Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ where each $P_\theta$ has density $p(x, \theta)$ with respect to some dominating measure $\nu$. Suppose that $X_1, \ldots, X_n$ are i.i.d. $P_{\theta_0}$ for $\theta_0 \in \Theta$, and let $\hat{\theta}_n$ denote any sequence of maximum likelihood estimators of $\theta$.

Wald’s consistency theorem for the sequence $\{\hat{\theta}_n\}$ yields consistency under the following hypotheses:

(a) Compactness of $\Theta$.
(b) Upper semi-continuity of the maps $\theta \mapsto p(x, \theta)$ for all $x$.
(c) Existence of an integrable upper envelope $F$ for the class of functions $f_{\theta}(x) \equiv \log p(x, \theta) - \log p(x, \theta_0)$.
(d) Measurability of $x \mapsto \sup_{|\theta - \theta'| < \rho} p(x, \theta')$ for all $\theta \in \Theta$ and all small $\rho$.
(e) Identifiability of the model $\mathcal{P}$: i.e. $p(x, \theta) = p(x, \theta_0)$ a.e. $\nu$ implies $\theta = \theta_0$.

Here we give a counterexample illustrating that the hypothesis (c) cannot be dropped. Suppose that

$$p(x, \theta) = \frac{(1 - \theta)}{\delta(\theta)} \left(1 - \frac{|x - \theta|}{\delta(\theta)}\right) 1_{[\theta - \delta(\theta), \theta + \delta(\theta)]}(x) + \frac{\theta}{2} 1_{[-1,1]}(x)$$

for $\theta \in [0, 1] \equiv \Theta$ where $\delta(\theta) = (1 - \theta) \exp(-(1 - \theta)^{-c} + 1)$. Note that $0 \leq \delta(\theta) \leq 1$. Let $f_{\theta}(x) \equiv \log p(x, \theta) - \log p(x, \theta_0)$, and define the class of functions $\mathcal{F}$ by

$$\mathcal{F} = \{f_{\theta} : \theta \in \Theta\}.$$ 

Note that $\Theta = [0, 1]$ is compact. Furthermore the functions $\theta \mapsto p(x, \theta)$ are continuous for every $x \in [-1, 1] \equiv \mathcal{X}$. Since a continuous function on a compact set attains its supremum, an MLE $\hat{\theta}_n$ exists for every $n$. The model is clearly identifiable. But Ferguson (1982), *J. Amer. Statist. Assoc.* 77, 831-834, shows that $\hat{\theta}_n \to_{a.s.} 1$ no matter what $\theta_0$ is true.

The key hypothesis of Wald’s theorem violated in Ferguson’s example is the existence of an integrable upper envelope. To see this, note that for $0 \leq x \leq 1$

$$\sup_{0 \leq \theta \leq 1} p(x, \theta) \geq p(x, x) = \frac{1 - x}{\delta(x)} \left(1 - \frac{|x - x|}{\delta(x)}\right) + (x/2) \geq \exp((1 - x)^{-c} - 1).$$
Therefore
\[ \sup_{0 \leq \theta \leq 1} \log \frac{p(x, \theta)}{p(x, \theta_0)} \geq \log \frac{p(x, x)}{p(x, \theta_0)} = (1 - x)^{-c} - 1 - \log p(x, \theta_0), \]
and hence, if \( \theta_0 \in (0, 1] \) we have
\[
E_{\theta_0} \left\{ \sup_{0 \leq \theta \leq 1} \log \frac{p(X, \theta)}{p(X, \theta_0)} \right\} \geq \frac{\theta_0}{2} \int_0^1 (1 - x)^{-c} dx - a \text{ constant} = \infty
\]
for all \( c \geq 1 \), while
\[
E_0 (1 - X)^{-c} 1_{(0,1]}(X) = \int_0^1 (1 - x)^{-c}(1 - x) dx = \infty
\]
if \( c \geq 2 \). But any envelope function must be at least as big as \( \sup_{0 \leq \theta \leq 1} \log \frac{p(X, \theta)}{p(X, \theta_0)} \). Hence an integrable envelope function does not exist for any \( \theta_0 \in [0, 1] \) when \( c \geq 2 \).

To see that the sequence of MLEs \( \{\hat{\theta}_n\} \) is inconsistent, first note that the log-likelihood function is given by
\[
l_n(\theta) = \sum_{i=1}^n \log p(X_i, \theta) = N_{\theta} \log(\theta/2) + n \mathbb{P}_n \left\{ \log \left( \frac{1 - \theta}{\delta(\theta)} \right) \left( 1 - \frac{|X - \theta|}{\delta(\theta)} \right) + \frac{\theta}{2} \right\} 1_{[\theta - \delta(\theta), \theta + \delta(\theta)]}(X) \]
where \( N_{\theta} \) denotes the number of observations in \( [\theta - \delta(\theta), \theta + \delta(\theta)] \). For each fixed \( \alpha < 1 \)
\[
\sup_{0 \leq \theta \leq \alpha} n^{-1} l_n(\theta) \leq \sup_{0 \leq \theta \leq \alpha} \log \left( \frac{1 - \theta}{\delta(\theta)} + \frac{\theta}{2} \right) \leq \log \left( \frac{1}{\delta(\alpha)} + \frac{1}{2} \right).
\]
But we will show that \( \sup_{0 \leq \theta \leq \alpha} n^{-1} l_n(\theta) \rightarrow_{a.s.} \infty \), which will therefore imply that \( \hat{\theta}_n \rightarrow_{a.s.} 1 \).
To do this, let \( M_n = \max\{X_1, \ldots, X_n\} \). Then \( M_n \rightarrow_{a.s.} 1 \) for each \( \theta_0 \in [0, 1] \) and
\[
\sup_{0 \leq \theta \leq 1} n^{-1} l_n(\theta) \geq n^{-1} l_n(M_n) \geq \frac{n - 1}{n} \log \frac{M_n}{2} + \frac{1}{n} \log \left( \frac{1 - M_n}{\delta(M_n)} + \frac{M_n}{2} \right).
\]
Thus
\[
\liminf_{n \to \infty} \sup_{0 \leq \theta \leq 1} n^{-1} l_n(\theta) \geq \liminf_{n \to \infty} \left\{ \frac{1}{n} \log \left( \frac{1 - M_n}{\delta(M_n)} \right) \right\} - 2 \rightarrow_{a.s.} \infty
\]
since $\delta(\theta)$ decreases to 0 very rapidly as $\theta \to 1$. To show this explicitly note that $\log[(1 - \theta)/\delta(\theta)] = (1 - \theta)^{-c} - 1$, so
\[
\frac{1}{n} \log \frac{1 - M_n}{\delta(M_n)} = \frac{1}{n(1 - M_n)^c} - \frac{1}{n} \to_{\text{a.s.}} \infty
\]
if $n(1 - M_n)^c \to_{\text{a.s.}} 0$. But for $x \geq x_0(c)$, $P_{\theta_0}(X \leq x) \leq P_0(X \leq x)$ and hence for $n \geq \text{some } N_{\epsilon,c}$,
\[
P_{\theta_0}(n(1 - M_n)^c > \epsilon) = P_{\theta_0}(X_1 < 1 - (\epsilon/n)^{1/c})^n \leq P_0(X_1 < 1 - (\epsilon/n)^{1/c})^n
\]
\[
= (1 - P_0(X_1 \geq 1 - (\epsilon/n)^{1/c})^n
\]
\[
= \left(1 - \frac{1}{2}(\epsilon/n)^{2/c}\right)^n \leq \exp(-(1/2)\epsilon^{2/c}n^{1-2/c})
\]
which has a finite sum on $n$ if $c > 2$. Thus by Borel-Cantelli, $P_{\theta_0}(n(1 - M_n)^c > \epsilon \ ; \ i.o.) = 0$ and $n(1 - M_n)^c \to_{\text{a.s.}} 0$ if $c > 2$.

![Figure 1: $\delta_c(\theta) \equiv (1 - \theta) \exp(-(1 - \theta)^{-c} + 1)$ for $c = 2, 3, 4$ and $0 \leq \theta \leq 1.$](image)
Figure 2: $\delta_c(\theta) \equiv (1 - \theta) \exp(-(1 - \theta)^{-c} + 1)$ for $c = 2, 3, 4$ for $0.5 \leq \theta \leq 0.7$.

Figure 3: $p(x, \theta)$ for $c = 2$, $\theta \in \{0, .25, .30, 1\}$, $x \in [-1, 1]$