

Statistics 583; Wald consistency theorem counterexample

Jon Wellner

Version 2, 8 April 2016

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ where each P_θ has density $p(x, \theta)$ with respect to some dominating measure ν . Suppose that X_1, \dots, X_n are i.i.d. P_{θ_0} for $\theta_0 \in \Theta$, and let $\widehat{\theta}_n$ denote any sequence of maximum likelihood estimators of θ .

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

- (a) Compactness of Θ .
- (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

$$f(x, \theta) \equiv \log p(x, \theta) - \log p(x, \theta_0).$$

- (d) Measurability of $x \mapsto \sup_{|\theta' - \theta| < \rho} p(x, \theta')$ for all $\theta \in \Theta$ and all small ρ .
- (e) Identifiability of the model \mathcal{P} : i.e. $p(x, \theta) = p(x, \theta_0)$ a.e. ν implies $\theta = \theta_0$.

Here we give a counterexample illustrating that the hypothesis (c) can not 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 $\widehat{\theta}_n$ exists for every n . The model is clearly identifiable. But Ferguson (1982), *J. Amer. Statist. Assoc.* **77**, 831-834, shows that $\widehat{\theta}_n \rightarrow_{a.s.} 1$ no matter what θ_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 $0 \leq x \leq 1$

$$\begin{aligned} \sup_{0 \leq \theta \leq 1} p(x, \theta) &\geq p(x, x) = \frac{1 - x}{\delta(x)} \left(1 - \frac{|x - x|}{\delta(x)}\right) + (x/2) \\ &\geq \exp((1 - x)^{-c} - 1). \end{aligned}$$

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

$$\begin{aligned} E_{\theta_0} \left\{ \sup_{0 \leq \theta \leq 1} \log \frac{p(X, \theta)}{p(X, \theta_0)} \right\} &\geq E_{\theta_0} (1-X)^{-c} 1_{(0,1)}(X) - \text{a constant} \\ &\geq \frac{\theta_0}{2} \int_0^1 (1-x)^{-c} dx - \text{a constant} \\ &= \frac{\theta_0}{2} \cdot \infty - \text{a constant} = \infty \end{aligned}$$

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

$$l_n(\theta) = \sum_{i=1}^n \log p(X_i, \theta) = N_\theta \log(\theta/2) + n \mathbb{P}_n \left\{ \log \left(\frac{1-\theta}{\delta(\theta)} \left(1 - \frac{|X-\theta|}{\delta(\theta)} \right) + \frac{\theta}{2} \right) 1_{[\theta-\delta(\theta), \theta+\delta(\theta)]}(X) \right\}$$

where N_θ denotes the number of observations in $[\theta - \delta(\theta), \theta + \delta(\theta)]^c$. 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 $M_n = \max\{X_1, \dots, X_n\}$. Then $M_n \rightarrow_{a.s.} 1$ for each $\theta_0 \in [0, 1]$ and

$$\begin{aligned} \sup_{0 \leq \theta \leq 1} n^{-1} l_n(\theta) &\geq n^{-1} l_n(M_n) \\ &\geq \frac{n-1}{n} \log \frac{M_n}{2} + \frac{1}{n} \log \left(\frac{1-M_n}{\delta(M_n)} + \frac{M_n}{2} \right). \end{aligned}$$

Thus

$$\liminf_{n \rightarrow \infty} \sup_{0 \leq \theta \leq 1} n^{-1} l_n(\theta) \geq \liminf_{n \rightarrow \infty} \left\{ \frac{1}{n} \log \left(\frac{1-M_n}{\delta(M_n)} \right) \right\} - \log 2 \rightarrow_{a.s.} \infty$$

since $\delta(\theta)$ decreases to 0 very rapidly as $\theta \rightarrow 1$. To show this explicitly note that $\log[(1 - \theta)/\delta(\theta)] = (1 - \theta)^{-c} - 1$, so

$$\frac{1}{n} \log \frac{1 - M_n}{\delta(M_n)} = \frac{1}{n(1 - M_n)^c} - \frac{1}{n} \xrightarrow{a.s.} \infty$$

if $n(1 - M_n)^c \xrightarrow{a.s.} 0$. But for $x \geq x_0(c)$, $P_{\theta_0}(X \leq x) \leq P_0(X \leq x)$ and hence for $n \geq$ some $N_{\epsilon, c}$,

$$\begin{aligned} P_{\theta_0}(n(1 - M_n)^c > \epsilon) &= P_{\theta_0}(X_1 < 1 - (\epsilon/n)^{1/c})^n \leq P_0(X_1 < 1 - (\epsilon/n)^{1/c})^n \\ &= (1 - P_0(X_1 \geq 1 - (\epsilon/n)^{1/c}))^n \\ &= \left(1 - \frac{1}{2}(\epsilon/n)^{2/c}\right)^n \leq \exp(-(1/2)\epsilon^{2/c}n^{1-2/c}) \end{aligned}$$

which has a finite sum on n if $c > 2$. Thus by Borel-Cantelli, $P_{\theta_0}(n(1 - M_n)^c > \epsilon \text{ i.o.}) = 0$ and $n(1 - M_n)^c \xrightarrow{a.s.} 0$ if $c > 2$.

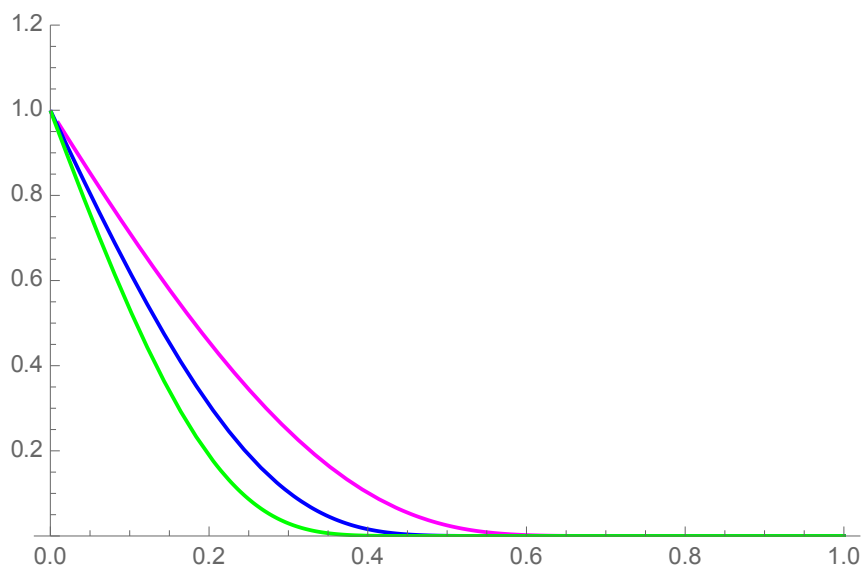


Figure 1: $\delta_c(\theta) \equiv (1 - \theta) \exp(-(1 - \theta)^{-c} + 1)$ for $c = 2, 3, 4$ and $0 \leq \theta \leq 1$.

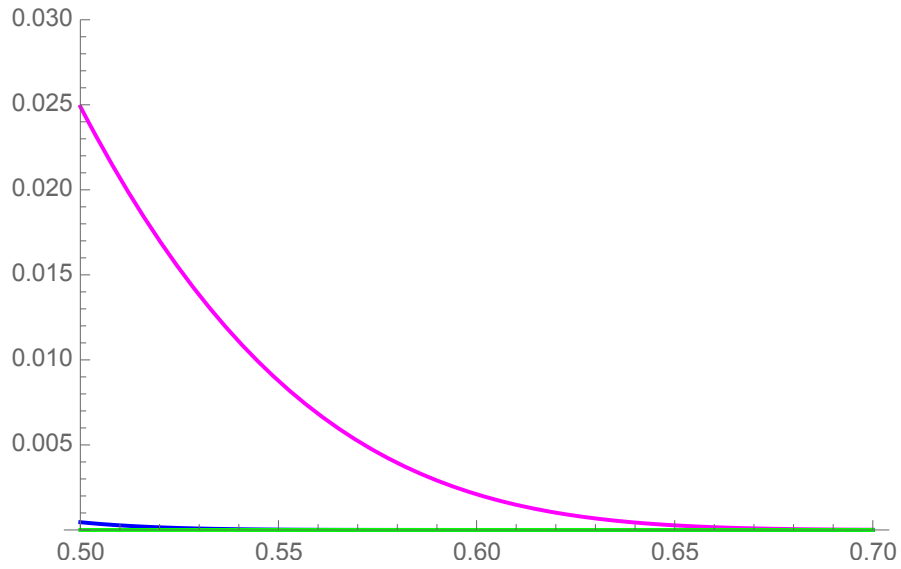


Figure 2: $\delta_c(\theta) \equiv (1 - \theta) \exp(-(1 - \theta)^{-c} + 1)$ for $c = 2, 3, 4$ for $.5 \leq \theta \leq .7$.

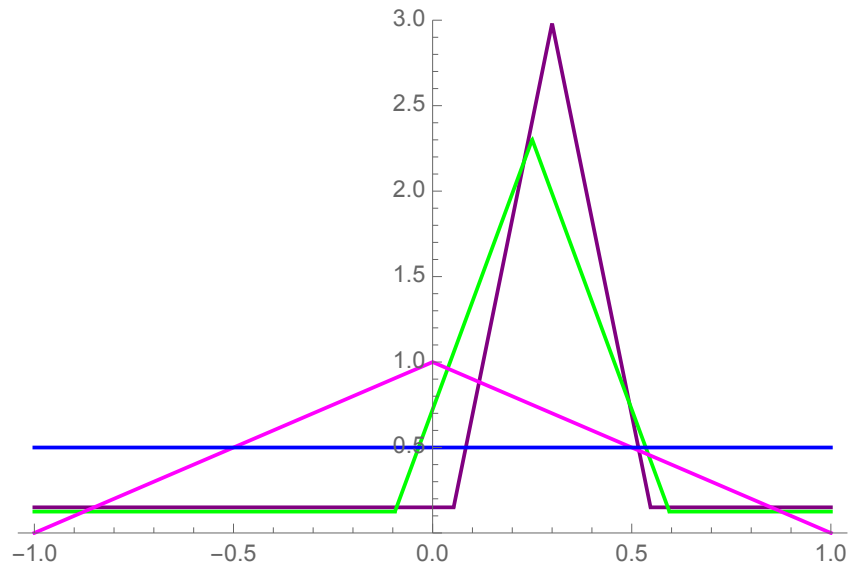


Figure 3: $p(x, \theta)$ for $c = 2$, $\theta \in \{0, .25, .30, 1\}$, $x \in [-1, 1]$