^n 1\{Z_i \le t\} = n^{-1}\sum_{i=1}^n 1\{Z_i > t\}.
$$

(d) Combining (b) and (c), the natural estimator $\hat{\Lambda}_n$ of Λ is just

$$
\widehat{\Lambda}_n(t) = \int_{[0,t]} \frac{1}{1 - \mathbb{H}_n(s-)} d\mathbb{H}_n^{uc}(s).
$$

(e) If Λ is the cumulative hazard function corresponding to F, then

$$
1 - F(t) = \exp(-\Lambda_c(t)) \prod_{s \le t}(1 - \Delta\Lambda(s))
$$

where $\Lambda_c(t) \equiv \Lambda(t) - \sum_{s \le t} \Delta\Lambda(s)$ is the continuous part of Λ.
(f) Since $\widehat{\Lambda}_n$ given in (d) is discrete with $\widehat{\Lambda}_{n,c} = 0$, it follows from the relation in (e) that the Kaplan-Meier estimator $1 - \widehat{F}_n$ of $1 - F$ is given by

$$1 - \widehat{F}_n(t) = \prod_{s \le t} (1 - \Delta\widehat{\Lambda}_n(s)) = \prod_{s \le t} \left(1 - \frac{\Delta \mathbb{H}_n^{uc}(s)}{1 - \mathbb{H}_n(s-)}\right).$$

4. (42 points) Suppose that $X \sim \text{Uniform}(0,1)$.
(a) Find the hazard rate function $\lambda(t) = f(t)/(1 - F(t))$ where $f = f_X$ and $F(t) = F_X(t) = P(X \le t)$.
(b) Find the cumulative hazard function $\Lambda(t) = \int_{[0,t]} \lambda(s) ds$.
(c) State a general formula expressing a survival function $1 - F(t)$ to its corresponding cumulative hazard function Λ.
(d) Specialize the formula in (c) to the particular case in (a) and (b).
(e) Explain how this continues to hold if we replace F as in (a) by $G = .5F + .51_{[1/2,1]} = .5 + .5\delta_{1/2}$; compute the cumulative hazard function Λ_G corresponding to G, and show that the identity in (c) holds.

Solution: (a) When $X \sim \text{Unif}(0,1)$, $f(t) = 1_{[0,1]}(t)$ and $1 - F(t) = 1 - t$, $0 \le t \le 1$. Thus $\lambda(t) = 1/(1-t)$ for $0 \le t < 1$, $\lambda(t) = \infty$ for $t \ge 1$, and $\lambda(t) = 0$ for $t < 0$.
(b) The cumulative hazard function

$$\Lambda(t) = \int_{[0,t]} (1 - F(y))^{-1} dF(y) = -\log(1 - F(t)) = -\log(1 - t),$$

for $0 \le t < 1$.
(c) $1 - F(t) = \exp(-\Lambda_c(t)) \prod_{s \le t} (1 - \Delta\Lambda(s))$ where $\Delta\Lambda(s) = \Lambda(s) - \Lambda(s-)$ and $\Lambda_c(t) = \Lambda(t) - \sum_{s \le t} \Delta\Lambda(s)$.
(d) For the case in (a) and (b) Λ is continuous, so $\Lambda_c = \Lambda$, $\Delta\Lambda(t) = 0$ for all t, and we have, for $0 \le t < 1$,

$$1 - t = 1 - F(t) = \exp(-\Lambda(t)) = \exp(-(-\log(1 - t))) = 1 - t,$$

so the identity holds.

(e) When $G = .5F + .5\delta_{1/2}$, we compute

$$\begin{aligned}
\Lambda_G(t) &= \int_{[0,t]} \frac{1}{1 - .5s - .51_{[1/2,1]}(s)} dG(s) \\
&= \begin{cases} \int_{[0,t]} \frac{1}{1-.5s}(.5)ds, & 0 \le t < 1/2, \\ \Lambda_G(1/2-) + .5\frac{1}{1-.5^2}, & t = 1/2, \\ \Lambda_G(1/2) + 2^{-1}\int_{(1/2,t]} \frac{1}{1-s/2-1/2}ds, & \end{cases} \\
&= \begin{cases} -\log(1-t/2), & 0 \le t < 1/2, \\ -\log(3/4) + 2/3, & t = 1/2, \\ -\log(3/4) + 2/3 - \log(1-s)|_{1/2}^{t}, & 1/2 < t < 1, \end{cases} \\
&= \begin{cases} -\log(1-t/2), & 0 \le t < 1/2, \\ -\log(3/4) + 2/3, & t = 1/2, \\ -\log(3/4) + 2/3 - \log[2(1-t)], & 1/2 < t < 1. \end{cases}
\end{aligned}$$

Then

$$\begin{aligned}
\Lambda_{G,c}(t) &= \Lambda_G(t) - \sum_{s \le t} \Delta\Lambda_G(s) \\
&= \begin{cases} -\log(1-t/2), & 0 \le t \le 1/2, \\ -\log(3/4) - \log[2(1-t)], & 1/2 \le t < 1, \end{cases}
\end{aligned}$$

so that

$$1 - G(t) = \begin{cases} 1 - t/2, & 0 \le t < 1/2, \\ (1-t)/2, & 1/2 \le t \le 1, \end{cases}$$

while, on the other hand,

$$\begin{aligned}
&\exp(-\Lambda_{G,c}(t)) \prod_{s \le t} (1 - \Delta\Lambda_G(s)) \\
&= \begin{cases} \exp(\log(1-t/2)), & 0 \le t < 1/2, \\ (1-1/4)\cdot(1/3) = 1/4, & t = 1/2, \\ \exp\left(\log(3/4) + \log(2(1-t))\right)\cdot(1/3), & 1/2 \le t \le 1, \end{cases} \\
&= \begin{cases} (1-t/2), & 0 \le t < 1/2, \\ 1/4, & t = 1/2, \\ (3/4)\cdot 2(1-t)\cdot(1/3) = (1-t)/2, & 1/2 \le t \le 1. \end{cases}
\end{aligned}$$

Thus the identity stated in (c) holds.

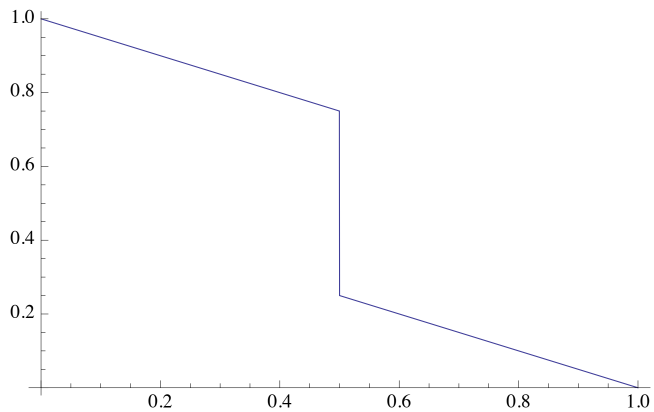

Figure 1: Survival function $1 - G$, problem 4

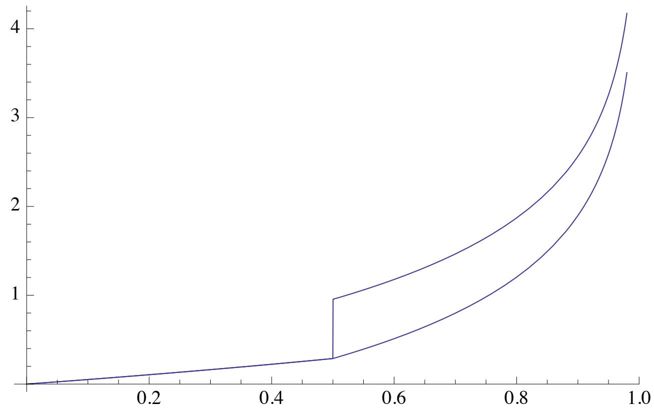

Figure 2: Hazard function Λ_G and its continuous part $\Lambda_{G,c}$, problem 4

Do **one** of problems 5-7.

5. (36 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. P_{θ_0} where $\mathcal{P} = \{P_\theta : \theta \in [1, \infty)\}$ and where P_θ is the Pareto(θ, α) distribution with density

$$p(x; \theta) \equiv p_\theta(x) = \frac{\alpha}{\theta} \left(\frac{\theta}{x} \right)^{\alpha+1} 1_{[\theta,\infty)}(x);$$

here α is assumed to be known.
(a) Find the MLE $\widehat{\theta}_n$ of θ.
(b) Give a proof of consistency of $\hat{\theta}_n$ using Wald's theorem and assuming that $\Theta = [1, M]$ for some $1 < M < \infty$.
(c) Give a direct proof that $\widehat{\theta}_n \rightarrow_p \theta_0$.

Solution. (a) The likelihood for θ of $X_1, \ldots, X_n$ in the given model is

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^{n} \left(\frac{\alpha}{\theta}\right) \left(\frac{\theta}{X_i}\right)^{\alpha+1} 1_{[\theta,\infty)}(X_i) \\ &= \frac{\alpha^n}{\prod_1^n X_i^{\alpha+1}} \theta^{n\alpha} 1\{\theta \le X_{(1)}\} \end{aligned}$$

where $X_{(1)} = \min_{1 \le i \le n} X_i$. This is maximized by $\hat{\theta}_n = X_{(1)}$.
(b) $\Theta = [1, M]$ is compact, and the function $\theta \mapsto p(x;\theta)$ is upper semi-continuous for each fixed x. Furthermore the functions

$$\begin{aligned} f(x,\theta) &= \log \frac{p_\theta(x)}{p_{\theta_0}(x)} \\ &= -\infty \cdot 1\{\theta_0 \le x < \theta\} + \alpha \log(\theta/\theta_0) \cdot 1_{[x \ge \theta]} \\ &\le F(x) \equiv \alpha \log(x/\theta_0) \end{aligned}$$

which satisfies

$$EF(X) = \int_{\theta_0}^{\infty} \alpha \log(x/\theta_0) \frac{\alpha}{\theta_0} \left(\frac{\theta_0}{x}\right)^{\alpha+1} dx < \infty$$

for every $\alpha > 0$. The function $\sup_{\theta:|\theta-\theta_0|\le\rho} p(x,\theta)$ is measurable for all small ρ by the same argument we used in the Uniform$(0,\theta)$ case. Since the model is identifiable, Wald's theorem yields $\hat{\theta}_n \to_{a.s.} \theta_0$.
(c) Now $P_{\theta_0}(\hat{\theta}_n < \theta_0) = P_{\theta_0}(X_{(1)} < \theta_0) = 0$ and, for every $\epsilon > 0$,

$$P_{\theta_0}(X_{(1)} > \theta_0 + \epsilon) = P(X_1 > \theta_0 + \epsilon)^n = (1 + \epsilon/\theta_0)^{-\alpha n} \to 0$$

as $n \to \infty$ since $1 - F_{\theta_0}(x) = P_{\theta_0}(X_1 > x) = (x/\theta_0)^{-\alpha}$. Thus $\hat{\theta}_n \to_p \theta_0$. Since $\sum_1^\infty P_{\theta_0}(X_{(1)} > \theta_0 + \epsilon) < \infty$ for every $\epsilon > 0$, we also have $\hat{\theta}_n \to_{a.s.} \theta_0$.

6. (36 points). Consider the (counter-)example given by Ferguson ACILST, pages 116-117 and Ferguson (1982):

$$p_\theta(x) = \frac{(1-\theta)}{\delta(\theta)} g_1\left(\frac{x-\theta}{\delta(\theta)}\right) + \theta g_0(x)$$

for $0 \le \theta \le 1$ where

$$\begin{aligned} &g_1(x) = (1-|x|)1_{[-1,1]}(x), \quad \text{the triangular density on } [-1,1], \\ &g_0(x) = 2^{-1}1_{[-1,1]}(x), \quad \text{the uniform density on } [-1,1], \\ &\delta(\theta) = (1-\theta)\exp(-(1-\theta)^{-c} + 1) \end{aligned}$$

for some $c > 0$. Consider the class of functions $\mathcal{F} = \{\log p_\theta(x) - \log p_{\theta_0}(x) : \theta \in [0,1]\}$. Show that any envelope function F of the class $\mathcal{F}$ satisfies $EF(X) = \infty$

if $c \geq 1$.

Hint: Note that: (i) $p_{\theta_0}(x) \leq p_{\theta_0}(\theta_0) =$ a constant; (ii) $\theta \mapsto p_\theta(x)$ is biggest when $\theta = \max\{0, x\}$. Use these two facts together with the form of δ to get a lower bound for

$$\sup_{\theta \in [0,1]} (\log p_\theta(x) - \log p_{\theta_0}(x))$$

which is not integrable under the stated condition.

Solution: Since $p_{\theta_0}(x) \leq p_{\theta_0}(\theta_0) = (1-\theta_0)/\delta(\theta_0) + \theta_0/2$, a constant, it follows that

$$\begin{aligned}
\sup_{\theta \in [0,1]} \log \frac{p_\theta(x)}{p_{\theta_0}(x)} &\geq \sup_{\theta \in [0,1]} \log \frac{p_\theta(x)}{p_{\theta_0}(\theta_0)} \\
&\geq \log \frac{p_x(x)}{p_{\theta_0}(\theta_0)} 1_{[0,1]}(x) \\
&= \log \left\{ \frac{1-x}{\delta(x)} \right\} 1_{[0,1]}(x) - \log p_{\theta_0}(\theta_0) \\
&= \left\{ (1-x)^{-c} + 1 \right\} 1_{[0,1]}(x) - \log p_{\theta_0}(\theta_0).
\end{aligned}$$

Therefore,

$$\begin{aligned}
E\left\{ \sup_{\theta \in [0,1]} \log \frac{p_\theta(X)}{p_{\theta_0}(X)} \right\} &\geq E_{\theta_0}[\left\{ (1-X)^{-c} + 1 \right\} 1_{[0,1]}(X)] - \log p_{\theta_0}(\theta_0) \\
&\geq \frac{\theta_0}{2} \int_0^1 (1-x)^{-c} dx - \log p_{\theta_0}(\theta_0) \\
&= \frac{\theta_0}{2} \cdot \infty = \infty \quad \text{if } \ c \geq 1.
\end{aligned}$$

7. (36 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. with mixture density

$$p(x; \mu, \nu, \theta) = \frac{\theta}{2} \exp(-|x-\mu|) + \frac{1-\theta}{2} \exp(-|x-\nu|), \qquad x \in \mathbb{R},$$

where $0 < \theta < 1$, $\mu, \nu \in \mathbb{R}$, $\mu \neq \nu$. In other words, p is the mixture of two Laplace distributions with parameters μ and ν respectively.

(a) Describe an EM - algorithm for estimation of (μ, ν, θ).

(b) What hypothesis do you need to conclude that the EM algorithm in (a) converges to stationary points of the incomplete data log-likelihood? Does it hold in this case?

Solution: (a) Here it is natural to let the "complete data" $\underline{X}$ be $(X_1, \delta_1), \ldots, (X_n, \delta_n)$ where $\delta_i \in \{0, 1\}$ and (X_i, δ_i) are i.i.d. with density

$$p(x, \delta; \theta, \mu, \nu) = \left(\frac{\theta}{2} \exp(-|x-\mu|) \right)^\delta \left(\frac{1-\theta}{2} \exp(-|x-\nu|) \right)^{1-\delta}$$

for $(x, \delta) \in \mathbb{R} \times \{0, 1\}$. Then the incomplete $\underline{Y}$ is $X_1, \ldots, X_n$, which are i.i.d. with the mixture distribution

$$p(x; \mu, \nu, \theta) = \frac{\theta}{2} \exp(-|x - \mu|) + \frac{1-\theta}{2} \exp(-|x - \nu|).$$

It follows that conditional on $X = x$, δ is Bernoulli$(p(x))$ where

$$p(x) \equiv p(x; \theta, \mu, \nu) = \frac{\theta \exp(-|x - \mu|)}{\theta \exp(-|x - \mu|) + (1-\theta) \exp(-|x - \nu|)}. \tag{1}$$

Hence $E(\delta|X) = p(X)$; this is the basis of the E - step of an EM algorithm.

To find the M - step, note that

$$\begin{aligned} l(\theta, \lambda, \mu | X, \delta) &= \delta\{\log \theta - |X - \mu|\} + (1-\delta)\{\log(1-\theta) - |X - \nu|\} \\ &\quad + \text{constant}, \end{aligned}$$

so that the scores (for a sample of size one) are

$$\begin{aligned} \dot{l}_\theta(X, \delta) &= \frac{\delta}{\theta} - \frac{1-\delta}{1-\theta}, \\ \dot{l}_\mu(X, \delta) &= \delta \left\{ 1_{[X<\mu]} - 1_{[X>\mu]} \right\} \overset{a.s.}{=} \delta \left\{ 2 1_{[X \le \mu]} - 1 \right\}, \\ \dot{l}_\nu(X, \delta) &= (1-\delta) \left\{ 1_{[X<\nu]} - 1_{[X>\nu]} \right\} \overset{a.s.}{=} (1-\delta) \left\{ 2 1_{[X \le \nu]} - 1 \right\}. \end{aligned}$$

Thus the score equations are solved by

$$\widehat{\mu}_n = \text{median}\{X_i : \ \delta_i = 1\}, \qquad \widehat{\nu}_n = \text{median}\{X_i : \ \delta_i = 0\}, \qquad \widehat{\theta}_n = \frac{\sum \delta_i}{n}.$$

This is the basis of an M - step.

Set $\theta^{(0)} = 1/2$, $\widehat{\mu}^{(0)} = \widehat{\nu}^{(0)} = \text{median}(X_i's) = \mathbb{F}_n^{-1}(1/2)$. Then, for $m = 0, 1, \ldots$, define

$$\widehat{\delta}_i^{(m)} \equiv p(X_i; \widehat{\theta}^{(m)}, \widehat{\mu}^{(m)}, \widehat{\nu}^{(m)}) \tag{2}$$

where $p(x; \theta, \mu, \nu)$ is given by (1). Note that a slight complication here is that our estimators of the δ_i's take on values in $(0, 1)$, and are not exactly zero or one. Thus instead of the medians of the X_i's with $\delta_i = 1$ and 0 respectively, we will estimate the empirical distributions involved in the score equations by

$$\begin{aligned} J^{(m)} &\equiv \sum_{j=1}^n \hat{\delta}_i^{(m)}, \\ \mathbb{F}_{J^{(m)},1}(x) &\equiv \frac{1}{J^{(m)}} \sum_{j=1}^n \hat{\delta}_i^{(m)} 1_{[X_i \le x]}, \\ \mathbb{F}_{J^{(m)},0}(x) &\equiv \frac{1}{n - J^{(m)}} \sum_{j=1}^n (1 - \hat{\delta}_i^{(m)}) 1_{[X_i \le x]}, \end{aligned}$$

and then set

$$\widehat{\mu}^{(m+1)} = \mathbb{F}^{-1}_{J^{(m)},1}(1/2)\,, \tag{3}$$

$$\widehat{\nu}^{(m+1)} = \mathbb{F}^{-1}_{J^{(m)},0}(1/2)\,, \tag{4}$$

$$\widehat{\theta}^{(m+1)} = \frac{\sum \widehat{\delta}_i^{(m)}}{n}\,. \tag{5}$$

Iteration of (2) and (3,4,5) yields an EM algorithm for estimation of (θ, μ, ν).

(b) Let $l(\underline{\theta}|\underline{X})$ denote the complete data log-likelihood where $\underline{\theta} = (\theta, \mu, \nu)$, and set

$$Q(\underline{\theta}|\underline{\theta}_0, \underline{Y}) = E_{\underline{\theta}_0}\{l(\underline{\theta}|\underline{X})\big|\underline{Y}\}$$

denote the conditional expectation of $l(\underline{\theta}|\underline{X})$ given the incomplete data $\underline{Y}$ under $\underline{\theta}_0$. If Q is continuous in both $\underline{\theta}$ and $\underline{\theta}_0$, then the EM algorithm converges monotonically to a stationary point of $l(\underline{\theta}|\underline{Y})$. In the present case we can write, using $X_i = T(X_i, \delta_i)$ for $i = 1, \ldots, n$,

$$\begin{aligned}
Q(\underline{\theta}|\underline{\theta}_0, \underline{Y}) &= E_{\underline{\theta}_0}\left\{\sum_{i=1}^{n} (\delta_i[\log\theta - |X_i - \mu|] + (1-\delta_i)[\log(1-\theta) - |X_i - \nu|])\,\big|\underline{Y}\right\} \\
&= \sum_{i=1}^{n} \big(E_{\underline{\theta}_0}(\delta_i|X_i)[\log\theta - |X_i - \mu|] + E\{(1-\delta_i)|X_i\}[\log(1-\theta) - |X_i - \nu|]\big) \\
&= \sum_{i=1}^{n} \{p(X_i; \theta_0, \mu_0, \nu_0)\cdot[\log\theta - |X_i - \mu|] \\
&\qquad + (1 - p(X_i; \theta_0, \mu_0, \nu_0))\cdot[\log(1-\theta) - |X_i - \nu|]\}\,,
\end{aligned}$$

where $p(x; \theta, \mu, \nu)$ is given by (1). This is a continuous function of $\underline{\theta}_0$ and $\underline{\theta}$ on $\Theta\times\Theta$ with $\Theta = (0,1)\times\mathbb{R}^2$. Thus the EM algorithm will converge to a stationary point of the incomplete data log-likelihood.